Analysis and comparison of Cloud Radiation Data sets of Long Term Cloud Coverage in the ISCCP, MODIS & SSM/I

Bachelor's Thesis

Usman Butt
Jesper Jørgen Gaardhøje
Anders Richard Pankoke Holm
June 2017

Supervisors:
Martin Andreas Bødker Enghoff
Henrik Svensmark

DTU Space
National Space Institute

Abstract

Global coverage of clouds have a great impact on the radiation balance of the Earth and is yet to be studied in detail. The main purpose of this thesis is to analyze and compare cloud coverage measurements based on readings from three different satellite systems; ISCCP, MODIS and SSMI. Based on radiance data from visible, infrared and microwave imagery a number of cloud properties are obtained.
The cloud coverage is analyzed in detail and categorized in high-, middle- and low-level clouds. The programs and scripts used in the analyze and for data processing are also discussed in the thesis. A regressional correction of the data is performed and applied. Discussion of global coverage of the different systems is made in order to determine the meaningfulness of radiance data obtained. The analysis part includes an intercomparison of data from areas off the coast of Angola and off the coast of Peru. These areas are chosen as they are known for having a high amount of low-clouds compared to middle and high. A possible link between the amount of galactic cosmic ray and the cloud formation on Earth is also discussed in the analysis.

The analysis shows that the ISCCP low cloud amount and the SSMI data is well-comparable in areas populated almost solely by clouds located in the lower parts of the atmosphere. Unfortunately only one area on Earth show these properties (an area off the coast of Angola, Africa).
The MODIS and ISCCP total cloud amount showed a rather weak correlation of 0.68, when compared in a overlapping period of 9 years. This was expected to be the most obvious comparison of all data sets.
A possible relationship was found between the ISCCP low cloud data and the galactic cosmic rays, after having done a regressional correction to the mid level cloud amount, and adding it to the ISCCP low level cloud amount (correcting a downwards trend from 1998-2009) we found a correlation of 0.34. Without this correction, the low cloud data only showed signs of a connection to galactic cosmic rays, in the first years after 1983.
We conclude overall that ISCCP, MODIS and SSMI show too many differences in their measurements for a comprehensive study of the physical effects, using all the data to be carried out at this time.

Preface

Following parts of the report was written and developed by:

Usman Butt

- Section 1: Introduction (with Anders and Jesper)
- Section 2: Clouds
- Section 3: Satellite Imagery
- Section 4: Satellite Systems
- Section 8: Discussion (with Jesper)

Anders Richard Pankoke Holm

- Section 1: Introduction (with Jesper and Usman)
- Section 5: Introduction to the data from the satellites
- Section 6: Data processing via MATLAB
- Development of MATLAB-scripts and functions (included in appendix A) for processing satellite data
- Drawing of figure 11, 12 and 13
- Visualization (through MATLAB) of figure 14 till 58

Jesper Jørgen Gaardhøje

- Section 1: Introduction (with Anders and Usman)
- Section 7: Analysis and Comparison
- Section 8: Discussion (with Usman)
- Section 9: Conclusion

With supervision of:

Henrik Svensmark
Martin Andreas Bødker Enghoff

Acknowledgement:

Thanks to Jacob Svensmark for providing the ISCCP D2 data for low, mid, high and total cloud amount, which is the one used and processed throughout the entire report.

Thanks to Rich Pawlowicz who, with the help of E. Firing, D. Byrne, M. Mann, J. Pringle, and J. E. Nilsen, have developed the `m_maps` package for MATLAB, which we use for plotting data onto world maps.

Thanks to Jens Jørgen Gaardhøje for proofreading of and idea inputs for the report.

Contents

1 Introduction

This paper contains a comparative analysis of data from 3 different satellite systems; ISCCP (International Satellite Cloud Climatology Project), SSMI (Special Sensor Microwave Imager) and MODIS (Moderate Resolution Imaging Spectroradiometer), which are all working with mapping of the cloud coverage of the Earth. The satellite systems are working with infrared imaging, visible image spectrometry and microwave radiometry, and thus have different ways of interpreting natural emission. We wish to get a better general image of the cloud coverage.

The main goal of this work is to estimate the likelihood of the change in observations in the low- and mid level cloud amounts measured by the ISCCP satellite system. For this reason the SSMI and MODIS data are included in this analysis. We wish to analyze if the SSMI and MODIS systems are showing the same tendencies.

Many factors influence the cloud coverage of the Earth. One of these is believed to be caused by the absorption of galactic cosmic rays in the atmosphere causing it to become ionized, thus creating "seeds" for new cloud for clouds to be formed. We wish to investigate this influence of the cloud coverage measured (by ISCCP and MODIS) on Earth in the period from 1983-2017 where data is available.

Section 2 covers the basics in cloud theory. The influence of clouds on the radiation budget is discussed along with a summary of the basic cloud types and information on how satellite system discovers these.

Section 3 discusses the different forms of imagery and how these are applied on satellites to obtain contrast photos which can be used for climatology purposes, like detection of cloud coverage.

Section 4 includes a description of the satellite systems which are investigated (ISCCP, MODIS and SSMI), the data, orbits and satellite coverage for each of the systems are discussed. Spatial resolution for the three systems are compared.

When understanding what data is collected and how this is done, in section 5 an overall look of the data will be presented to the reader, including how it is arranged as well as what we extract, period in which the individual satellites recorded and how the data is processed before doing any further work in MATLAB.

Section 6 takes the reader through various methods used in order to perfectly visualize the data given. Time series with overall representation of the data will be generated from a weighted mean algorithm and monthly means will be removed leaving us with the anomaly. Additionally, the reader will be taken through the methods in which we investigate trends and correct the data.

In doing this assessment of the satellite systems, each system will be analyzed individually before comparisons between them are made.

The SSMI data measures the "liquid water path", which is the total cloud liquid water contained in a vertical column of the atmosphere. This differs from ISCCP that measures the total mean cloud amount (and low- mid & high level cloud amounts individually), and MODIS that measures the total cloud fraction.

In order to be able to compare the SSMI data to the ISCCP data, two specific areas are chosen for further investigation. These areas are chosen because of their large fraction of clouds located in low altitudes. Because the temperatures in these lower altitudes provides for liquid water, every cloud measured by the ISCCP should,in theory, be directly comparable to the SSMI readings.

This analysis in section 7, will end up in a discussion of the uncertainties involved with measurements of the Earths cloud coverage, and in an overall assessment and conclusion of how comparable these satellite systems are.

2 Clouds

As a consequence of cooling of atmospheric water vapor due to vertical air motions (heated water vapor travels upward), clouds are formed. The water vapor condenses on microscopic particles in the air, sea salt, dust particles or bits of organic matter moving around in the air. Chemical aerosol particles provide the possibility for water vapor to condense in the same way as small particles in the air and is often made up by sulfuric acid. As the atmosphere is transparent to solar radiation (except for when clouds are present) the surface absorbs 70 percent of the solar radiation, causing the air near the surface of the Earth to be warmer than it is at higher altitudes.

The formation, evolution of motion of clouds is determined by interaction of cloud microphysical processes with atmospheric radiation and motion. As air moves past the particles in a cloud, a frictional force is exerted, as with anything on the Earth. The particles inside the cloud causes the air to move around the cloud, instead of moving through the cloud, this causes movement in the clouds. Differences in nature and different behaviour of the clouds in certain meterological situations produces different cloud types.

2.1 The effect of clouds on the radiation balance

Clouds may play a very important role in the radiation balance as the albedo of a cloudy part of the Earth is around 0.5 whereas the mean albedo of a cloudless part of the Earth is around 0.14 [1], which leads to a higher reflection of the radiation. Clouds can therefore affect the radiation balance of a vertical column underlying the particular cloud. The presence of clouds does also change the upward infrared radiation flux in the atmospheric temperature [4]. The opacity of the clouds in terms of planetary infrared radiation, would mean that an increase in clouds both in extent (horizontal size) and an increase in the effective cloud top heights, results in a reduction of the upward flux of the infrared radiation from the atmosphere of the Earth into space. This leads to two problems with increase of cloud cover on the Earth, first one being an increase of the albedo of the Earth, resulting in a decrease in the solar energy available at the surface of the Earth, and the second one being a decrease in the loss of the infrared radiation into the space. In order to measure changes in the cloud coverage which is actually significant to the changes happening to the climate and differences in the radiation balance, satellite systems which can provide high temporal and spatial resolution are required. The statistical accuracy required for a cloud observing system can be determined by estimating the magnitude of changes in cloud properties, that would lead to changes in the radiation balance which exceeds $0.5 W m^{-2}$, about 25% of this forcing is already produced by the abundances of greenhouse gases in our atmosphere [2]. Changes in the cloud properties could possibly explain some of the forcing happening, a change in the radiation balance with a forcing at the same order of magnitude as mentioned, would be produced by a change in cloud amount, top pressure, optical thickness and particle radius of around 1%, 10mb, 0.15 and 0.3μm [3].

2.2 Cloud Types

The different cloud types are characterized as following [13]:

2.2.1 High-level clouds

High-level clouds occur above about 6000 meters of height and are given the prefix “cirro”. Temperatures are freezing in the troposphere, which means that the clouds in this area are composed of mostly ice crystals.

- **Cirrus** clouds are thin, very opaque and composed entirely of ice crystals. They are often a sign of an approaching warm front or upper-level jet streak.
- **Cirrostratus** clouds form a more spread, veil-like layer. Cirrostratus look like a ”cover” on the sky. As a warm front approaches, cirrus tend to thicken into cirrostratus, which may thicken and lower into mid-level clouds.

2.2.2 Mid-level clouds

The clouds in the middle level of the troposphere are given the prefix “alto”. These appear between 2000 and 6000 meters of height. The temperature structure of the troposphere, altitude and time of year, decides which phase the water in the clouds at this height are in, these clouds can be composed of liquid water droplets, ice crystals or a combination of the two, including supercooled droplets.

- **Altostratus** clouds are “strato” type clouds (see section on low-level clouds). They can often indicate the approach of a warm front and may thicken and lower into stratus, then nimbostratus resulting in rain or snow.

  Although altostratus clouds can sprinkle a little water or very light showers, these do not precipitate in general.
- **Altocumulus** clouds are characterized as “cumulo” type clouds (low-level cloud behaviour) in mid-levels, as they tend to be heap like clouds, with a tendency of convection.

  Altocumulus align in rows of clouds, which indicates areas in the cloud where moist air is ascending, and air in between which is caused by drier air descending in these areas.

2.2.3 Low-level clouds

Low clouds occur below 2000 meters, and normally consist of liquid water droplets or supercooled droplets, the weather does also affect these, and freezing temperatures and cold winter storms where the clouds are comprised of mainly ice crystals and snow.
The two main types of low clouds include **stratus**, their development is horizontal, and **cumulus**, which have an vertical development. Stratus clouds are uniform and flat, and do not precipitate (light precipitation may occur). The stratus clouds form a gray coloured cloud cover.

Stratocumulus are clouds clumped in thick and thin areas.They appear frequently in the atmosphere, ahead or behind a frontal system. Thick, dense stratus or stratocumulus clouds producing steady rain or snow are called **nimbostratus** clouds. Unlike layered, horizontal stratus, cumulus clouds are more individual in nature and spread on the sky, these develop vertically.

2.3 ISCCP Cloud Classification

The ISCCP data sets are obtained from passive measurements of radiation reflected and emitted by the clouds.
This means that the data set can not provide information about the vertical mass or size of the clouds.
They can show the correlated variations of cloud properties that are characteristic of different kinds of atmospheric motions and indicate the vertical distribution of cloud top locations. The distribution of cloud properties and their correlated variations is best illustrated by two-dimensional frequency distributions of the cloud top pressure (PC) and visible optical thickness (TAU), the ISCCP dataset divides the PC and TAU range into nine basic categories as shown on figure 1.

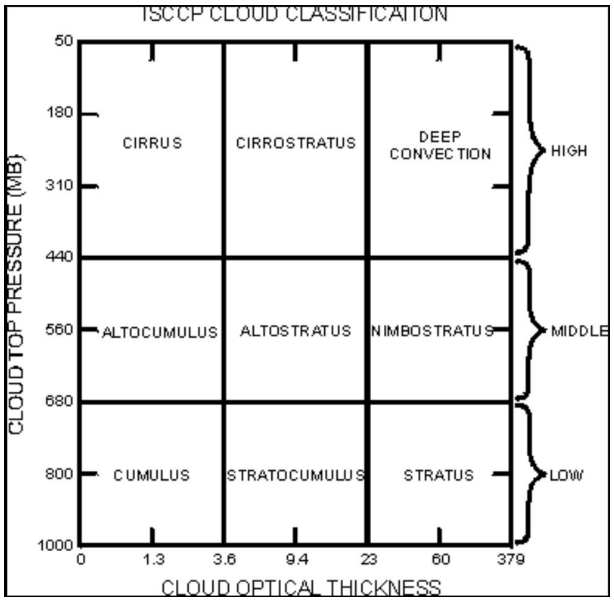


Figure 1: Overview of cloud types by their Optical thickness and cloud top pressure. High-, middle- and low-level clouds are classified in these categories in the ISCCP data sets.
Credit: ISCCP official site 20/6-2017

The PC is a name for the pressure of the part of the cloud with the highest altitude, PC describes the pressure at this height. TAU is the optical thickness or optical depth of the clouds. TAU is calculated by taking the natural logarithm of the ratio of the incident radiant power to the transmitted radiant power and is dimensionless. The optical depth is given by:

$$\tau = \ln\left(\frac{\Phi_e^i}{\Phi_e^t}\right) \tag{1}$$

Where Φ_e^i is the radiant flux received by the cloud and Φ_e^t is the radiant flux transmitted.

3 Satellite Imagery

3.1 Visible Radiance Images

Data has been operationally collected and processed since July 1983 by NASA and NOAA. [6] Since that time, visible (VIS) and infrared (IR) radiance images have been provided from all operational ISCCP weather satellites. These imaging principles are the same as the ones used in the MODIS instrument.

In order to properly interpret visible satellite imagery, knowledge of the surface characteristics in the scene being analyzed is required. [7] Water, land, snow, ice and various types of soil and vegetation all have signatures that can be used for identification.
The characteristics of the surface determines how much of the radiation reaches the satellite. Furthermore, elevation, latitude and climate information can be used to explain certain features on the image.
For example; cloud types found over water are often different from those found over land mass. The elevation changes over land can be much more sharp than those over water. Thermal characteristics of land and water produce varying background radiance properties over the diurnal cycle.
Moreover, the characteristics of water and land are different in each spectral channel and can also vary with the composition of the surfaces of these. A recognizable geographic location is required for the manual navigation and gridding of an image. For these purposes a land/water boundary is often used, as the features of these are often very clear on both visible and infrared imagery.
Generally, the land appears brighter than the water areas. Land areas can have an albedo of 0.10 to as high as 0.35 over sand, whereas a sea surface with a high sun angle has an albedo of 0.05-0.10, forests typically lie between 0.03 and 0.10, and thus are not as good for contrasting.
For ISCCP visible imagery, a radiance background image is constructed by comparing the image to other images of the same location(described further in section 3.1.3.1). MODIS uses the same method for imaging.

3.2 Infrared Images

According to the black body radiation law, all bodies with a temperature above 0K emit infrared radiation. The same principles as visible imagery apply for infrared imaging. A warm land surface next to a cold water surface creates another form of contrast, but as water temperatures change slowly compared to land temperatures, the contrast can be poor at some time of day and sharp at an other time of day. Since clouds would have a different temperature than the underlying surface of the Earth, a spacial (the pixel compared with other pixels in same area) and temporal (the pixel compared with pixels of the same location, at other times of day or other days) comparison can provide a map over the observed area. By having both a visible and a thermal image of the scene observed, the two methods provides a good overall picture of the surface and the clouds as missing or bad information from one of the data measurements, can be supplied by the other to obtain results which are closer to the reality of the scene. At night the infrared measurements are preferred, as the imagery from VIS imaging are not as clear.

3.3 Microwave Radiometry

For all matter with temperatures above 0K, thermally caused electromagnetic radiation occurs. This radiation is addressed with microwave radiometry [17]. With the help of sensitive receivers with multiple receiving channels, microwave sensors are used to derive the characteristic emission spectrum of matter in the atmosphere or extraterrestrial objects. As the atmosphere appears semi-transparent in the microwave spectral region, this brings the possibility to study water vapor and the cloudy atmosphere with microwave radiometry. As the spectral range is from 1 to 300 GHz, microwaves can be used to provide additional information to the visible and infrared spectral range [16]. Due to the impact of reflected sky radiation of cosmic origin, significant contrasts can be observed between reflective and absorbing materials. For Earth observations, an approximate observable brightness temperature measure range is 3K to over 300K. The brightness temperature can be used as a daytime and almost weather independent indicator for many physical phenomena in the microwave region. This makes it an interesting tool for analysis of especially climatology. Microwave radiometry is the method used in the SSMI system.

3.4 Brightness temperature

By measuring the noise power in the microwave region, the physical temperature of objects can be calculated. This is done instead of measuring the voltage or the current in the microwave region. A microwave radiometer has the same features as a low-noise highly amplifying power meter
The temperature of a resistor can be calculated as the noise power is proportional to the physical temperature of this, following Nyquist's law. The physical temperature of thermally radiating matter can also be calculated, following Rayleigh-Jeans law, considering the resistor as a lossless receiving antenna in thermal equilibrium The received radiation is not necessarily generated in the observed matter due to emission, as is the case for a blackbody, but it can be composed of additional partly reflected and partly translucent radiation generated in the same fashion elsewhere. The observed temperature is apparent and is not the same as the physical temperature T_0, it is called the brightness temperature T_B and the composition is illustrated on figure 2.

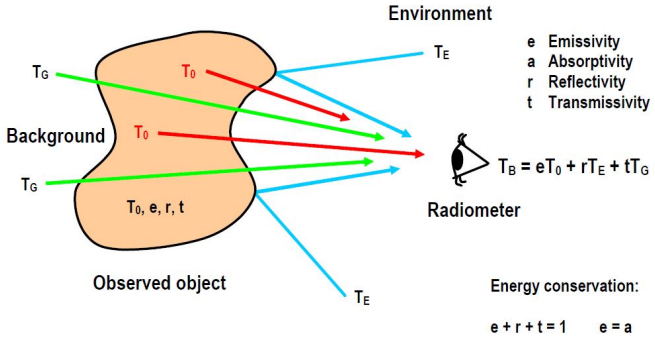


Figure 2: Simplified illustration of T_B. The environment radiates the object, resulting in thermal electromagnetic radiation, which is captured by the sensor. Credit: WFMN07. M. Peichl, S. Dill, M. Jirousek and H. Süß

Chemical and physical properties of matter, including the permittivity, surface roughness and shape of the object and polarization and viewing angle are included in emissivity e, reflectivity r and transmissivity t.
As the penetration depth in the millimetre-wave (MMW) region is very low, and the transmittivity is close to zero, the difference in emissivity in matter allows the sensor to see differences among various matter and can separate these from the background. An image of the surface and the matter above the surface is created based on microwave radiometry, which is the big difference between SSMI and the satellites operating with VIS and IR imaging.

3.5 Sky Brightness Temperature

Significant contrasts can be observed when microwave radiometry is applied on Earth studies. These contrasts are a consequence of the reflected sky radiation of cosmic origin. The space outside the terrestrial atmosphere acts like a blackbody radiator with a physical temperature of about 3K.
The cold appearance of the space can be observed more or less strongly dependant on the reflective properties of an object and the frequency dependant absorption and self-emission of the atmosphere.

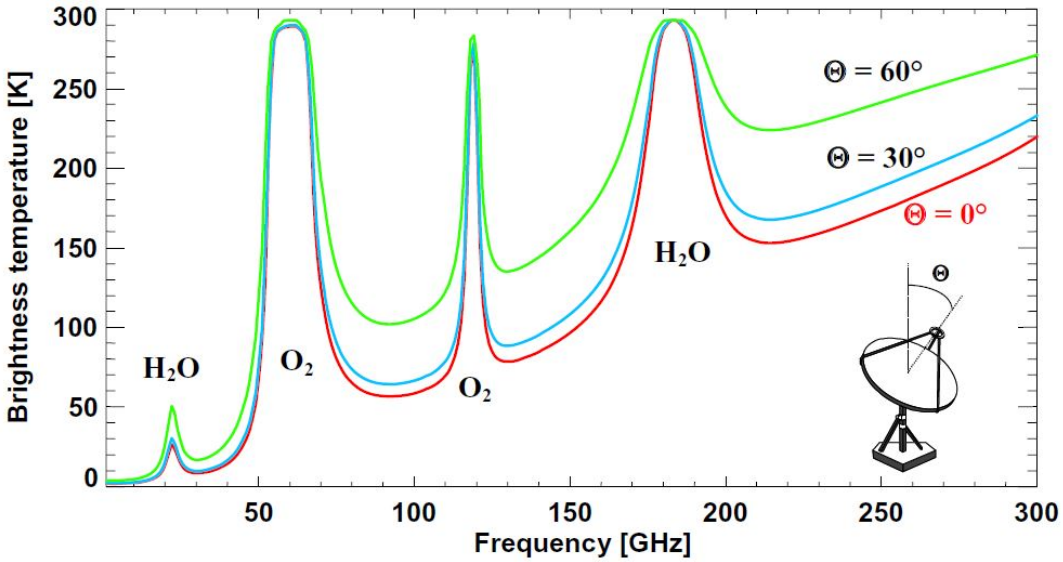


Figure 3: Main MMW windows for sufficient atmospheric penetration and low sky brightness temperatures. The three different colors represent the readings at three different viewing angles. It can be seen that the windows are wider at higher viewing angles. Credit: WFMN07. [17]

Figure 3 shows the main MMW windows for sufficient atmospheric penetration.
Disturbances from scattering of energy by the atmospheric molecules takes place along with interference from frequency selective absorption in these windows, this is a problem for MMW imaging, but these areas are still preferred in atmospheric research.

The brightness temperature increases with the increasing observation angle due to the longer path of the cosmic radiation through the atmosphere and it can be seen that the bandwidth of the windows increase as the viewing angle increases.

3.5.1 Antenna Temperature

As seen on figure 3 the observable brightness temperature of a scene depends on the direction of the observation. It can be measured by a microwave radiometer, which is an antenna connected to a low-noise receiver. The actual measured brightness temperature is given by the antenna temperature $T_A(\Theta', \Phi')$, which is a convolution of the true brightness temperature and the antenna power pattern $P(\Theta, \Phi)$, which can be a real or a synthetic beam (passive/active). This convolution represents the movement of the antenna beam in spherical coordinates for which $d\Omega = Sin\Theta d\Theta d\Phi$ is valid.

$$T_a(\Theta', \Phi') = \frac{1}{\Omega_A} \int \int T_B(\Theta, \Phi) P(\Theta - \Theta', \Phi - \Phi') d\Omega \tag{2}$$

The antenna beam degrades the true brightness temperature as it has a finite beam width. The receiver adds its own noise to the observable noise power of the scene, therefore the radiometer can only measure with a finite sensitivity.

3.5.2 Spatial Resolution and Sensitivity

The ratio of the wavelength λ and the aperture diameter D of the imaging antenna, determines the angular resolution $\Delta\Theta$. The constant k_A depends on the field distribution in the aperture plane of the antenna.

$$\Delta\Theta = k_A \frac{\lambda}{D} \tag{3}$$

This means that the antenna size has to be improved in order to improve the spatial resolution of the microwave radiometer for a given wavelength.

The antenna temperature T_A, the receiver noise temperature T_E the radiometer bandwidth Δf and the integration time τ determines the sensitivity or temperature resolution ΔT.

$$\Delta T = k_R \frac{T_A + T_E}{\sqrt{\Delta f \tau}} \tag{4}$$

The constant k_R depends on the receiver type. For a high sensitivity, the antenna and the receiver noise have to be low and the bandwidth and the integration time have to be large.

4 Satellite Systems

4.1 ISCCP

One of the data sets we are working with is from the International Satellite Cloud Climatology Project(ISCCP).
ISCCP was established in 1982 as part of the World Climate Research Programme(WCRP) [5] to collect and analyze satellite radiance measurements to get an overview of the global distribution of clouds and their variations. Data collection began 1st of July 1983 and ended 30th of June 2010. The system is at network of satellites from different countries, and is as the name states, an international collaboration. Data are still collected today, but is not published yet. ISCCP is a multi-satellite system, consisting of different satellites from different countries. The data sets and products are being used to improve modeling and understanding of the role of clouds in our climate system. The primary focus is on the effects of clouds on the radiation balance, which is done by collection of radiance data, used to produce data which contain information of cloud properties and general physical properties of the atmosphere.

4.1.1 ISCCP Satellite Coverage

For the latest 18 years the Earth has been covered on all longitudes by ISCCP satellites. [6] Satelittes which are still operating are named NOAA-17, NOAA-18, MTSAT-2, FY-2E, GOES-13, GOES-15, METEOSAT-7, METEOSAT-9 with METOP-1, and METEOSAT-8 in reserve. Figure 4 shows the coverage of the satellites, for later comparisons and discussion of general adequate coverage of the Earth. As mentioned ISCCP is a multi-satellite system and consists of various geostationary satellites and polar-orbiters.
As geostationary satellites have the same view of the Earth at all times, this means that the viewing angle on some locations are higher than others. This leads to a longer path from the satellite to the surface of the Earth, this causes higher readings at a high viewing angle (as explained with radiometric viewing angle in section 3.5).
Polar orbiters, orbit the poles to cover all longitudes of these. The polar orbiters are used for calibration of the geostationary satellites through relative radiometric normalization and as providers of information in the polar areas.

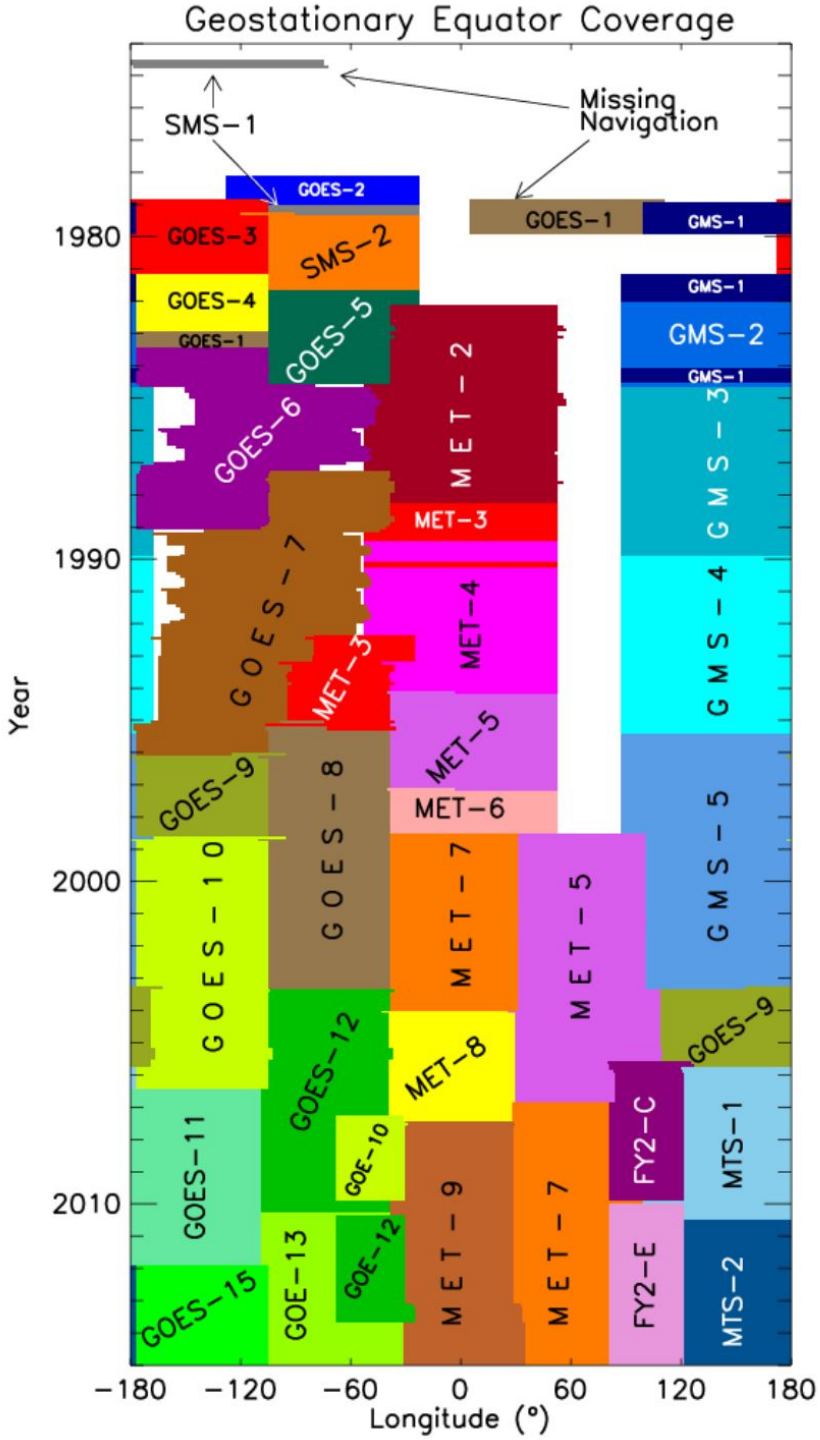


Figure 4: ISCCP satellite coverage of the Earth. The figure shows the geostationary satellites that have been orbiting and which longitudes they have been covering. The Earth has been covered at all longitudes the past 18 years. Credit: NOAA, ISCCP project site 20/6-2017

The coverage of ISCCP satellites is good compared to MODIS and SSMI satellites. Figure 5 explains the setup with geostationary satellites and polar orbiters for calibration purposes.

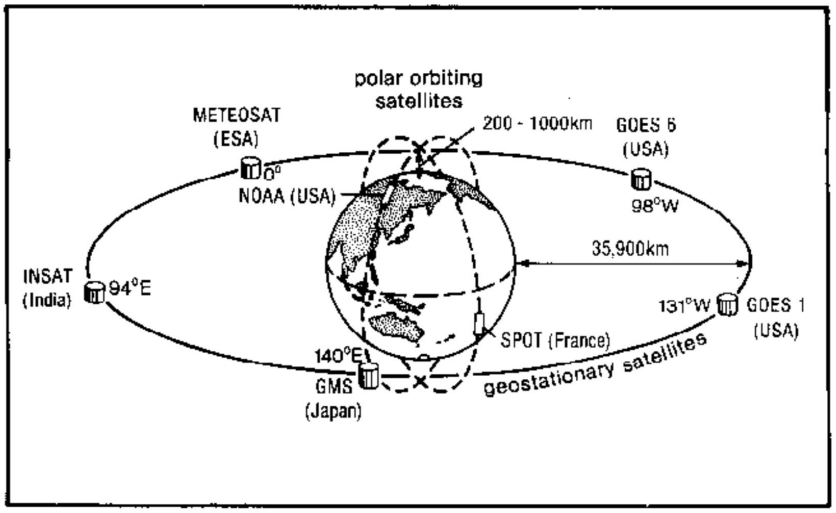


Figure 5: Figure illustrating the orbital path of the geostationary satellites and the polar orbiting satellites of ISCCP. Various geostationary satellites are used to obtain data which are calibrated with polar orbiters. Credit: ISCCP project site 20/6-2017

4.1.2 Data stages

To differentiate between raw image radiance data, reduced volume radiance data and derived cloud data, three stages have been defined for the ISCCP infrared (IR) radiance and visible (VIS) imagery radiance data. [12]

4.1.2.1 Stage A - Full resolution digital satellite image data

Stage A data are full resolution narrow band radiometer image data taken for every 3 hours. Included in this are the satellite orbital navigation (earth location) and calibration parameters. Within this data category are two sub-categories:

- AC-Data - Data for satellite intercalibration
  The AC-Data is full resolution data for both visible and IR imagery, with the navigation and calibration parameters appended. These are for geographical areas of about 2000 km x 2000 km.
- AS-Data - Special full resolution data sets
  The AS-Data is full resolution image data, or 20 pre-defined geographical areas of about 250 km x 250 km for every 3 hours.

4.1.2.2 Stage B - Reduced volume digital satellite image data

Stage B data is radiometric image data which is reduced by averaging and sampling. This stage is applied to reduce the data volume to a more manageable size.

- B1-Data - Nominal 10 km spacing
  B1-Data are compressed versions of stage A data produced by averaging the visible radiances to the resolution of the infrared radiances and sampling the matched resolution data to a nominal 10 km spacing.
- B2-Data - Nominal 30 km spacing
  B2-Data is further compressed versions of the stage A data, produced as B1 with sampling of the matched resolution pixels to a nominal 30 km spacing.
- B3-Data - Globally merged calibrated radiances
  The B3-data are B2-data which have been normalized to the polar orbiter radiometer responses. These are placed in a standard format with navigation and calibration information appended.
- BC-Data - Normalization coefficients
  BC-Data are coefficients calculated from AC-data used to convert individual radiometric data into SI-units of radiance.
- BS-Data - Special reduced volume data sets
  BS-data are B1-data for a number of special geographical regions of about 2000 km x 2000 km obtained every 3 hours for 4 days each month.

4.1.2.3 Stage C - Correlative and cloud data

The Stage C data are cloud parameters obtained from B3-data and correlative data using an experimental algorithm. For specification of the radiative properties of the surface and atmosphere, the correlative data (CD-data) are used. The cloud parameters are computed for an equal area grid with 250 km resolution.

4.1.3 Data handling

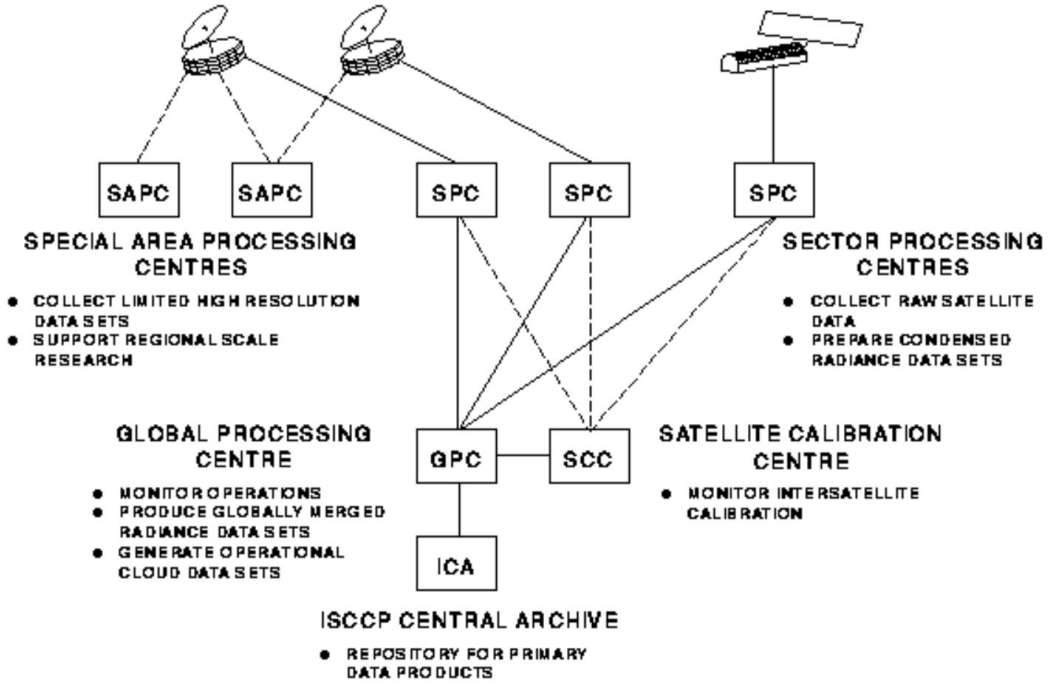


Figure 6: Overview of the data collection method, and the different operators of the data, for each data step there is a new handler of the processed data. Credit: ISCCP official site 20/6-2017

Data are collected from weather satellites operated by several nations and processed by government agencies, laboratories and universities. For each satellite a satellite processing center(SPC) collects the raw data and sends it to the global processing center(GPC). The correlative data center(CDC) delivers the other satellite and conventional weather data to the GPC. The satellite calibration center(SCC) normalizes the calibration of the geostationary satellites with respect to a polar orbiter satellite standard. All ISCCP data products are archived at the ISCCP central archive(ICA) and at NASA Langley Research Center(LARC).

4.1.4 Cloud detection

The sets of data we are working with(D2 data) are made from B radiation data sets and are improvements of the C data sets.
The ISCCP cloud detection procedure is applied to each month of satellite data at eight times of day(hours divisible by 3) and consists of five steps [8]:

1. Space contrast test(applied to individual IR images).
2. Time contrast test(three consecutive IR images at constant diurnal phase).
3. Cumulation of space/time statistics(both IR and VIS images).
4. Construction of clear-sky composites for both IR and VIS(once every 5 days at each diurnal phase and location).
5. Radiance threshold(both IR and VIS images).

By comparison, all pixels with a lower IR radiance than the the one with the highest IR radiance in an area is classified as cloudy. This is the first step in the classification process and is a spatial contrast test. The pixel with the highest IR radiance is therefore not classified as either clear or cloudy.
To avoid confusion between diurnal variations of the suface temperature, the second step is perfomed for each time of day. Pixels with sharply lower IR radiances than the day before or after at the same location is classified as cloudy. The pixels with little variation are classified as clear. By combining the results of the first to steps, it is possible to label pixels as clear, if little or no variation is shown in both of these.
For statistics, the variations of the IR and VIS radiances over larger spacial and temporal domains are used. This is the third step and is used to describe data on a larger scale .
Along with the results from the spacial and temporal tests in the first two steps, the statistics from step three are used to estimate the values of clear IR and VIS radiances for each location and diurnal phase for every 5 days.
The first four steps are used to derive clear values for both IR and VIS radiances. The fifth step now compares all of the original IR and VIS radiances to the deduced clear-sky values (lower IR or higher VIS). If the observations differs from the values by more than an estimated uncertainty, the pixels are classified as cloudy. The possibility of having a marginally cloudy pixel is also present. Marginally cloudy pixels have radiance values close to the values dividing clear pixels from cloudy. All other pixels than cloudy or marginally clody pixels are classified as clear.
The following sections will explain in detail how the radiative analysis is performed and how time statistics for the data are created.

4.1.4.1 Radiative Analysis

A comparison of the measured radiances with radiative transfer model calculations, which include the effects of the atmosphere, clouds and surface is performed when it is classified whether the pixel is clear or cloudy.
Availability of correlative data sets restricts the number of parameters which can be determined from the observations. The satellites only measure radiances at 2 wavelengths, this restricts the number of parameter products additionally. In spite of these restrictions, 130 parameters of physical cloud properties, and attributes of the atmosphere is represented in the data sets. An overview of these variables can be seen in section 4.1.5.

By comparision and use of the correlative data, the analysis strategy isolates the cloud effects and attributes all remaining radiance variation changes in two cloud properties, other parameters are assigned climatological average values (and assigning -1000 as value for missing data).
At the wavelengths which VIS and IR operate at (0.6 and 11 microns) atmospheric effects are small and are dependant on ozone, water abundances, the temperature profile and aerosol optical thickness.
Correlative data is used to obtain information regarding the atmosphere. Ozone, water abundances and temperature profile data as mentioned before are among the additional information needed to correctly interpret data. As aerosol property information is sparse, their effects are neglected; however, since most aerosol occurs near the surface, and has an effect on the upwelling radiation from there, the use of surface properties obtained from clear radiance values with the same radiative model, incorporates the primary aerosol effects into these surface properties.
That is, except after a large volcanic eruption, since large dust storms are generally detected as clouds, and their properties retrieved as those of a water cloud.
The radiance values obtained from the surface are the visible reflectance and the brightness temperature
The amount of sunlight reflected by the surface at 0.6 microns and at a particular sun and satellite geometry, is represented in the data for visible reflectance.
The anisotropical reflection of sunlight for different surfaces, results in a variation of visible reflectance, which depends on satellite position, time of day and season.
The temperature values obtained for the IR radiances at the surface are correctly described, but does not directly translate into the actual physical temperature, as the IR emissivities on most surfaces are slightly less than one.
As land and vegetated area emission information is inadequate, all temperatures are retrieved assuming an emissivity of one.

Clouds are represented in the model as a thin average sized (10 microns) and averagely distributed layer of water droplets, which are uniformly covering the image pixel. All variations of cloudy radiances are attributed to changes in visible optical thickness (0.6 microns) and a temperature.
The amount and angular distribution of sunlight reflected by the cloud layer is determined by the optical thickness parameter, and the temperature is a brightness temperature, which is interpreted to represent the physical temperature at the top of an opaque cloud layer.
As vision is impaired at night, only IR radiances are measured, variations in these are associated with the cloud top brightness temperature. A consequence of this is that no optical thickness is reported. An IR optical thickness is related to the cloud optical thickness at day, to correct for emissivities of clouds which are less than one. Both of these values (corrected cloud top temperature and original opaque cloud value) is reported.

Particle size, phase(water or ice) and size distributions are other cloud properties that may vary. Multiple layering caused by clouds on top of each other or small scale brokenness caused by vertical and horizontal inhomogeneities can affect the radiances (see section 7).
The retrieved optical thickness and the temperature parameters in the ISCCP data are corrected by comparison. The radiative model radiances which are calculated with and without the variation of these additional parameters, are used to produce a transform from one model to the other, that is, if the role of these other cloud properties can be formulated correctly and their value measured.

4.1.4.2 Statistics

For each image pixel, the radiative analysis and the basic detection is performed. However the overwhelming cloud climatology data need to be reduced to a manageable volume. By summarizing the pixel values in the stage C1 data (3-hourly) this is done.
Projection of the pixel data into a standard map grid with and resolution about 280km, and averaging of cloud and surface property values as well as their standard deviations. An equal area map grid is used to maintain approximately statistical significance.
Furthermore the certain distributions of cloud optical thickness and top pressures are reported.
All of these statistics are collected separately for each satellite.
The final C1 data sets are assembled by merging the results from all of the satellites.
The merging is done by selection to preserve the statistical character and uniformity. Two criteria is the base for settling which satellite should be used. 1) a strong preference for time records produced from a single satellite. 2) limiting the zenith angle.
At low latitudes nearby geostationary satellites generally report the data, these are replaced with a geostationary satellite which covers part of the same area if data is missing or a polar-orbiter. In the polar regions polar-orbiters provide the coverage.

4.1.5 Data Content

The ISCCP data sets consist of 130 variables, which are listed in table 1.

Table 1: Overview of the variables in the ISCCP D2 data sets

Variable number	Description
1-7	Box identification
8-19	Cloud amount
20-22	Mean cloud top pressure (PC)
23-25	Mean cloud top temperature (TC)
26-28	Rate of liquid water precipitation
29-31	Mean cloud water path
32-40	IR cloud types
41-70	LOW cloud types
71-100	MIDDLE cloud types
101-115	HIGH cloud types
116-117	Mean surface temperature
118	Mean surface reflectance
119	Mean snow/ice amount
120-130	TOVS atmospheric information

For the analysis we are interested in the box information (1-7) for location purposes. The information about date/month is specified by the filename. The cloud amount (8-19) is used for general comparison with cloud fractions and cloud liquid water from MODIS and SSMI. LOW, MIDDLE and HIGH-cloud types (41-115) are used for detailed analysis of cloud behaviour in specific regions and to check variability in the different cloud regions (various heights).

4.2 MODIS

MODIS (Moderate Resolution Imaging Spectroradiometer) is an instrument carried onboard the Terra (EOS AM-1) and Aqua (EOS PM-1) satelittes. The satellites are in a sun-synchronous, near-polar, circular orbit.
The main purpose of the data collection from MODIS is to improve the understanding of global dynamics, which means data for studies of processes occuring on land, in the oceans and in the lower atmosphere. Data is acquired for 36 spectral bands. The MODIS data is well-used in modelling of interactive Earth system models, to help making predictions of global climate changes. We are interested in MODIS data as the data includes a cloud fraction variable, which is comparable to both ISCCP and SSMI.
The system consists of four refractive objective assemblies, these cover a spectral range of 0.4 to 14.4 microns, which means that MODIS does not only capture images in VIS and IR, but also NIR (near infrared), SWIR (short-wave infrared), MWIR (mid-wave infrared) and LWIR (long-wave infrared). This leads to a wide coverage in the IR domain.

4.2.1 MODIS Satellite Coverage

Terra MODIS and Aqua MODIS circles the whole surface of the Earth in one to two days. The coverage is not as substantial as ISCCP or MODIS, which does add uncertainty to the measurements. Terra and Aqua images the same area of the Earth, three hours apart from each other.

4.2.1.1 Terra

Terra has a syn-synchronous, near-polar circular orbit, wwith crosses equator at 10:30 AM local time. Terra has a descending node, which means it crosses equator from north to south. Figure 7 shows the orbital path of Terra.

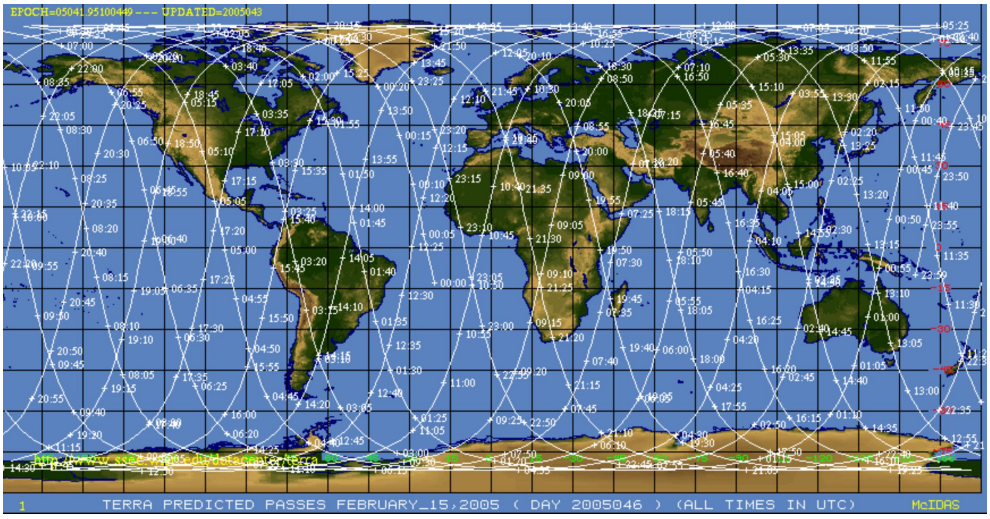


Figure 7: Figure illustrating the orbital path of the Terra MODIS satellite. The satellite crosses equator at 10:30 AM local time, from north to south. Credit: NASA, National Snow Ice Data Center 20/6-2017

4.2.1.2 Aqua

Aqua has a syn-synchronous, near-polar circular orbit, which crosses equator at 1:30 PM local time. Terra has a ascending node, which means it crosses equator from south to north. Figure 8 shows the orbital path of Aqua.

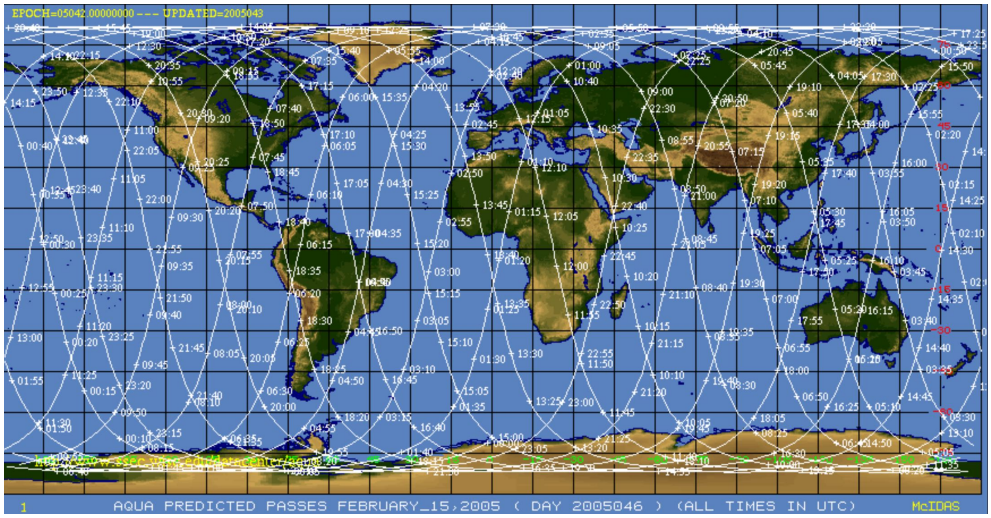


Figure 8: Figure illustrating the orbital path of the Aqua MODIS satellite. The satellite crosses equator at 1:30 PM local time, from south to north Credit: NASA, National Snow Ice Data Center 20/6-2017

4.2.2 Data Content

As the MODIS intruments covers a large spectral range, various measurements are made. Level 1 measurements include "raw" measurements. These data contain raw radiances, calibrated radiances and geolocational information. MODIS data also include land information including information about surface reflection, vegetation areas and albedo of the surface. The data we are interested in is the atmospheric data, from these sets it is possible to obatin information about cloud properties, cloud mask and atmospheric profiles. For our analysis we have used the variable cloud fraction (0.0-1.0) which indicates how many of the pixels in an image are covered with clouds. The data set does not include other variables, as it was possible to extract only the cloud fraction variable directly from the data as a ".csv" file. The cloud fraction can be directly compared with ISCCP's cloud amount, which is given in a percentage (CA).

4.3 SSMI

The Special Sensor Microwave Imager (SSMI) and the Special Sensor Microwave Imager Sounder (SSMIS) are satellite passive microwave radiometers.[14] Which will be used in the analysis to be compared with both ISCCP and MODIS. The instruments has been carried onboard Defense Meteorological Satellite Program (DMSP) satellites since 1987.
SSMI operates at four frequencies: 19.35, 22.235, 37 and 85.5 GHz [15]. These channels make it possible to retrieve three important geophysical parameters over the ocean: near-surface wind speed W, columnar water vapor V and columnar cloud liquid L. The latter is why we are interested in using SSMI for our analysis. When rain is not present, the relationship between the ocean brightness temperature T_B measured by SSMI and W, V and L is relatively simple given by the radiative transfer equation (RTE). The RTE can be approximated by a closed form expression, which is called the T_B model function. The retrieval of the parameters we are looking for is accomplished by varying W, V and L until the model matches the SSMI observations.
The brightness temperature function (T_B) is derived based on well-accepted microwave theory and is denoted by:

$$T_B = f(T_s, U_*, V, L, T_a) \tag{5}$$

The arguments in the function are; Sea-surface temperature T_s (K), sea-surface friction velocity U_* (cm/s), columnar water vapor content V (g/cm^2), columnar liquid water content L (g/cm^2) and surface air temperature T_a (K)

For the derivation of the equation for the brightness temperature upwelling from the sea surface through the intervening atmosphere, radiative emission and scattering are considered alongside with radiative absorption and emission of the atmosphere. This does not include the radiative scattering in the atmosphere caused by raindrops.

4.3.1 SSMI Satellite Coverage

SSMI and SSMIS are carried onboard near polar orbiting satellites. 10 instruments have been deployed in total since 1987. The first 6 deployments are SSMI instruments, these are named F08, F10, F11, F13, F14 and F15. F16, F17, F18 and F19 are SSMIS instruments. of these 10 instruments, the latest 5 are still orbiting. Figure 9 illustrates the coverage of one SSMI satellite for one day, as seen on the figure, measurements are only made over oceans, areas close to land or sea ice are not included in the measurements.

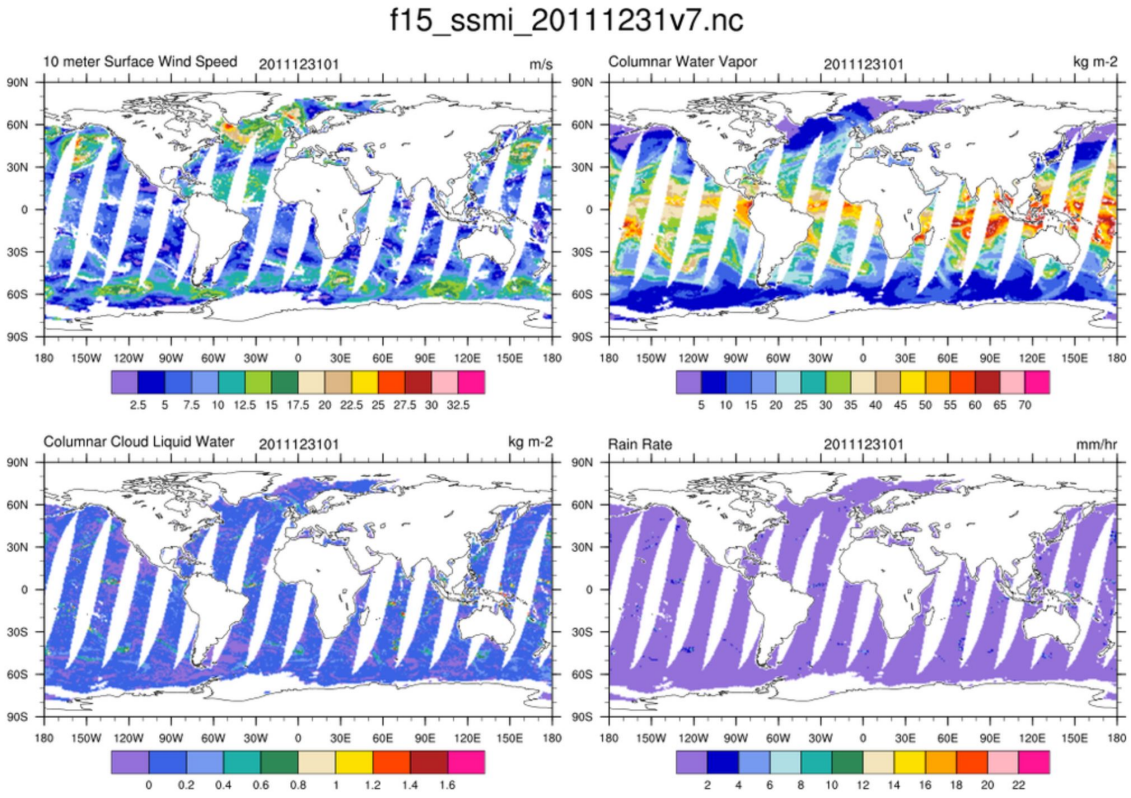


Figure 9: Figure showing the coverage of a SSMI satellite for one day, these data are used for compiling the monthly mean data used in the analysis. Data from more satellites and orbits provide a full coverage. One orbit from the F15 satellite is shown. Credit: NCAR (https://www.ncl.ucar.edu/Applications) 19/6-2017

Figure 10 shows the averaged data made from daily data of the SSMI satellites. These data products are used for the analysis when comparing cloud coverage of the Earth.

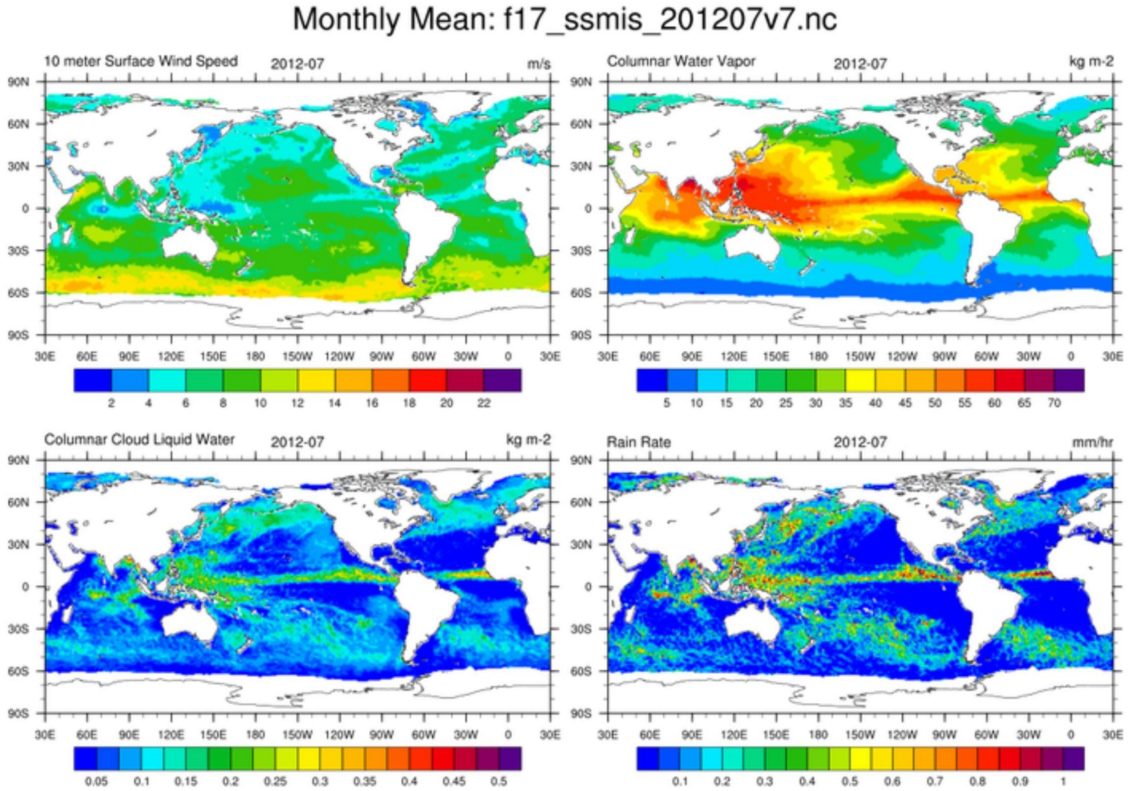


Figure 10: Figure showing mean value for a month for SSMI data. These data are compiled by averaging of the daily data. Here is monthly mean maps presented for the F17 satellite, for the month of July in 2012. Credit: NCAR (https://www.ncl.ucar.edu/Applications) 19/6-2017

4.3.2 Data Processing Approach

The process of calculating the brightness temperature (T_B) from raw radiometer counts is very complex and is a process with various steps which require accounting for a number of accurately characterized effects. [19] Radiometer non-linearity, emission from the primary antenna, imperfections in the calibration targets and antenna pattern adjustments. RSS T_B are calibrated consistently and thus makes sure that the measurements for all antennas can be used to construct a multi-decadal time series.
The SSMI radiative transfer model (RTM) of the ocean and the intervening atmosphere in the absence of rain is very accurate, it can predict T_B in the top of the atmosphere to a very high degree of accuracy. The RTM has been continually developed and refined for 30 years and works well in the microwave spectrum (1-100 GHz) for ocean observations. For the absolute calibration reference, a rain-free ocean is used.
The calibration steps include a geolocational analysis, along-scan correction, absolute calibration, hot load correction and an antenna emissivity correction, an attitude correction

would have been optimal for SSMI, but is not present. For each measurement the exact location of the readings has to be known for subsequent comparisons. By using ascending and descending values and comparing the position of small islands with each other, the SSMI calibration makes sure that the geolocation is exact.
By comparison of antenna temperatures with simulated radiative transfer model values the rest of the calibration steps are performed.
The along scan corrections makes sure that there are no obstructions of the mirror. As it rotates the satellite itself or parts of the cold mirror could obstruct the view. These obstructions are accounted for in this step.
The absolute calibration includes an antenna pattern correction. Before the launch spill over and cross-polarization values are determined. These values are adjusted after launch so that the antenna temperatures which are measured, matches the temperatures simulated in the radiative transfer model.
The early SSMI instruments does not need hot load thermal gradient correction, which means that this is only done for the SSMIS instruments. The T_B is determined from microwave radiometer counts by using two known temperatures to obtain a value which is inferred to being the surface temperature of the Earth scene. As the temperature of the cold space is known (2.7K) this is used as a reference temperature, a mirror reflecting cold space is used, combined with the temperature of a hot absorber measured by several thermistors, this information is used in the antenna feedhorns for each scan. The temperature of the Earth scenes are the determined by fitting a slope to the measurements which are known from the cold space mirror and the hot absorber, this is done assuming that the response is linear. Variations in the radiometer gain and noise temperatures from external sources are conpensated for by this 2-point calibration system. The simple calibration methods has few difficulties. The cold mirror is affected by the moon, which means that when moonlight is present and reflected in the cold mirror, the temperature of the mirror will change. When the moonlight is present, the moon-affected values are removed. The hot absorber is more probematic as large thermal gradients develop in the hot load due to solar heating, this makes it difficult to determine an average effective temperature of the hot absorber. The thermistors measure the changes in the temperature and these might vary with up to 15K, this brings a uncertainty to the measurements.
The main reflector is assumed to be perfect, with an emissivity of 0.0, but in some cases the reflector has an emissivity. At some altitudes degradation of the primary antenna leads to oxidization of the coating of the primary antenna. The measured radiation therefore consists of both readings from the Earth scene and antenna emissions. The calibration procedure is used to deduce the emissivity of the antenna of around 3.5%. For the SSMI systems the emissivity appers to increase as a function of frequency, and the amount of disturbance lies between 0.5 and 3.5%
For correcting of spacecraft pointing errors, attitude adjustment is required. Different methods are used to determine the spacecraft pointing, but the preferred method is a star tracker. This correction is not made for SSMI, but for other systems that look a lot a like, for SSMI data it is assumed that the pointing is correct.

4.3.3 Data Content

The SSMI data sets contain various information, table 2 shows these.

Table 2: Contents of the data included in the SSMI data sets

Acronym	Product name	Valid data range
TIME	Time	0 to 1440 0.0 to 24.0
WSPD_MF	10-m surface wind speed	0 to 50.0 m/s
VAPOR	Total gaseous water contained in a vertical column of the atmosphere	0 to 75.0 mm
CLOUD	Total cloud liquid water contained in a vertical column of the atmosphere	-0.05 to 2.45 mm
RAIN	Rate of liquid water precipitation	0 to 25.0 mm/hr

The TIME data variable is the time given in minutes since midnight GMT and in fractional hour of day GMT, and is one of the most important parameters for our analysis. As we are using averaged data, the exact time stamp for each data point is not used, but a mean for the month is considered.
WSPD_MF is a near surface wind speed (10 meter height), possible reasons for missing data in the wind speed measurements are sun glint, rain, ocean reflected radio frequency interference (RFI) or being near sea ice or land ($\sim$50 km).
The VAPOR variable tells us about water contained in the atmosphere. Heavy rain interferes strongly with these measurements and will therefore cause incorrect data and leads to missing data points. Measurements close to land are also removed from the data ($\sim$25 km).
CLOUD is the variable that we are interested in. By measuring the liquid water path (LWP) the total cloud liquid water is calculated in a vertical column of the atmosphere. These measurements are made over oceans, missing data is a result of measurements made near land ($\sim$ 25 km). The data is scaled, which means that it is multiplied with 10 in our analysis.
RAIN is the data measured when heavy rain is present, as this causes heavy disturbances in the VAPOR variable. RAIN describes the rate of precipitation per hour.

4.4 Discussion of Coverage

The main reason for using satellite based systems for cloud cover observations, is that these can provide a larger spatial resolution of data. Surface based instruments are also used, but they only provide coverage in areas which are populated, although these measurements are good, they can not provide information about clouds above the oceans. Aircraft based instruments can provide a large spatial resolution, but a system of airplanes would take up more resources and more man power than an orbiting satellite to obtain large-resolution images.

A satellite system in polar orbit can provide a full global coverage. The time it takes for a satellite to cover the Earth compared to other systems is also a factor in why this is the most effective way to gather data. For a single polar-orbiter a orbit takes around 12 hours. [20] This results in a quite low time resolution but these measurements can be supplemented by geostationary orbiters. They deliver results in a very high time resolution compared to the polar orbiters, and can provide data from the area they are covering for every 15-30 minutes. The different time resolutions mean that descriptions of cloud variations in the different orbits does not have the same spatial scales for meaningful data. As the polar orbiter has a low time

resolution it would only make sense to describe cloud coverage on a large scale (>1000 km) as it would not be possible to go into detail with a coverage of the entire area for every 12 hours. The geostationary orbits on the other hand, makes it possible to go into detail. They provide full coverage of the location for every 30 minutes, this brings a lot of information concerning the entire area, and gives the possibility for small scales to be investigated (∼5km).

4.4.1 ISCCP, MODIS and SSMI Resolutions

As mentioned in previous section, different spatial resolutions are optimal for different time resolutions. This section highlights the differences in the resolutions in the systems.

4.4.1.1 ISCCP

The resolution of the data provided from ISSCP is 30 km per 3 hours, this is the format of the original radiance data. The 2.5 x 2.5 degree data that are used in our analysis, covers a 280 km area per pixel at equator. As ISCCP has both geostationary and polar orbiting satellites, the spacial resolution and the time resolution are reasonable compared to each other. More geostationary satellites would provide a coverage which would be good when looking at the resolution, five satellites for each hemisphere would be optimal minimum coverage [20]. The calibration of data from geostationary against polar orbiting satellites gives reason to believe that the data obtained from ISCCP is more meaningful than the data obtained from MODIS in terms of coverage.

4.4.1.2 MODIS

MODIS has different spatial resolutions for the different channels the spatial resolution for the different bands are: 250 m (bands 1-2), 500 m (bands 3-7), 1000 m (bands 8-36). The data used in this thesis is from band 8-36. This gives us a spatial resolution of 1000 m, the data comes in 0.1 x 0.1 degree grids, which means a very high resolution. As MODIS are placed on 2 near polar orbiting satellites, this resolution seems quite high, the orbital period for MODIS is one to two days which is quite high for polar orbiters, and the spatial resolution should therefore be lower for MODIS to obtain more meaningful data (larger spatial area). The coverage is global, but inserting more polar orbiters with the same time displacement compared to each other (∼3 hours) would justify a high spatial resolution.

4.4.1.3 SSMI

SSMI systems have different spatial resolutions for each channel, 12.5 km resolution for 19.3 and 22.2 GHz, 7.5 km for 37.0 GHz and 2.5 km for 85.5 GHz channel. The resolution is also quite high for the SSMI data, but the SSMI system consist of 5 polar orbiters. This provides better coverage than MODIS and the orbital period is shorter. which could make SSMI a better candidate than MODIS for measuring the liquid water path. Coverage over oceans are good in SSMI data sets, but comparisons of the systems can only be made in these areas as SSMI only has valid data over above sea. The latter is why we still need other systems to obtain data over land areas. For the comparisons, areas above oceans are used when SSMI data are included.

5 Introduction to the data from the satellites

In order to have sufficient information to work with, data from 3 different missions have been acquired: ISCCP, SSMI and MODIS. From these a total of 9 data sets will be processed. These are all monthly data with respective data points per longitude/latitude (XXXX×XXXX), giving a resolution of (X.X°×X.X°) per data point:

- **ISCCP D2** (144×72) → (2.5°×2.5°)
- **SSMI** (144×72) → (2.5°×2.5°)
- **SSMI f08, f10, f11, f13, f14 and f15** (1440×720) → (0.25°×0.25°)
- **MODIS** (3600×1800) → (0.1°×0.1°)

All data processing is done in MATLAB.

5.1 Introduction to the data

ISCCP D2

Originally the ISCCP D2 data, acquired through the Global Hydrology Resource Center (GHRC), had to be run through a Fortran program in order to generate data in a format, which MATLAB was able to read. Luckily, Jacob Svensmark was able to provide these files as a `.txt`-format, making it very easy to extract the desired data. From the original data, 4 subsets were picked: Low cloud amount (CA), middle CA, high CA and total CA. These are arranged in a 144×72 matrix, where every element represents an area of 2.5°×2.5°. The cloud amount is measured as percentage of the area that is covered in the respective cloud type.

SSMI

The first of the SSMI data set is given in the `.NC`-format (NetCDF - Network Common Data Form), and is, when loaded into MATLAB, similarly to ISCCP D2 arranged as a 144×72 matrix representing same areas.

f08-, f10-, f11-, f13-, f14- and f15-satellites

These 6 satellites have all been operating in different time periods (some of them overlapping each other) as part of the SSMI mission. The data set used are of a much higher resolution than the formerly mentioned SSMI data. They in fact offer a 1440°×720° resolution, i.e. the matrix we import, is 100 times larger, and so the areas represented by the matrix elements are of the size 0.25°×0.25°, which of course means much more data per set. In order to load the data, 3 functions and 2 scripts were provided along with the data.

MODIS

At last we have the MODIS data, given in the `.CSV`-format. These are heavy sets of data in that they have a resolution of 3600°×1800°, making the represented areas 0.1°×0.1°. Since data of this size is tough for the computer to handle, a one-time transformation from 0.1°×0.1° to 1°×1° is performed during the import of the files. This is done by taking all the 360×180 areas of size 10×10 data points, and calculating the weighted mean hereof giving a single value representing the whole area. This means taking longitude and latitude into

account, which will be explained further in the section about the generated time series. The data represents cloud fractions, which just like ISCCP D2 tells of the cloud cover in the respective area, with only difference being ISCCP given as a percentage(0-100%) and MODIS as fraction(0.0-1.0).

As mentioned, the 9 sets of data were all recorded over different time periods, generating a new set every month in their respective operating periods. This results in the following amount of monthly data:

Table 3: Satellite operating period

Satellite data	Start of recording	End of recording	Months of data
ISCCP D2	July 1983	December 2009	318
SSMI	August 1987	August 2012	282
f08	July 1987	December 1991	54
f10	December 1990	November 1997	84
f11	December 1991	May 2000	102
f13	May 1995	November 2009	175
f14	May 1997	August 2008	136
f15	December 1999	December 2009	121
MODIS	February 2000	March 2017	206

We see that MODIS has very recent data, which is merely the latest recorded data at the time of extraction. The satellite is still operating.

5.2 Representation of the data

The data for each month is packed as a $i \times j$ matrix. This gives us the longitude and latitude vectors of respective lengths i and j. These are given:

$$Longitude_{res} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{i-1} \\ x_i \end{bmatrix} \quad \text{and} \quad Latitude_{res} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{j-1} \\ y_j \end{bmatrix} \tag{6}$$

Here, the individual x- and y-values represents a data point from the set. The interval between each value is given as the length of the sides of the area the respective data represents, hence why it is noted according to the resolution (res). The values in the 2 vectors are the latitudinal and longitudinal coordinates given in degrees. We wish to pack the data from every satellite into a single $n \times m$ array. This is done by sorting the $i \times j$ matrix, so for every month, the data will be represented in a single row vector of the length $i \times j = m$, and the final set of data will contain n of the monthly rows of data. In order to keep track of the individual values, 2 vectors (longitude and latitude), also of length m, are created with a system to pair up one value of longitude and one of latitude with their respective value from the data set. Each of the i longitude values will first be given for the first latitude value, then for the second and so forth till the j'th value like so:

Table 4: Pairwise arrangement of longitude and latitude vectors

$Long_{sat}$	x_1	x_2	$\cdots$	x_{i-1}	x_i	x_1	x_2	$\cdots$	x_{i-1}	x_i	x_1	x_2	$\cdots$	x_{i-1}	x_i
$Lati_{sat}$	y_1	y_1	$\cdots$	y_1	y_1	y_2	y_2	$\cdots$	y_{j-1}	y_{j-1}	y_j	y_j	$\cdots$	y_j	y_j

These vectors are created to ease the plotting process, as well as providing an easy overlook when comparing individual data points from month to month . In so, we will have arranged our data as such:

$$data_{sat} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,m} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,m} \end{bmatrix} \tag{7}$$

6 Data processing via MATLAB

All the data just introduced, will throughout the "Analysis and comparison"-section be handled, limited according to certain areas and visualized as both time series and distribution of data points onto maps. The following section will be a walk-through of the most important data processing methods used to represent relevant data. All of the data processing is performed in MATLAB. Essential MATLAB scripts and functions, created in connection with the collecting and visualization of the data, can be found in the appendices.

6.1 Generating time series

The cloud amount changes from month to month for the entirety of the Earth. In order to get a grasp of these changes over time, we will generate a time series for each satellite, giving a weighted mean for the entire Earth for each month. This is, as teased earlier in the section about MODIS, done through the dependence of both longitude and latitude.

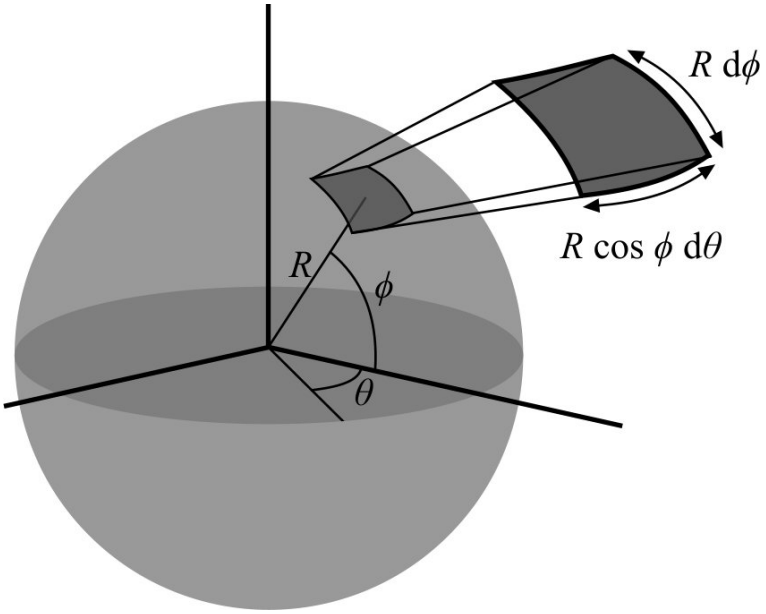


Figure 11: Area of a sphere

6.1.1 Surface of a sphere

As seen in figure 11 an area element on a sphere depends on the 2 angles: ϕ (polar angle = latitude) and θ (azimuth = longitude). Though the actual shape of the Earth is an oblate spheroid, the difference between the semi-major axis a and the semi-minor axis c, in regards to Earth, is very small, i.e. $|a - c| \approx 0$. So, we consider Earth as a sphere with constant radius R. The area of a square is, as always, given by the product of width by height. Here:

$$\begin{aligned} height &= R\,\mathrm{d}\phi \\ width &= R\,\cos\phi\,\mathrm{d}\theta \end{aligned} \tag{8}$$

The area is calculated:

$$\begin{aligned} \mathrm{d}A &= height \cdot width \\ &= R^2 \cos\phi \,\mathrm{d}\theta \,\mathrm{d}\phi \end{aligned} \tag{9}$$

Through this, the surface of the entire Earth can be calculated:

$$\begin{aligned} A &= \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R^2 \cos\phi \,\mathrm{d}\phi \,\mathrm{d}\theta = 2\pi R^2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} R^2 \cos\phi \,\mathrm{d}\phi \\ &= 2\pi R^2 \big[\sin\phi\big]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 2\pi R^2 \left(\sin\frac{\pi}{2} - \sin -\frac{\pi}{2}\right) \\ &= 4\pi R^2 \end{aligned} \tag{10}$$

We recognize the sphere surface formula, which we will use in our algorithm for calculating the weighted mean.

6.1.2 Weighted mean algorithm

This information will now be used to formulate an algorithm, that'll take the input of the area we wish to investigate (entire data set for the entire surface of the earth), and the longitude/latitude vectors corresponding to this area. Standardizing the process like this is what makes us able to take individual limited areas of the sphere and get a mean of said area, like when we transform MODIS or when we investigate areas with interesting activity.

As mentioned earlier, an area $\mathrm{d}A$ is given by the resolution as $res \times res$ ($0.25° \times 0.25°$ for the 6 SSMI satellites, $1° \times 1°$ for MODIS and $2.5° \times 2.5°$ for ISCCP and the overall SSMI data). Each of these $\mathrm{d}A$-areas has a respective data point, where the product of the 2 gives the total area of the surface covered by clouds within that area, given as $\mathrm{d}AC$ (area covered), i.e. the 3 are connected through:

$$\mathrm{d}AC = data \cdot \mathrm{d}A \qquad data = \frac{\mathrm{d}AC}{\mathrm{d}A} \qquad \mathrm{d}A = \frac{\mathrm{d}AC}{data} \tag{11}$$

Let's apply this to a larger area A containing $x \times y$ $\mathrm{d}A$-areas. These areas will be denoted with i and j, where $i = 1, 2, \cdots, x$ and $j = 1, 2, \cdots, y$.

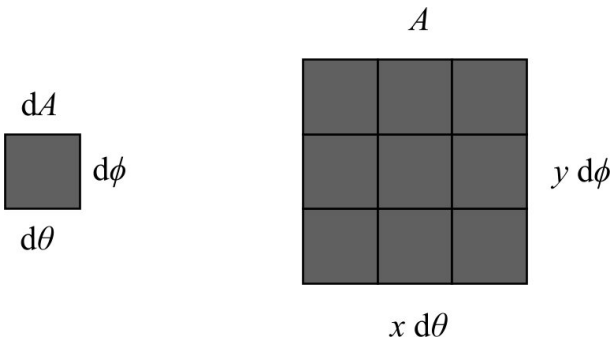


Figure 12: On the left we see a single area ($\mathrm{d}A$), to which a single data point is associated. On the right we see the larger area (A) composed of $x \times y$ $\mathrm{d}A$-areas.

The larger area A denotes a certain area of investigation (aoi). This could for example be the area off the coast of Angola or Peru, on which we focus in the analysis section 7.5.2. The area would be the surface of the sphere, if we're talking the entire Earth. Since the side of these $\mathrm{d}A$ squares are given as $res \times res$, we can write:

$$\begin{aligned} \mathrm{d}A &= \mathrm{d}\theta \times \mathrm{d}\phi & A &= x\,\mathrm{d}\theta \times y\,\mathrm{d}\phi \\ &= res \times res & &= x\,res \times y\,res \\ &= res^2 & &= x\,y\,res^2 \end{aligned} \tag{12}$$

The area of A (in m^2) depends on the latitude, since as we get closer to the poles, the individual $\mathrm{d}A$ areas shrink, even though the sides are of the same degrees (res). An area at a fixed latitude, will be of same size as every other area around the sphere, i.e. $\mathrm{d}A_1 = \mathrm{d}A_2$ as illustrated on figure 13:

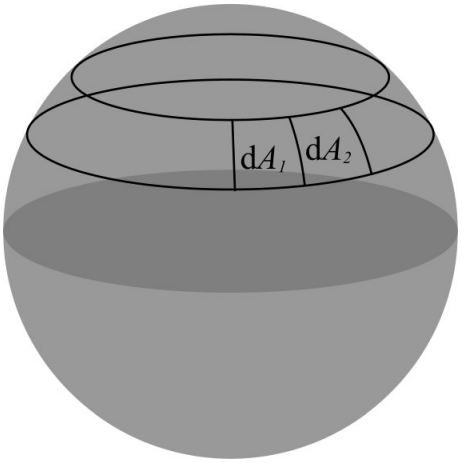


Figure 13: At a given latitude, every $res \times res$ area will be of equal area.

We can create a latitude matrix ($Lati_{i,j}$) similar to the vector presented in eq. 6. As explained before the large area A holds $x \times y$ smaller areas $\mathrm{d}A$ indexed with i and j. The $Lati_{i,j}$ matrix will then contain latitude coordinates in degrees for the respective $\mathrm{d}A_{i,j}$ areas. Using eq. 11, we can see 2 sums must be generated. First sum is A_{aoi}, which is the total area of the investigated area:

$$\begin{aligned} A_{aoi} &= \sum_{i=1}^{x}\sum_{j=1}^{y} \mathrm{d}A_{i,j} \\ &= \sum_{i=1}^{x}\sum_{j=1}^{y} R^2 \cos Lati_{i,j} \cdot \mathrm{d}\theta \cdot \mathrm{d}\phi \\ &= R^2 \cdot res^2 \sum_{i=1}^{x}\sum_{j=1}^{y} \cos Lati_{i,j} \end{aligned} \tag{13}$$

Second sum is AC_{aoi}, which is the total area covered by clouds within the investigated area:

$$\begin{aligned} AC_{aoi} &= \sum_{i=1}^{x}\sum_{j=1}^{y} \mathrm{d}A_{i,j}\, data_{i,j} \\ &= \sum_{i=1}^{x}\sum_{j=1}^{y} R^2 \cos Lati_{i,j}\, \mathrm{d}\theta \cdot \mathrm{d}\phi \cdot data_{i,j} \\ &= R^2 \cdot res^2 \sum_{i=1}^{x}\sum_{j=1}^{y} \cos Lati_{i,j} \cdot data_{i,j} \end{aligned} \tag{14}$$

The weighted mean for the investigated area, which will be used for our time series, is given as seen in eq. 11:

$$\begin{aligned} data_{aoi} &= \frac{AC_{aoi}}{A_{aoi}} \\ &= \frac{\sum_{i=1}^{x}\sum_{j=1}^{y} \cos Lati_{i,j} \cdot data_{i,j}}{\sum_{i=1}^{x}\sum_{j=1}^{y} \cos Lati_{i,j}} \end{aligned} \tag{15}$$

Some of the data sets contain extreme values, for example the ISCCP data for low cloud amount, that holds several -1000 values, or the SSMI satellite which only collected data above the sea, and in so have elements above land marked as 250-255 depending on type of land. The algorithm will have clearly defined individual limits, which will make them skip a data point, should it hold any of these extreme values, and in so not add anything to either sums (A or AC) for this particular data point. Eventually, the time series (TS) for the respective satellite in the investigated area will be generated through:

$$TS_{aoi,k} = \frac{AC_{aoi,k}}{A_{aoi,k}} \quad \text{for } k = 1, 2, \cdots, n \tag{16}$$

Here k denotes the respective month, of which there are n in total. Following, we see an example of the ISCCP total cloud amount time series.

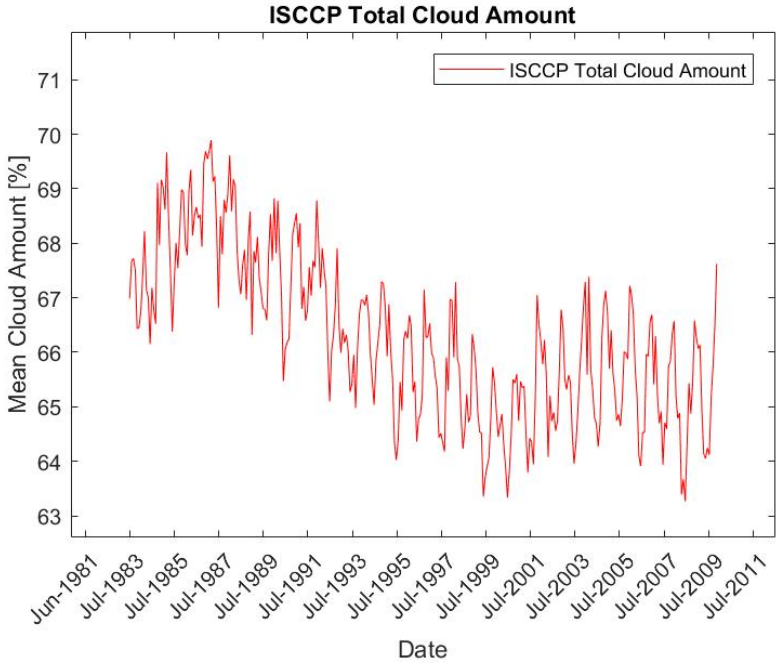


Figure 14: The generated time series for ISCCP total cloud amount in the period July 1983 to December 2009 for the entire Earth.

6.2 Removing seasonal variations (Anomaly)

In figure 14, we see several fluctuations beside the overall trend recorded over the period. These are seasonal variations throughout the year, which need to be removed in order to get some useful information, where we can see the actual recorded shifting trends. This will be further explained in the "Analysis and comparison" section. Initially 2 vectors of length 12 will be generated based on the time series and the months where data have been recorded.

$$S = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_{11} \\ s_{12} \end{bmatrix} \quad \text{and} \quad N = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{11} \\ n_{12} \end{bmatrix} \tag{17}$$

The elements of each vector represent the month of the year, i.e. the first element is January, second is February and so forth till 12'th element that is December. S denotes the sum of all data collected during that month and N denotes the total number of each month, in which data have been collected in the period. For example; in the period January 1983 till February 1985, data have been gathered for January and February 3 times (1983, 1984 and 1985) and 2 for the rest. Our N-vector's n_i-elements will then be 3 for $i = 1, 2$ and 2 for $i = 3, 4, \cdots, 11, 12$. The i'th element of S will then be the sum of all the n_i months of data collected for that month for every year.

Using these 2 vectors, a new vector containing the mean for every month will be generated. The mean for a data set x of length n is given by:

$$\mu_x = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{18}$$

Since S is the sum of the data collected for a respective month, and N is the length of the sets of data, we just have to divide element-wise:

$$\mu_{month} = S/N = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{11} \\ \mu_{12} \end{bmatrix} \quad \text{with yearly mean: } \mu_{year} = \frac{1}{12} \sum_{i=1}^{12} \mu_i \tag{19}$$

At last a new time series is created from the old one (TS_k) by subtracting the monthly mean μ_i from the respective month and adding the yearly mean μ_{year}:

$$TS_{ano,k} = TS_k - \mu_i + \mu_{year} \tag{20}$$

Here the i'th element is the month where the corresponding k'th element was recorded. $TS_k - \mu_i$ is then the deviation from the monthly mean. What we are left with in $TS_{ano,k}$ is the anomaly (ano). In figure 15, the original time series from figure 14 can be seen along with the new one, where the seasonal variations have been removed.

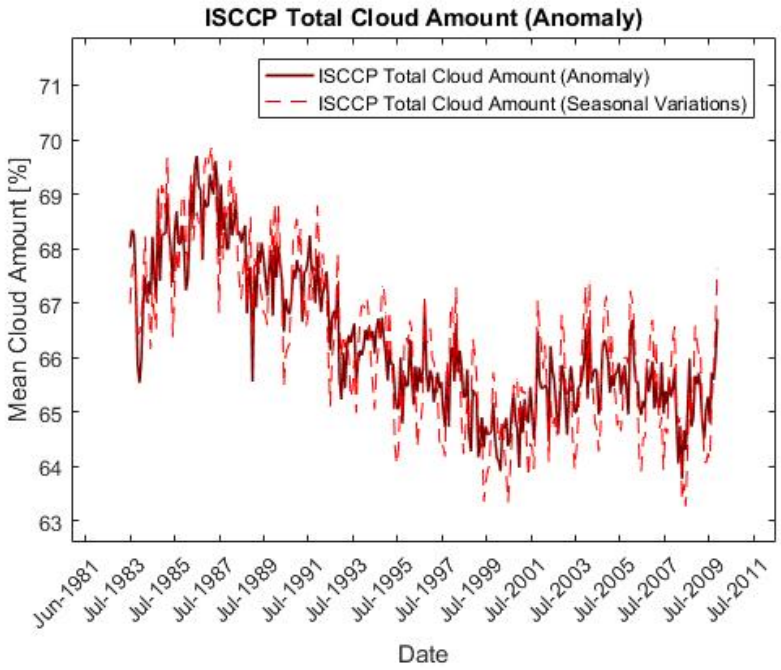


Figure 15: ISCCP total cloud amount with and wihtout seasonal variations.

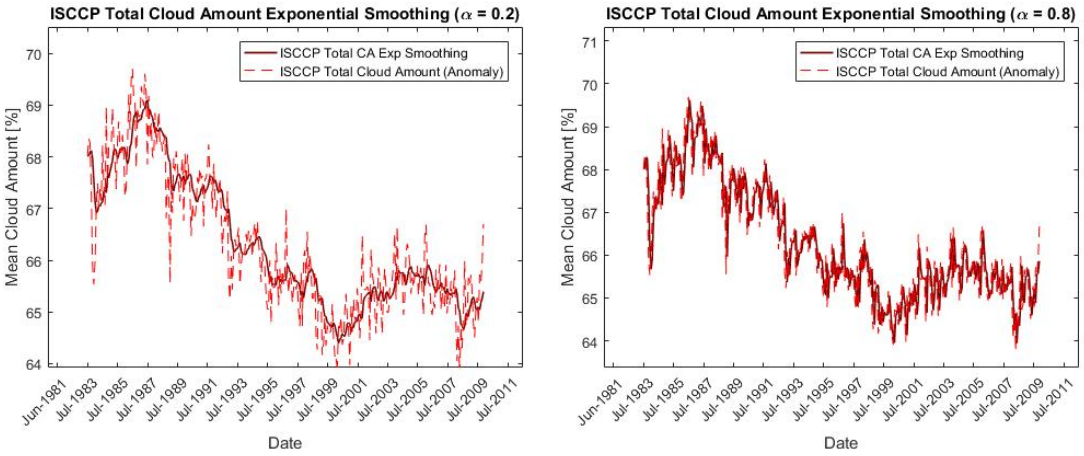
We see a lot of the more extreme fluctuations have been either removed or at least reduced. For all of the sets of data, we keep the time series both with and without the seasonal variations, since these tell us how different seasons of the year affect our data.

6.3 Exponential smoothing

Most of our data shows clear seasonal variations as just discussed. Even when filtering out the seasonal variations and only looking at the anomalies, some extreme values still show up, as can be seen in figure 15. In order to get an understanding of what happens over time, not just small fluctuations over 2-3 months, we will use exponential smoothing. This is a tool used in the analysis of time series in order to predict future values based on earlier values. How big this effect is depends the smoothing factor α. With the original data x_t and the exponential smoothing s_t, the smoothed series is given:

$$\begin{aligned} s_0 &= x_0 \\ s_t &= \alpha\, x_{t-1} + (1-\alpha)\, s_{t-1} \quad \text{for } 0 < t \end{aligned} \tag{21}$$

α is given as a value between 0 and 1. With a small smoothing factor ($\alpha \to 0$) we will see a quite smoothed graph in comparison with the original data (see figure 16a), whereas with a higher factor ($\alpha \to 1$), the smoothing will resemble the original data a lot more (see figure 16b).



(a) Smoothing factor 0.2

(b) Smoothing factor 0.8

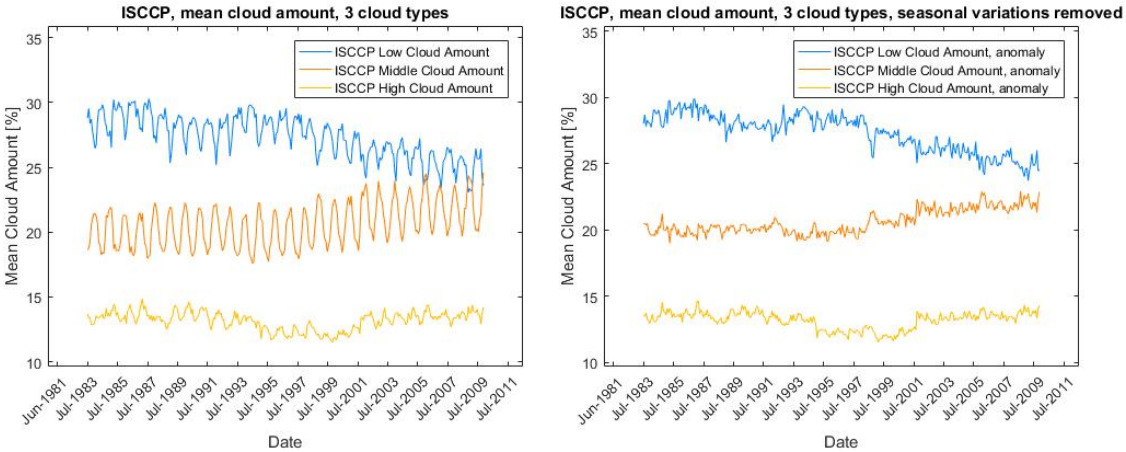
Figure 16: Example of exponential smoothing of the ISCCP Total Cloud Amount data

The exponential smoothing captures trends in the data. With a very low value of α we would see a limited delay on the tendency graph (the smoothing), in so we will keep this value of α whenever we perform the smoothing in the future. It follows the data's tendencies nicely, without giving too large fluctuations, and in addition doesn't have a noticeable delay. So when the exponential smoothing is done in section 7 we use:

$$\alpha = 0.2$$

6.4 Polynomial regression

We will initially take a look at the data collected for each cloud type from the ISCCP satellite:



(a) With seasonal variations

(b) Anomaly

Figure 17: The 3 cloud types from the ISCCP data. Left is before the seasonal variations are removed. Right is with the anomaly remaining.

Around 1998 we see spikes in the data for low and middle clouds. Afterwards the overall low cloud amount starts to decrease as the middle cloud amount simultaneously begins to increase. We will therefore investigate if this happens at the same pace by producing a regression model for either of the data. Looking at the data, we estimate that a polynomial of second order would be fitting in describing both the downwards tendency the low cloud data shows, as well as the upward tendency seen in the middle cloud data.

6.4.1 Regression by least squares method

Often when looking at data, we see a trend by which the data develops. We might see an exponential growth, a fall at a certain constant rate or possibly a trigonometric tendency. Using MATLAB, we can develop a model for the estimated in-/decrease just discussed. The polynomial regression model of n'th order is found using the least squares method [21]. It is given as:

$$F(x) = c_0 + c_1\, x + c_2\, x^2 + \cdots + c_n\, x^n \tag{22}$$

We can find the coefficients $c_0, c_1, \cdots, c_n$ in MATLAB by writing the basis functions $g_0, g_1, \cdots, g_n$, of which there are a total of $n+1$. For the polynomial regression these are simply given as:

$$g_0(x) = x^0, g_1(x) = x^1, g_2(x) = x^2, \cdots, g_n(x) = x^n \tag{23}$$

Eq. 22 can now be written more generally as:

$$y = \sum_{j=0}^{n} c_j\, g_j(x) \tag{24}$$

In the least squares method we find the values of c_j for which the errors are of the smallest possible magnitude. That is the difference between the regression model and the actual measured data squared:

$$\varphi(c_0, c_1, \cdots, c_n) = \sum_{k=0}^{m} \left[\sum_{j=0}^{n} \left(c_j\, g_j(x_k) - y_k \right)^2 \right] \tag{25}$$

Here x_k is our time vector and y_k is our actual data with $m+1$ data points. To minimize the errors we set the conditions:

$$\frac{\partial \varphi}{\partial c_i} = 0 \quad \text{for } i = 0, 1, 2, \cdots, n \tag{26}$$

Solving these partial derivatives and setting them equal to 0 we get:

$$\sum_{j=0}^{n} \left[\sum_{k=0}^{m} g_i(x_k)\, g_j(x_k) \right] c_j = \sum_{k=0}^{m} y_k\, g_i(x_k) \quad \text{for } i = 0, 1, 2, \cdots, n \tag{27}$$

These are also known as the normal equations, which are always given as:

$$(\mathbf{A}^T \mathbf{A})\mathbf{x} = \mathbf{A}^T \mathbf{y} \tag{28}$$

Where $\mathbf{A}$ is a $(m+1) \times (n+1)$ matrix, $\mathbf{x}$ is a vector of length $(n+1)$ and $\mathbf{y}$ is a vector of length $(m+1)$:

$$\mathbf{A} = \begin{bmatrix} g_0(x_0) & g_1(x_0) & \cdots & g_n(x_0) \\ g_0(x_1) & g_1(x_1) & \cdots & g_n(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ g_0(x_m) & g_1(x_m) & \cdots & g_n(x_m) \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_m \end{bmatrix} \tag{29}$$

This means that $\mathbf{x}$ is actually our coefficient vector, and by isolating $\mathbf{x}$ from eq. 29, MATLAB will calculate our coefficients $c_0, c_1, \cdots, c_n$, which can then be applied to eq. 22 for the final polynomial regression model.

$$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y} \tag{30}$$

6.4.2 Second order regression

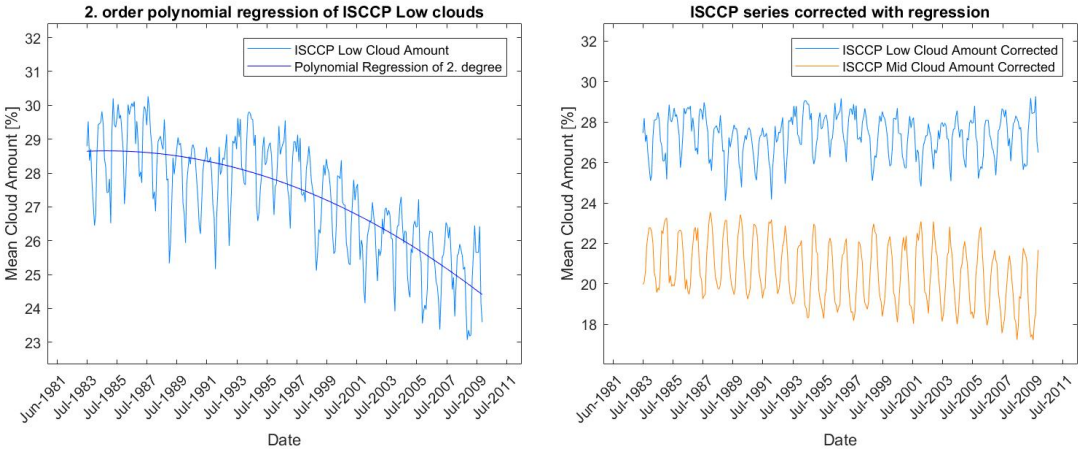
We will now apply this method to our time series of the low cloud data over n months. First of all we will make the second order regression on the data.

$$y_{low,k} = c_{0,low} + c_{1,low}\, x_k + c_{2,low}\, x_k^2 \quad \text{for } k = 1, 2, \cdots, n \tag{31}$$

Afterward the mean of the function over the period of measurement will be subtracted from the entire series. This gives us the deviation from the mean only, which we call δ.

$$\delta_{low,k} = y_{low,k} - \frac{1}{n}\sum_{k=1}^{n} y_{low,k} \tag{32}$$

To correct the data, we add $\delta_{low,k}$ to the entire mid cloud series and subtract it from the low cloud series. The result of this can be seen on figure 18b

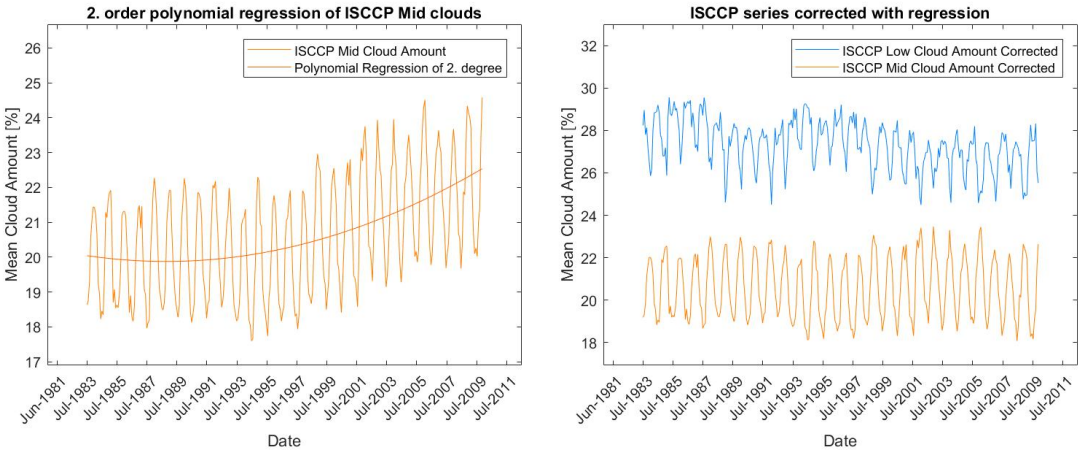


(a) Regression of low cloud

(b) Mid and low clouds corrected

Figure 18: The second order polynomial regression performed on the low cloud time series with seasonal variations, and afterwards corrected on both low and mid cloud data

On figure 19 we see the same procedure performed, both with the regression done on the mid cloud time series instead.



(a) Regression of mid cloud

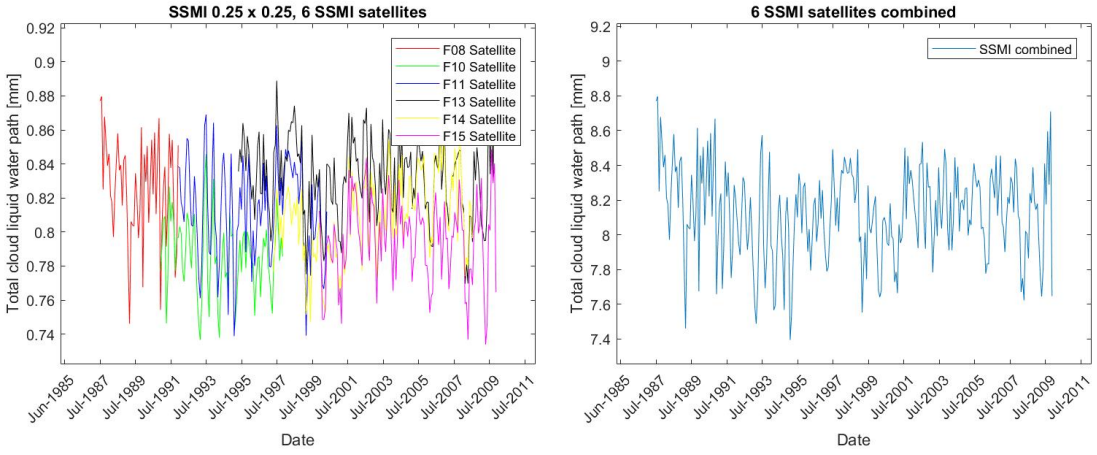
(b) Mid and low clouds corrected

Figure 19: The second order polynomial regression performed on the mid cloud time series with seasonal variations, and afterwards corrected on both low and mid cloud data

When comparing both of the corrected figures, we see a difference between the two. But our assumption that they develop equal opposite paces isn't completely wrong, given that they are both being straightened out, and not taking off further in a general direction. More factors would though have to be taken into account, an area that will be reflected further in section 7.1

6.5 Combined representation of data from the 6 SSMI satellites

We have, as already explained, 6 SSMI satellites collecting data at an resolution of $0.25° \times 0.25°$ operating in different periods. These provide us with different time series for the individual satellites, that can seem very confusing and chaotic when represented all at once in a single plot (see figure 20a). Thus, a new "combined" time series will be generated by looking at periods where 2 or more satellites overlap each others, sum together the data for every satellite operating the respective month, and hereafter dividing by the number of satellites, which data we just summed. Why we see this as an acceptable way to combine the 6 satellites' times, will be explained in section 7.3.2. The first data was collected by the f08 satellite in July of 1987 and the last by the f15 satellite in December of 2009. This gives a total of 270 months of data. When combining these, we get the time series seen in figure 20b.



(a) Time series for the 6 satellites, raw data

(b) Combined model, raw data

Figure 20: On the left we see the 6 satellites operating in their respective periods. On the right is the combined time series.

7 Analysis and comparison

In this section we investigate the 3 types of data (ISCCP, SSMI and MODIS), to find out if the data sets are actually consistent within the same time period and geographical zones. We would expect, as we plot the time series for the satellites, that they show the same patterns and seasonal variations.

7.1 ISCCP

We start by investigating the ISCCP data. ISCCP has measured the mean cloud amount for low-mid-high cloud types all over the earth in the time span 07/1983 - 12/2009 . In this period we would expect the average amount of low-mid-high clouds to be somewhat constant. Of course we would expect some seasonal variations for each cloud type depending on the season, but the overall cloud amount is thought to be stable as time progresses. This assumption is made, having read nothing indicating otherwise.

7.1.1 Unprocessed data

We are using 4 variables from the data set, "total cloud amount", and "mean low- mid- high-cloud amount" (see section 4.1 how these are measured). These variables are used in figure 21. The blue line shows 1 variable: "Total mean cloud amount" taken directly from the ISCCP data. The red line is the 3 variables low- mid- & high cloud amount from the data set summed up. We see that the sum of the 3 variables is a bit lower than the total mean cloud amount suggesting that some data might be lost in the process of splitting the data into subsets of data (around 7-8 % get lost in this process of splitting into 3 separate variables). We do not know why this sum of low, mid & high clouds does not add up to the total cloud amount.
The time evolution of the 3 cloud types are plotted as a funtion of time in figure 22. As with the total cloud coverage of the Earth, we would expect the amount of high clouds to be rather stable as a function of time.

From figure 22 we see that the high cloud amount is pretty stable over the 26 year period, as we would expect. This expectation comes from a general assumption that the ratio between low-mid and high level clouds are somewhat constant in time, this might not be the case as an undiscovered long term trend might influence these ratio's. The low and mid clouds, though, behave significantly different than expected. The low and mid cloud amounts are stable (with season variations) until 1997. After this point the low cloud amount drops significantly and the mid cloud amount increases. This sudden change in pattern needs to be investigated to try and explain what this change is caused by. Is it caused by changing hardware on board the satellites and/or bad calibrations as new satellites are sent in orbit, or is this change in pattern actually real evidence of the changing cloud coverage of the Earth, and that the cloud coverage is changing?
The main goal of this work will be to estimate the likelihood of the change in observations seen in figure 22. For this reason the SSMI and MODIS data are included in this analysis. Are SSMI and MODIS showing the same tendencies or is the ISCCP data the only data showing these tendencies?
By investigating the amount of low and mid level clouds we see that the seasonal patterns are anticorrelated. When the low clouds reaches a the maximum cloud amount in the yearly seasonal pattern, the mid level clouds are at a seasonal minimum. Further we see that the high

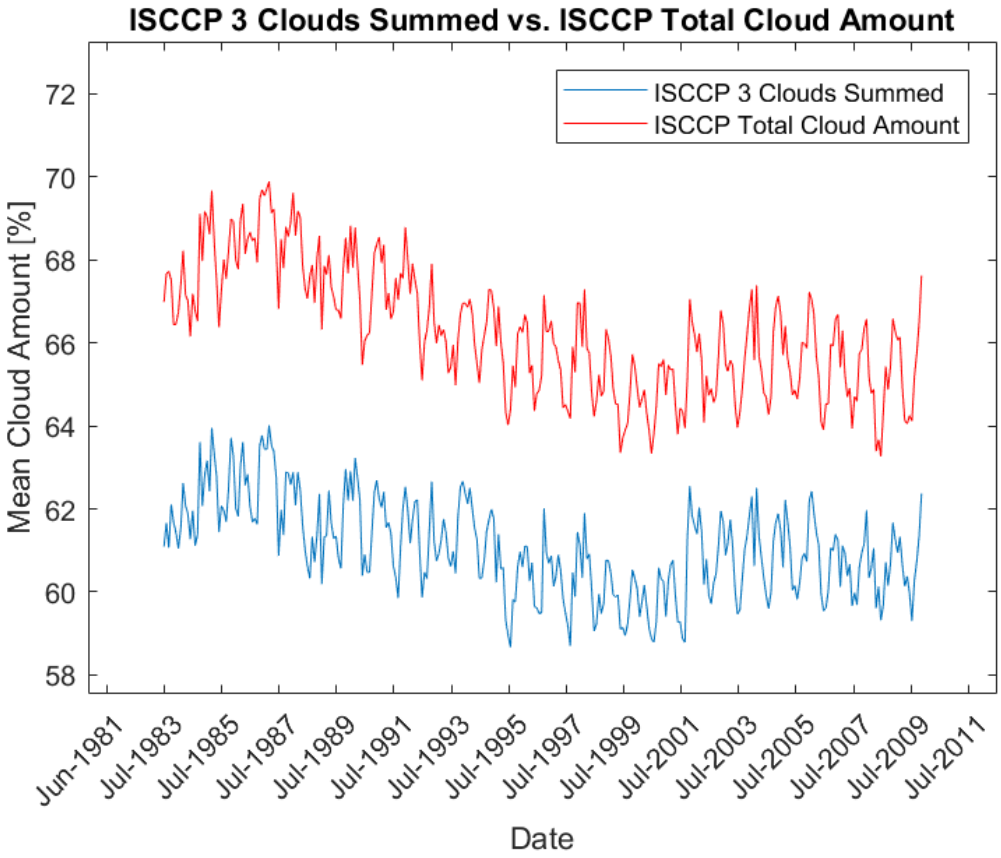


Figure 21: ISCCP, mean of low-mid-high clouds summed plotted against total mean cloud amount. The red line shows variable total mean cloud amount, the blue line is a sum of the 3 varables "low cloud amount" , "middle cloud amount" and "high cloud amount"

cloud pattern follows the low cloud pattern, . The later comparison between the three types of data (ISCCP,SSMI & MODIS) we wish to make, does not account for high clouds (which are ice clouds) since SSMI and MODIS only measures the liquid water path. Therefore we do not go into further details with the high level clouds.

The correlation coefficient between the amounts of low and mid level clouds is -0.8336. A correlation factor of this magnitude is a clear indication that the low cloud coverage is dependent of the amount of mid level clouds at a given time (or the mid level cloud amount depends on the low cloud amount) and that the data are strongly anti-correlated. We see that the low cloud amount reaches a maxima in the summer time (June,July) whereas the mid level clouds in these months are at a minima. In December/January the trends are opposite (low clouds at minima and mid clouds at maxima).
This large anti-correlation factor is probably occuring because the satellite measures from above. When the satellite is measuring from above, the satellite is viewing the top layer visible from the satellite. This is causing the low level cloud amount to drop when the mid level cloud

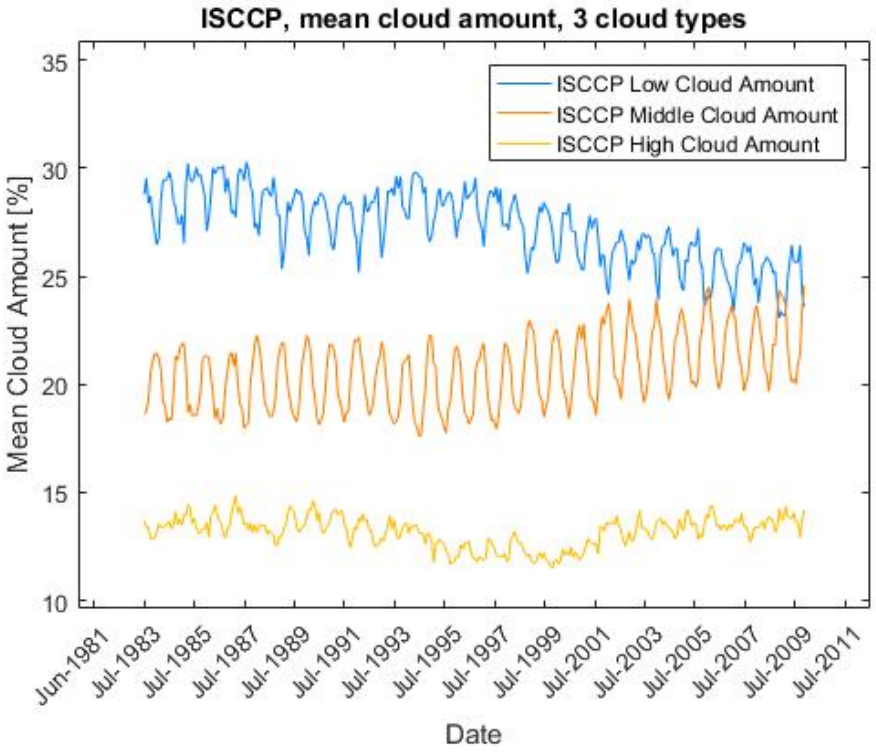


Figure 22: ISCCP, time series of the amount (in %) of low, medium, high clouds respectively. The low and middle cloud plots are anti-correlated and changing in amount over time. The high cloud amount seems independent of changes in the low and middle cloud amounts.

amount increases since mid level clouds are likely to be opaque due to the density of the mid level clouds.
This phenomen does not occur as much with high level clouds since high density clouds, that high in the atmosphere is less likely to occur. We would therefore expect the mid and low level clouds to be independent of the high clouds. This is also the case as analysis shows correlation factors of -0.1676 for low vs high clouds, and 0.2744 for mid vs high clouds. Because of this independence, the high level cloud data is not considered further. The high cloud amount is, however, included in the variable "Total cloud amount".

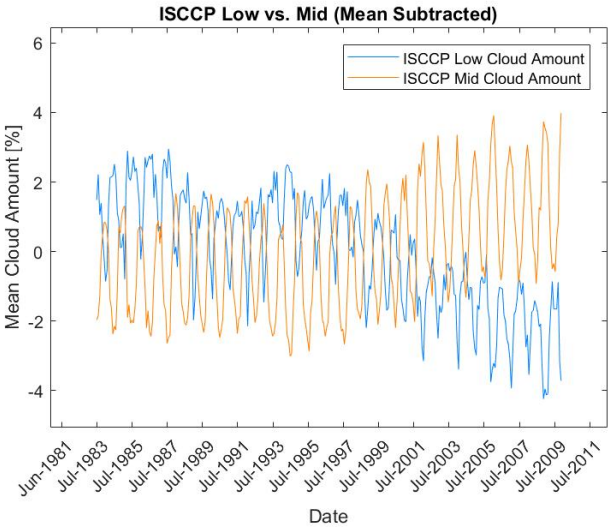


Figure 23: ISCCP, Low and mid level clouds with their respective means subtracted. The difference in trends is clearly shown. The low cloud amount drops as the mid level cloud amount increases.

Figure 23 shows a plot of the amounts of low and mid level clouds. The respective means have been subtracted from each of the two time series. In this plot the anti-correlation becomes very clear especially in the period after 1998 where the two types move in complete opposite directions (mid level clouds increase and low clouds decrease in amount) and the maximum/minimum is clearly phase shifted by half a period.

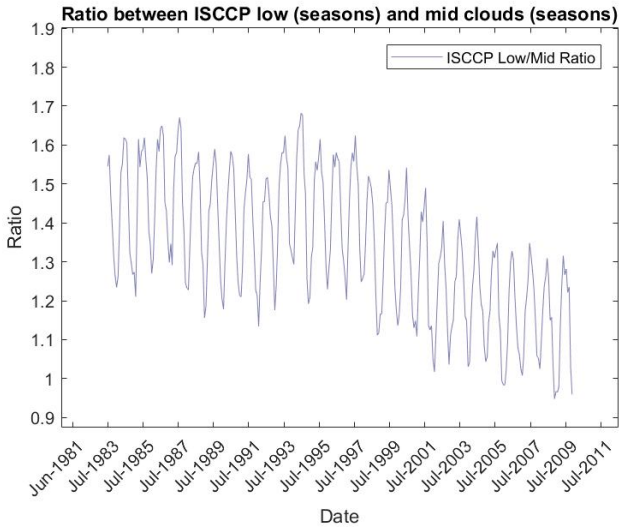


Figure 24: ratio plot between ISCCP low cloud amount and ISCCP middle cloud amount. The change in ratio are significant, we see a ratio between low level clouds and mid level clouds drops significantly, as the mid level cloud amount increases, causing the low level cloud amount to decrease. This further indicates an anti-correlated relationship between the low- and mid level cloud amounts. The spacing between the observed peaks is one year apart.

7.1.2 Processed data

As described in the section "regression" (section 6.4), we are fitting a second degree polynomial to the amount of mid level clouds time series. This polynomial fit (and upwards trend) is then added to the low cloud level data. If the two opposite moving trends are perfectly anticorrelated we should see the low and mid level clouds to be very stable in the 26 year period after having done the fit. This regression is plotted in figure 25. After having done the regression we see the pattern we were expecting to see going in to this project. The low and mid level clouds are pretty stable with yearly seasonal variations.

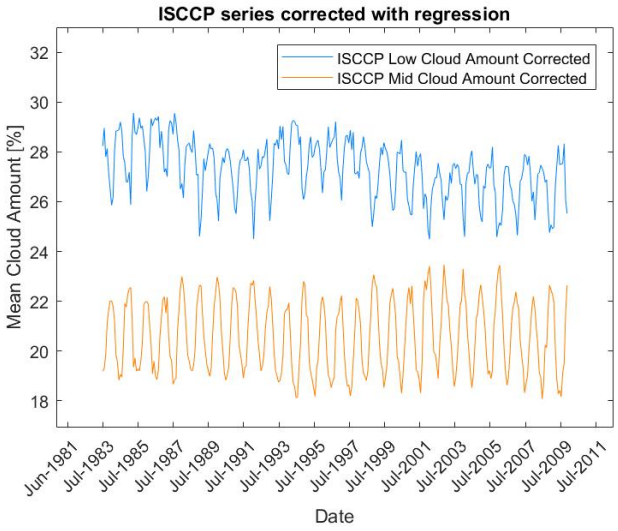


Figure 25: ISCCP, blue: Low level clouds (with added regression) and yellow: mid level clouds (with subtracted regression). With this regression added/subtracted the two plots are stable. This further indicates that changes in the middle cloud amount directly affects the low cloud coverage.

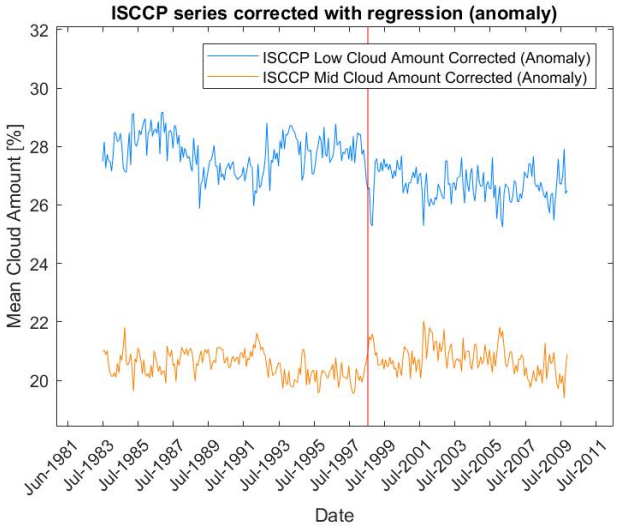


Figure 26: ISCCP corrected low & mid level clouds anomaly plot. The downwards trend for the low clouds. The anomalies shows some minor fluctuations, except for the big spike in late 1998 (red line) which is caused by a significantly stronger El Niño than usual.

In order to compare the two sets in more detail, we now subtract the respective means of each time series. This is shown in figure 27.

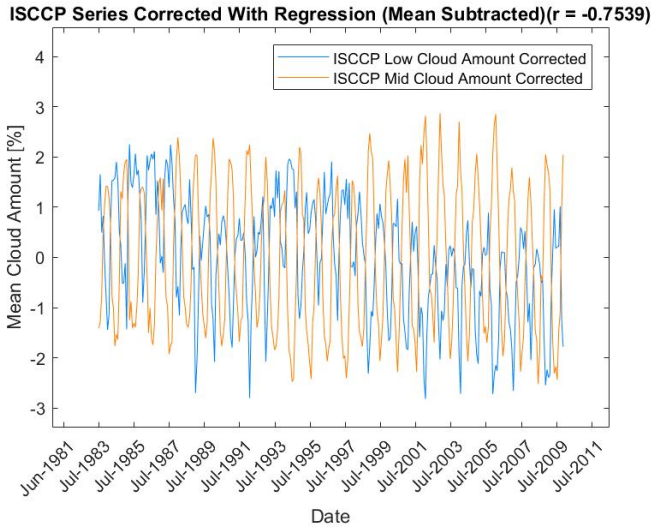


Figure 27: ISCCP, low and mid level clouds with the regression applied, and the mean subtracted (raw data). The mean of each cloud type subtracted individually. The plot clearly shows a strong anti-correlation between low and mid level cloud amounts.

In the further processing of the data, the seasonal variations are removed. With the seasonal variations removed in figure 28 the anomalies in the data is much clearer (removal of seasonal variations explained in section 6.2). Some of these anomalies can be explained by specific events such as the sudden increase in mid level clouds in mid 1997 - mid 1998, and drop of low level clouds. This drop was probably caused by a significantly stronger El Niño [22] in that period than observed in the earlier years. These sudden variations can be explained by natural phenomena, the general trend in the mid and low clouds cannot, with the knowledge known about the Earths climate so far.

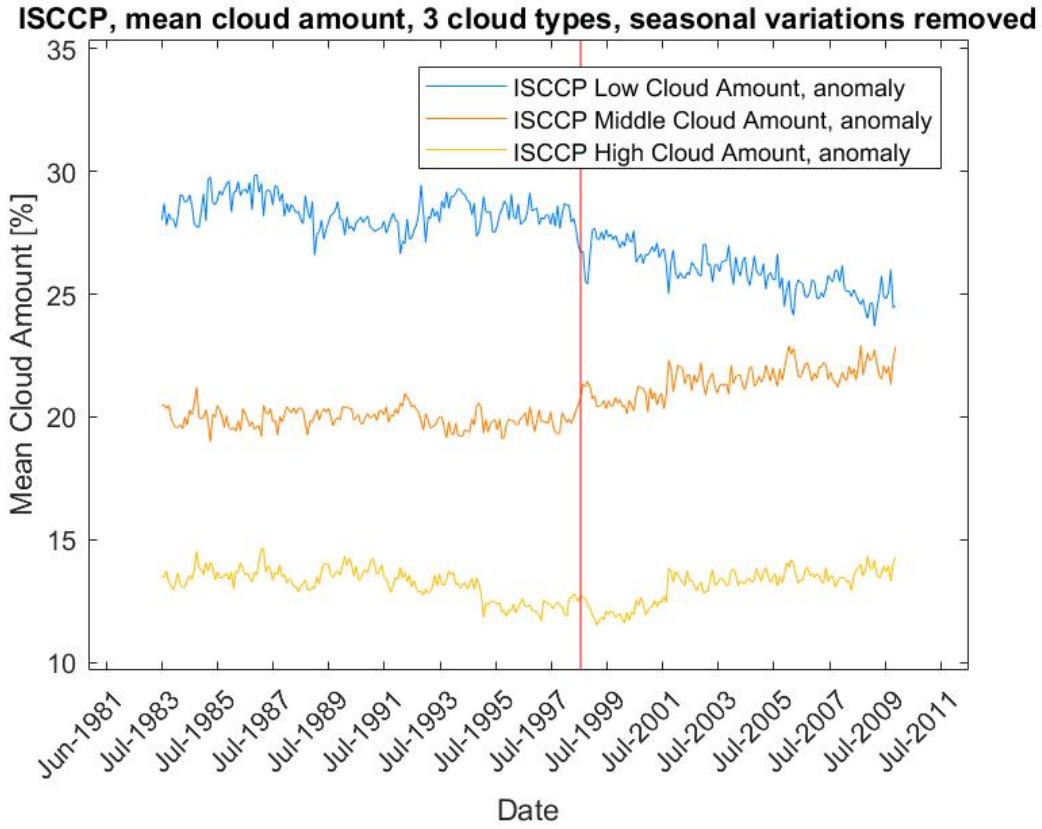


Figure 28: ISCCP, time series of low-mid-high clouds with seasonal variations removed. Note that the regressional correction is not applied in this plot. An especially large anomaly is seen in late 1998 (indicated by the red line). This anomaly is due to a significantly stronger El Niño than usual causing a drop in the low cloud amount and increase in the mid level cloud amount. Due to warmer temperatures during El Niño we would expect more convection to take place, increasing the altitude of the (former) low level clouds. This is most likely why we see more mid level clouds in this period. The high cloud amount is not affected as much as low-mid level clouds

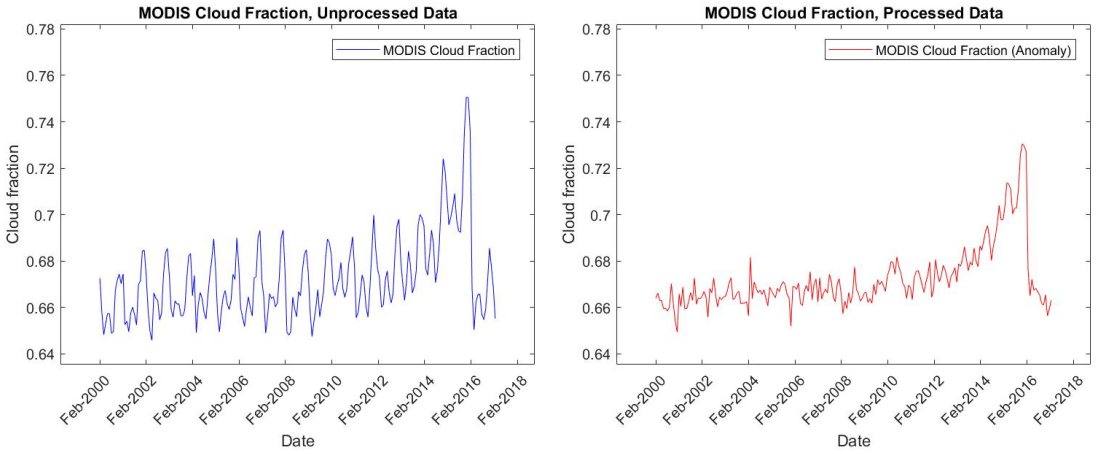
This concludes the data processing of the ISCCP data.

7.2 MODIS

In figure 29a we plot the unprocessed MODIS data. The data runs from February, 2000 to March, 2017. MODIS measures the total mean cloud amount of the Earth. As seen from the figure, the amount of clouds are somewhat constant with a slight increase in the amount of clouds in the period from 2000-2014. We see large yearly seasonal variations around June-July, and some smaller variations in December-January, indicating the MODIS data observes a half year seasonal variation and a yearly seasonal variation in the total cloud coverage.
In the period 2014 - 2016 a large increase in the cloud amount is seen before a very large and sudden drop. In the the last year or so of data we see (perhaps) the cloud amount stabilizing

again. Without further data this only is only speculation though.
This sudden increase and drop in the cloud amount is obviously a major deviation relative to the seasonal variations.

Figure 29b shows the same data with the seasonal variations removed. After 2014 a large sudden increase in the cloud amount is again seen, so this increase is not caused by regular seasonal variations. Like the sudden increase seen in the ISCCP data in section 6.1.2, this increase is most likely caused by a very strong El Niño happening from may 2014 - may 2016. El Niño has been especially strong in the years 1982-1983, 1997-1998 and 2015-1016 [22].



(a) (b)

Figure 29: (a) MODIS, time series of all available data. We see a very large increase and drop in the cloud amount from 2014-2016. This large increase is most likely caused by a very strong El Niño in that period. (b) MODIS, anomaly plot of all available data. The large increase in 2014-2016 is not part of any seasonal variations in the cloud coverage. The increase is caused by El Niño.

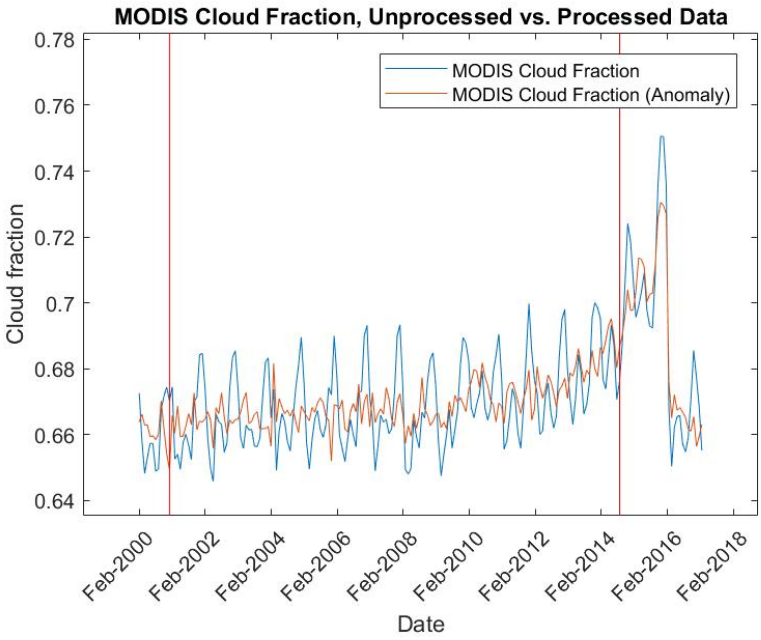


Figure 30: MODIS, unprocessed data vs anomaly data. In 2001 a smaller fluctuation is shown by the first red line. in 2014 the second red line indicates the beginning of El Niño and the large fluctuation this causes.

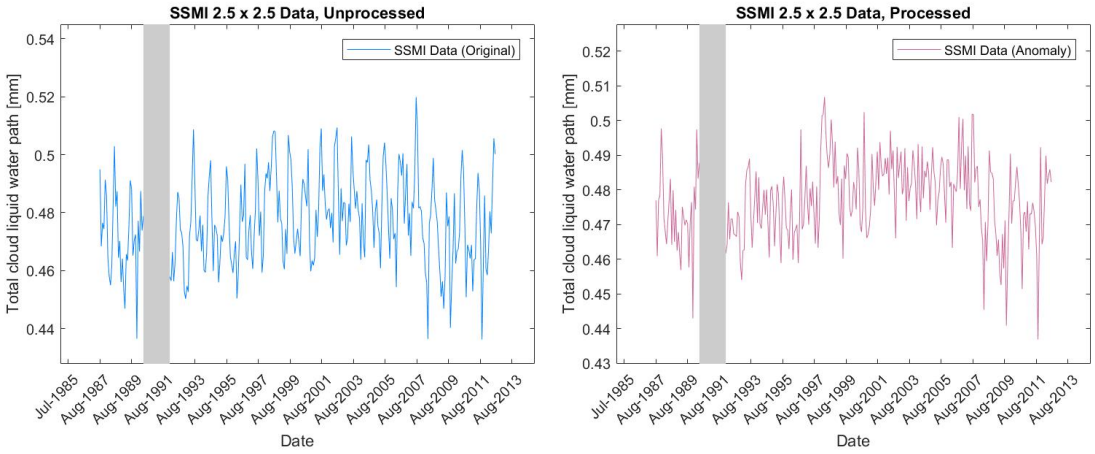
In figure 30 we plot the unprocessed data and the data with seasonalities removed. This further substantiates the claim that the event variations observed are caused by extraordinary events. Furthermore we see some smaller drops in the beginning of the year 2001 and again in late 2005 where the original data and the anomaly data move in opposite directions locally. The smaller fluctuations are most likely caused by. The first of these "anomaly drops" seems to be appearing because the cloud amount in beginning 2001 has a lower seasonal peak than the following seasonal peaks. These smaller fluctuations are tough to pinpoint exactly why they are observed, many factors have an impact on the Earth's climate, this smaller fluctuations doesn't seem to be a major event as we see a much larger increase for the later part of the series with El Niño active.

7.3 SSMI

From SSMI we have two different types of data. We have a high resolution (0.25 x 0.25 degrees) data set, and a low resolution data set (2.5 x 2.5 degrees). The SSMI measures the "total cloud liquid water path", a measurement of the total cloud liquid water contained in a vertical column of the atmosphere. This differs from the ISCCP and MODIS data, which measure the total cloud amount. For further explanation of how the SSMI measurements are done see section 4.3.
Although the SSMI data are not output in the same units as ISCCP and MODIS, we compare the SSMI to our collection of data, to investigate whether or not the overall trends are consistent. In this analysis we will further investigate two geographical areas known for having a majority of clouds in the lower levels of the atmosphere.

As the atmosphere is colder as the altitude increases, the water in the clouds (which SSMI measures) will freeze at higher altitudes. This influences the SSMI readings, measuring the "liquid water path", meaning SSMI doesn't make any measurements in the higher parts of the atmosphere where liquid water doesn't occur. In theory a good comparison should be possible between the SSMI data and the ISCCP low level cloud data, given the area of investigation is containing solely low clouds (or at least close to). This is because all clouds in this area consists of liquid water which is measurable by SSMI.
Given all clouds in the area are located in low altitudes a direct comparison between ISCCP low clouds and SSMI liquid water path should therefore be possible.

7.3.1 SSMI 2.5 x 2.5 degree resolution data



(a) (b)

Figure 31: (a) SSMI 2.5 x 2.5 degree resolution, original data. We see very large swings with in the data with fluctuations upwards of 17% from global minimum to global maximum. This might suggest the measurements are very uncertain. (b) SSMI 2.5 x 2.5 degree, anomaly data. After the seasonal variations are removed. We still see many large fluctuations. This could indicate that the SSMI 2.5 x 2.5 degree data contains large uncertainties connected with the measurements. The grey area indiciate the motnhs of missing SSMI 2.5 x 2.5 data.

In figure 31a the unprocessed 2.5 x 2.5 SSMI degree data are presented. A one year seasonal pattern is seen, just as in ISCCP and MODIS. The data shows large fluctuations from minimum to maximum. The data shows measurements from 0.43 global minimum to 0.52 global maximum, this is a difference from low to high of about 17%, a significant deviation. In the ISCCP we saw deviations upwards of 10 %, which is significant but less than the SSMI data.

In figure 31b the anomaly of the the SSMI 2.5 x 2.5 is plotted. The anomaly plot shows large deviations as discussed previously. These large variations seen after the seasonal variations is removed is an indicator that atmospheric water path might be tougher to measure than the cloud amount (which also has difficulties).

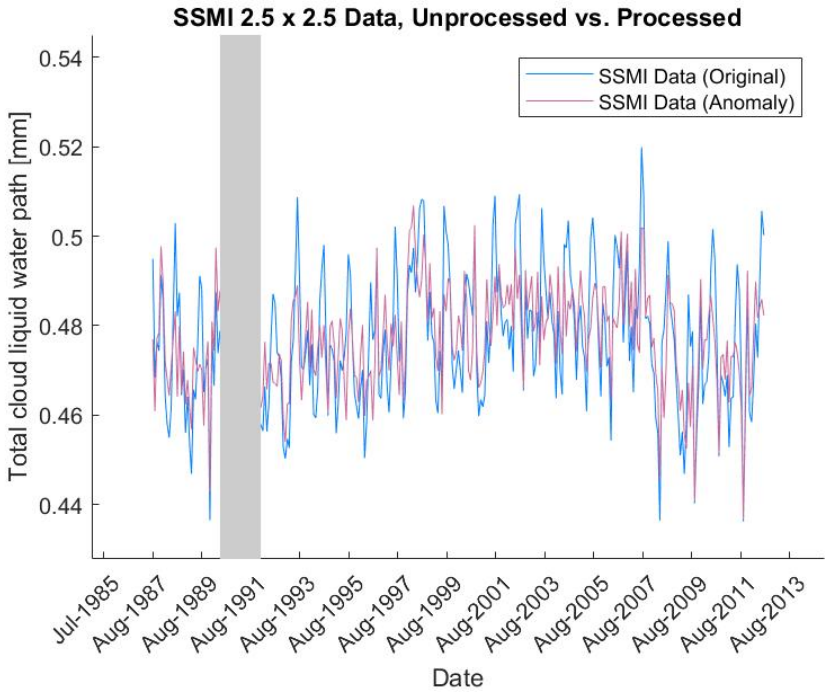


Figure 32: SSMI 2.5 x 2.5 degree, original data vs anomaly data. The large anomaly fluctuations is seen alongside the original data. The fluctuations in the data are decreased slightly by removing the seasonal variations. But the anomaly data is still very "noisy"

Figure 32 shows the data alongside the anomaly of the data. The deviations are so eye catching that it is tough to conclude anything. The trend line does seem to be somewhat stable as time progresses with a mean of 0.477. In the SSMI single events are hard to pinpoint due to the large underlying deviations. In this section we have seen large deviations in the SSMI data, this could make a comparison between SSMI and the ISCCP low cloud amount difficult. This comparison will be done in section 7.5.2.

7.3.2 SSMI 0.25 x 0.25 degree resolution data

As described in section 5.1 (Data processing) the higher resolution data (0.25 x 0.25 degrees resolution) is data collected from the f08-, f10-, f11-, f13-, f14- and f15-satellites in the period July 1987 - December 2009.

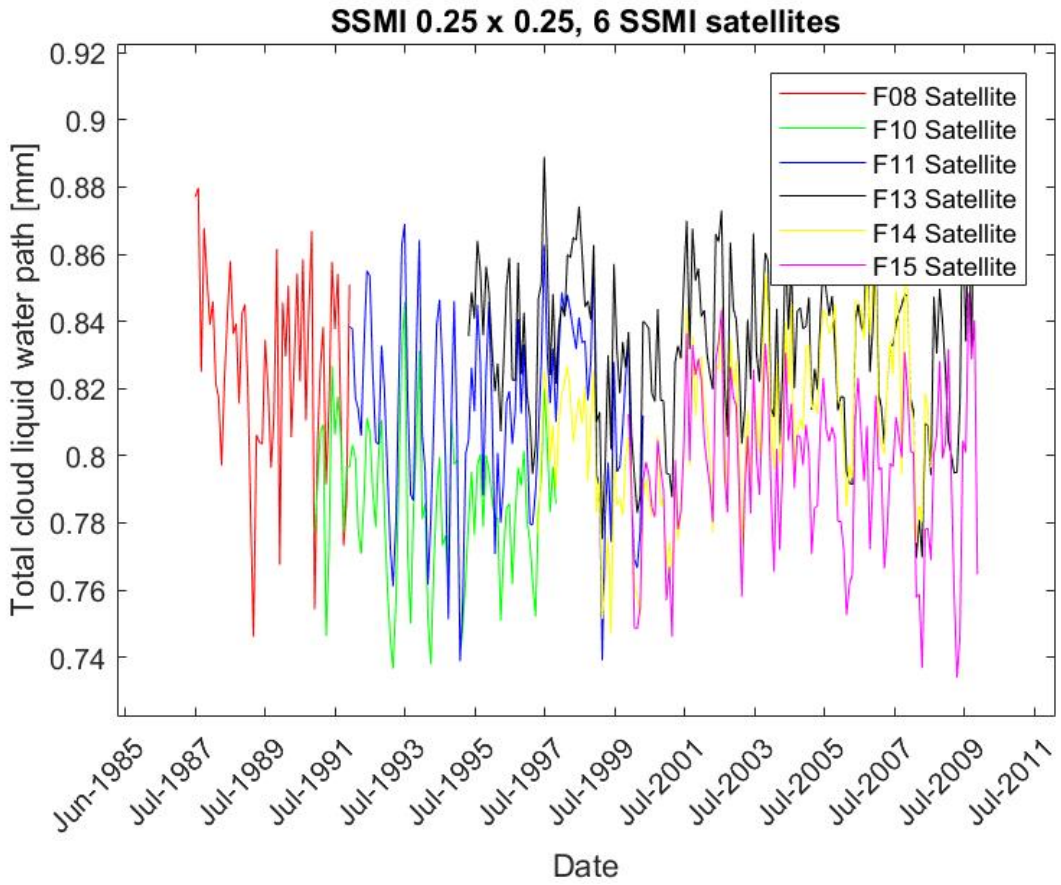


Figure 33: SSMI 0.25 x 0.25 degree, original data from the six satelittes. In this plot it is difficult to see exactly how well the patterns of each satellite is consistent with the patterns of the other satellites, due to small overlaps in some places (e.g. between red and green). Overall, the variations of the data sets apper consistent.

The data from these satellites are plotted in figure 33. The satellites lifespan overlap so the differences and similarities from satellite to satellite is clear. Most of the satellites seems to be showing the same tendencies with different amplitudes, except for the F13 satellite (black). The F13 satellite seems to deviate from the F11 (blue) and F14 (yellow) in late 1997 - early 1998.

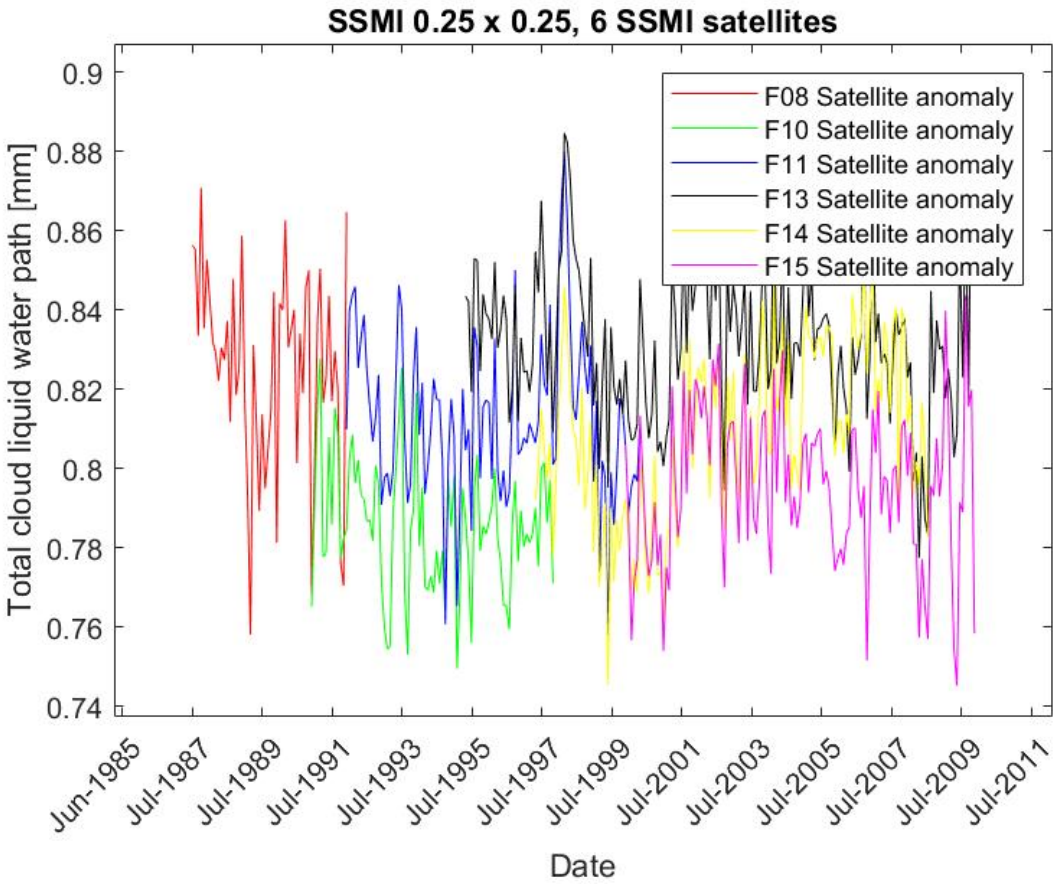


Figure 34: SSMI 0.25 x 0.25 degree, data from the six satellites with seasonal variations removed. Again specific details are hard to pinpoint, but the overall patterns of each satellite do seem to correspond with the other overlapping satellites. We do see large differences in amplitude especially between F10 (green) and F11 (blue), and between F10 (green) vs F13 (black) in the period they overlap. The differences in amplitudes from satellite to satellite is roughly around 5% ($\frac{0.04}{0.80} = 5\%$) since the difference in the measurements for most satellites is roughly 0.04 mm. Some satellites show larger differences e.g. the pink (F15), this satellites deviates by roughly 0.07 from the black (F13) which is roughly 9%.

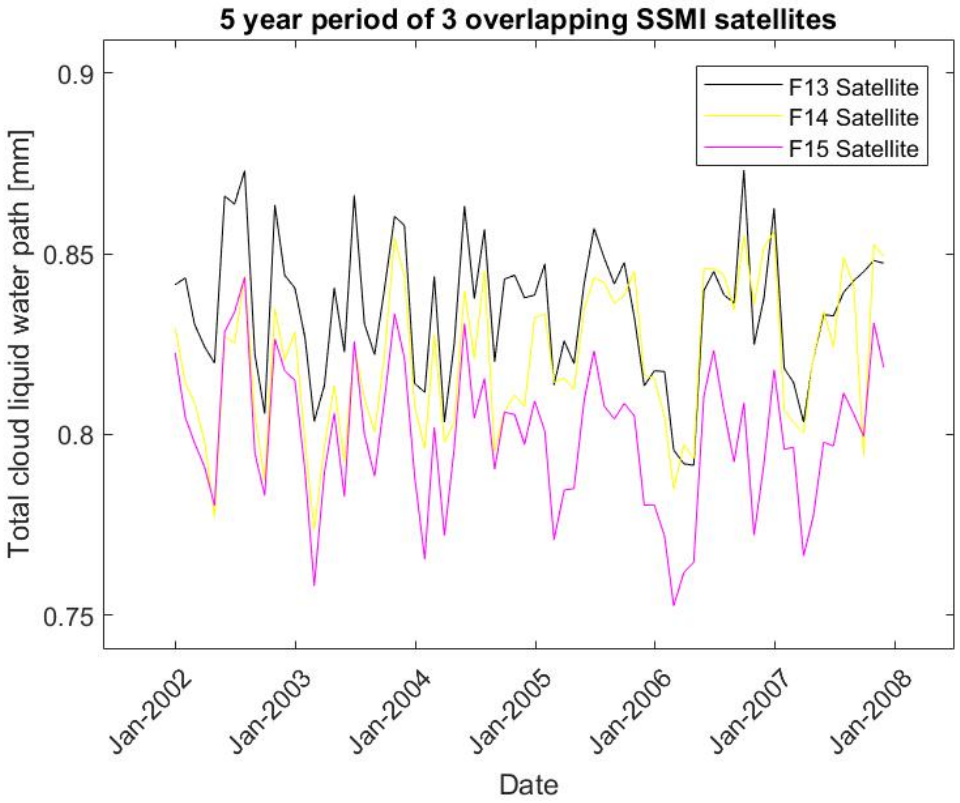


Figure 35: 5 year period of overlapping satellites enlarged. The 3 satellites overlapping in this observational period are showing the same seasonal tendencies but with differences ($\sim 5\%$) in amplitudes. Since the satellites show the same patterns an averaged plot of the six satellites seems valid.

In figure 34 we plot the anomaly of the satellites. The deviation for the F13 satellite we saw in the raw data is removed as it was a seasonal variation spike. The anomaly plot furthermore gives some insight into how difficult it can be calibrate subsequent satellites. These satellite should in theory measure the exact same thing should they be calibrated perfectly. This is clearly not a perfect calibration from satellite to satellite, and we see deviations ($\sim 5\%$ in most overlapping points of observations) in amplitudes across the field of satellites.

Table 5: Correlation factors between overlapping satellites. The table indicates difficulties in calibrating a new satellite to an "old" satellite already in orbit. We would expect every satellite to make the same observations should they be perfectly calibrated. For F14- and F15 satellites we find a correlation of 0.7575, this might be an indication of a bad calibration.

Satellite	Overlapping satellite	Correlation	Months of overlap
F08	F10	0.7020	12
F10	F11	0.8563	83
F10	F13	0.9188	31
F11	F13	0.8919	60
F11	F14	0.9061	36
F13	F14	0.8536	136
F13	F15	0.8610	119
F14	F15	0.7575	93

In table 5 the correlation factors between the overlapping satellites are listed.
In the table, two correlation coefficients are significantly lower than the others. They are for the F08 vs F10 satellites and F14 vs F15 satellites. The F08 vs F10 correlation is based on only 12 overlapping months of data, a large uncertainty for this low amount of overlapping data is to be expected. F14 vs F15 is more interesting, here we have 93 months of overlapping observations and we find a correlation of 0.7575. This low correlation is a clear indication of calibration difficulties, because the satellites measuring in the same period should measure the same patterns, and therefore be well correlated (if the calibration is done successfully). The other satellites shows better correlations from 0.85-0.92, not great, but at at least we can say with some confidence that the satellites actually measures most of the same events. It is concerning how large the uncertainties is for the SSMI data, and questions how use full this data is.

Figure 36 shows the measurements of the 6 satellites combined into one single plot. This is done by taking the average value for the overlapping data. As mentioned before the data is clearly showing large uncertainties. This combination of the six individual satellites is only valid if the data from the six satellites are consistent, which the data seems to be following the same seasonal patterns.

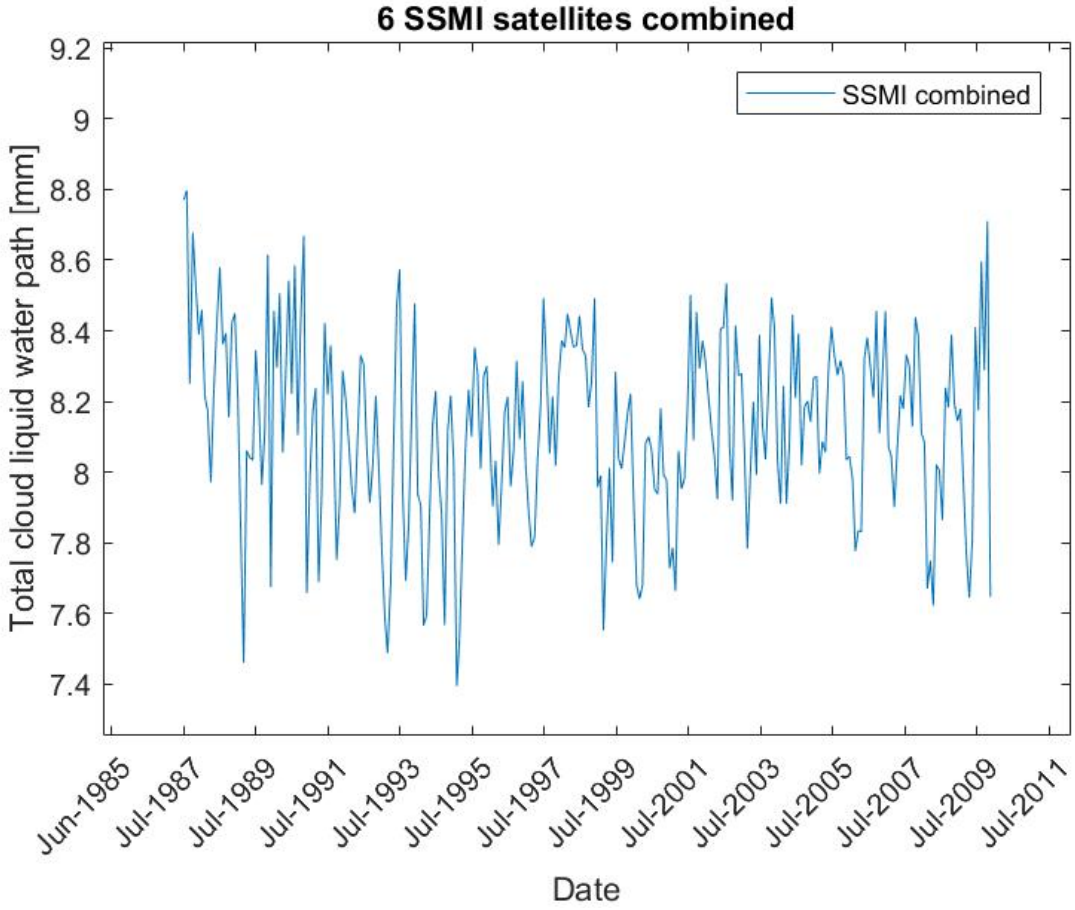


Figure 36: SSMI 0.25 x 0.25 degree, original data from six satelittes combined into one averaged plot. The averaged data shows large variations between the maximum and minimum value of about 16%

7.4 Galactic cosmic rays & sunspot activity

In this section we wish to analyze the possibility of a connection between Galactic Cosmic Rays (GCR) hitting the Earth's atmosphere and the effect on clouds arising from the ionization. The theory is that the ionized molecules can form a "seed" for clouds to be formed (or water droplets). These cosmic rays consists of protons mostly (H), and approximately 10% He [23]. This swarm of particles (GCR) hit the Earth constantly.The solar activity follows an

11-year cycle. When the intensity of the solar activity is high, the Heliosphere shielding of the solar system from GCR's increases (As the solar activity increases, more CGR's changes diretion (or gets blocked) by the Heliosphere).
When these GCR's hit the atmosphere of the Earth the lower energy particles get absorbed or refracted by the magnetic field. The higher energy charged particles ($\sim$ 1000 MeV and higher) generate new secondary particles as they interact in the atmosphere. These new particles travel further through the lower parts of the atmosphere causing these parts to become ionized. The intensity of this ionization reaches a maximum at an altitude of 15-30 km [23] depending on latitude and solar activity level.
This ionization takes place in the whole atmosphere with different intensities. Ilya G. USOSKIN et al 2009. [24] discusses this ionization in the atmosphere further stating the main energy loss takes place at altitudes below 30 km resulting in ionization, dissociation and excitation of molecules.
This ionization should take place in the whole atmosphere. If we were to observe some kind of relationship between the GCR's and the cloud formation pattern, we would expect to find a correlation between the cloud amounts in the whole atmosphere. The patterns of the solar cycle and the cloud amount is expected to show an anti-correlation, as less ionization should take place when the solar activity increases (and the Heliosphere shields more GCR's). Furthermore some altitudes may be more suited for the forming of aerosoles and thereby cloud forming

The most common method of measuring these cosmic rays is by using a neutron monitor. In this method steel tubes are filled with a gas called boron trifluoride [25]. The secondary particles then penetrate the steel casing. After penetrating, the high energy neutron gets absorbed by the Boron atom causing it to split into a helium ion and a lithium ion. These ions then strip electrons from other neutral atoms, causing a charge in the tube to be present.

Other detection methods could be the Pierre-Auger station in south America, which is a little more uncommon. The detection of GCR's happens in two parts. The first detection system is to station large water tanks which detects detects the GCR's [26]. The second detection system observes a faint glow caused by the collision of the GCR's with air molecules a collection of light sensors is placed pointing towards the sky observing these traces of light.

The GCR data used in this section is collected by OULU [27] and the sunspot activity data is measured by The Royal Observatory of Belgium [28]

7.4.1 ISCCP

In figures 37 and 38 the cosmic neutron count and the ISCCP total cloud amount anomaly data is shown.

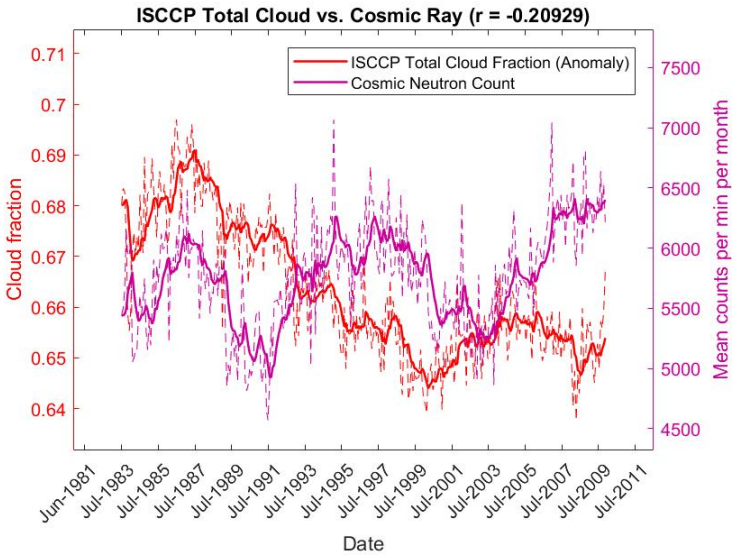


Figure 37: ISCCP total cloud coverage vs cosmic rays. The period 1983-1989 seems to shows clear signs of a relationship between cosmic rays and the total cloud amount. After 1989 the total cloud coverage deviates from the cosmic rays measured resulting in a overall correlation of -0.2 (not correlated at all).

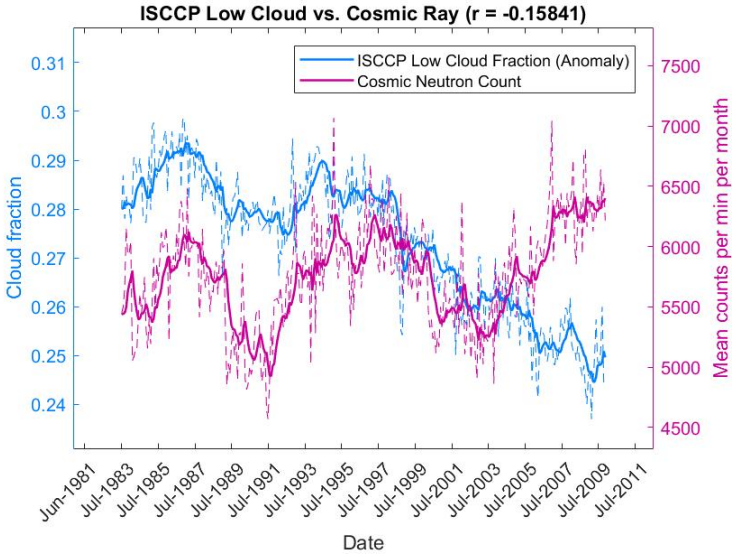


Figure 38: ISCCP low level clouds vs cosmic rays. The low cloud coverage seems to follow the pattern of the cosmic rays from 1983-1997, after this point the low cloud amount measured by ISCCP drops more or less linearly. The overall correlation of the two plots is at -0.15 indicating no relationship between the two series.

The most relevant of the two plots is expected to be the Total cloud coverage because we are interested in a possible ionization of the whole atmosphere (below 30km altitude). Although Marsh & Svensmark [29] found the most obvious link between low clouds and GCR's, in a selected data time range. The total cloud coverage plotted against the cosmic ray data in figure 37 shows strong signs of being almost completely independent with a correlation of -0.2. This is far from expected, as we were expecting to see some degree of similarity in the two graphs. In the beginning of the observation period (1983-1989) the cosmic ray dat and the total cloud coverage shows strong similarities (correlation) in their trend lines. With the exponential smoothing (the solid red line), this period shows a correlation of 0.707. In figure 38 the first period is, as with the total cloud coverage, well correlated in the first part of the period but drastically differs in the later observations. In section 7.1 we discovered a downwards trend in the low cloud coverage from 1998-2009.

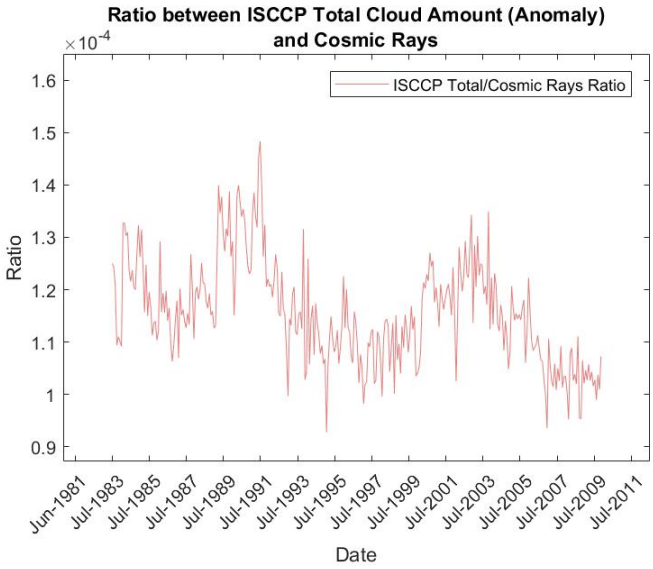


Figure 39: Ratio plot between ISCCP total cloud amount and Cosmic Rays. We see a strong variation in the ratio of the two data. If the two variables were to be perfectly correlated, the ratio plot would be constant in time. The ratio in this plot is certainly not constant in time.

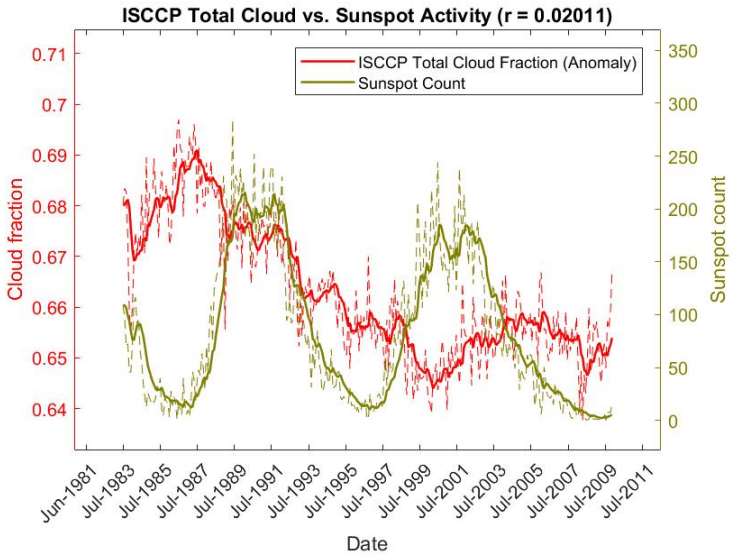


Figure 40: ISCCP total cloud coverage vs sunspot activity. In the first part of the period (1983-1989) we see signs of anti-correlation between the ISCCP total cloud amount and the sunspot activity. After 1989 the total cloud amount and the sunspot activity doesn't seem to be dependent of each other resulting in an overall correlation of 0.02.

The relationship between the sunspot activity and the ISCCP total cloud amount can be seen in figure 40. We would expect to find a strong anti-correlation. This anti-correlation is found in the early observations from 1983-1989 just as we found for the cosmic ray data vs ISCCP. After 1989 the data are more or less independent resulting in an overall correlation coefficient of 0.02 over the whole time period seen in figure (a).

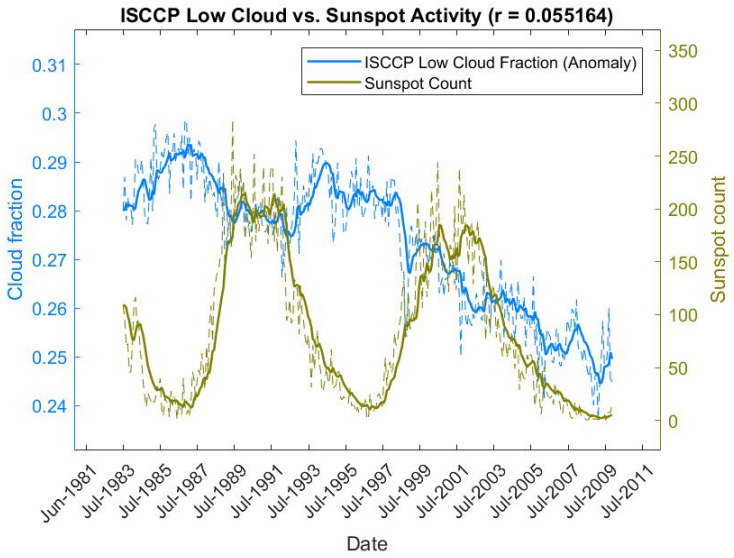


Figure 41: ISCCP low cloud coverage vs sunspot activity. With the low cloud amount we see signs of a anti-correlation between the cloud amount from 1983-1998. The later part seems independent. The ionization should take part in the whole atmosphere and not only in the lower parts of the atmosphere.

7.4.1.1 Regressional corrected low cloud data vs GCR

The ISCCP anomaly data of the total cloud amount and the low cloud amount didn't show signs of a correlation with the GCR's in the last part of the observational period.
As discussed in section 7.1 the mid level cloud amount showed an increase from 1998-2009, where the low level cloud amount decreased. A regressional correction was done removing these opposing trends, stabilizing the low and mid level clouds amounts in this 26 year period. This regression is now applied to the low cloud amount and compared to the GCR data in this period. Figure 42 shows this corrected data. The corrected data show many of the same tendencies as the GCR count. With this correction we might see some connection between the low cloud amount and the GCR's.

It is very important to stress that we do not know whether this drop in the amount of low clouds (before the correction) is caused by different (or faulty) measurements by the ISCCP instruments, or if this downwards trend actually is correctly measured, and the cloud coverage of the Earth is showing signs of a long term trend yet unknown.
The two graphs on figure 42 shows a correlation of 0.3444.This gives us an indication that there is a possibility that this ionization might be influencing the cloud amount in the lower parts of the Earth's atmosphere. The blue plot shows larger fluctuations than the purple plot indicating that the cloud coverage probably is influenced by other factors alongside the amount of cosmic rays ionizing the atmosphere.

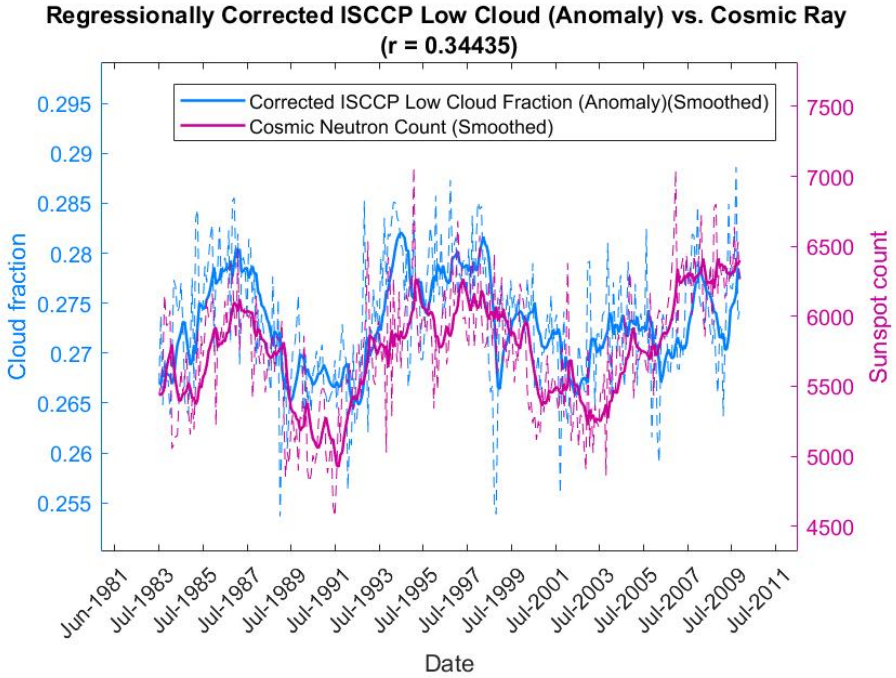


Figure 42: Anomaly of the ISCCP regressionally corrected low cloud coverage data vs GCR. With the corrected data we see a much better correlation between the low cloud amount and the cosmic rays over the whole period. The correlation coefficient is still low (0.34) but this indicates a probability of a connection between these two.

It is interesting that we see stronger signs of a correlation using the ISCCP low cloud amount, than using the ISCCP total cloud amount, as we expected this ionization to occur in the whole atmosphere (with the highest intensity in the high parts of the atmosphere). This should definitely be studied further in the future.

7.4.2 MODIS

The ISCCP data showed no clear indication of a relationship between the cloud coverage and the cosmic rays in the last part of the observational period.
MODIS provides measurements from 2000-2017, these are plotted against the cosmic rays in figure 43. As discovered in section 7.2 a large increase in the amount of clouds was measured due to a stronger El Niño than usual. This large increase is seen in the figure, the cosmic ray data does not show any signs of having any correlation to this event. The overall relationship of the two time series doesn't seem to have any strong relationship on each other, resulting in a, r coefficient of 0.1.

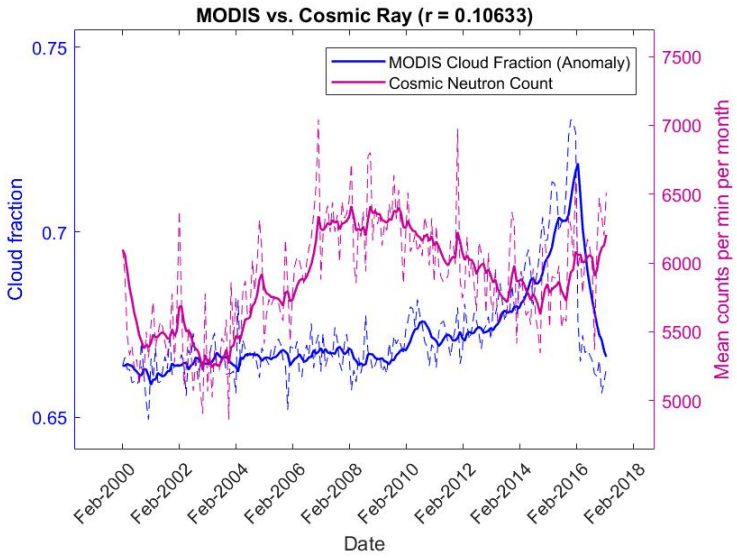


Figure 43: MODIS total cloud coverage vs cosmic rays. The MODIS data shows none of the patterns observed in the cosmic ray data. The overall correlation factor is at 0.11 confirming no dependence.

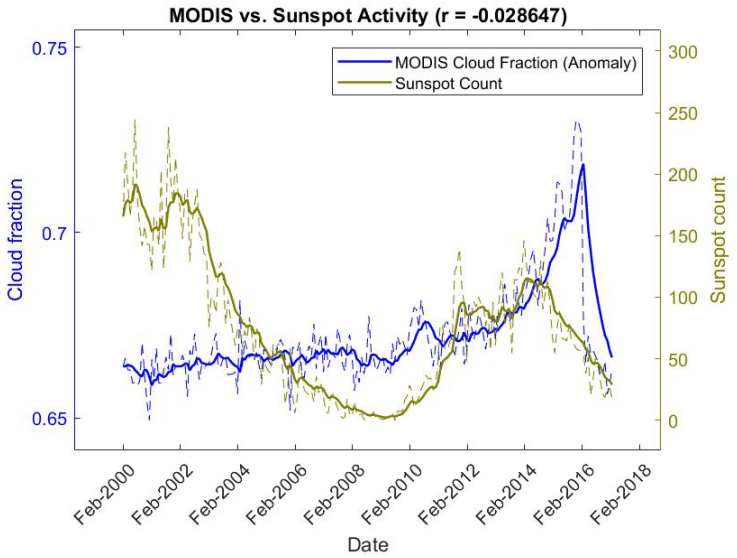


Figure 44: MODIS total cloud coverage vs sunspot activity. No signs of an anti-correlation. The MODIS data increases slightly in the first 14 years, with a big increase at the end of the period.

The overall time series with both ISCCP total cloud coverage and MODIS is displayed in figure 45. In section 7.5 the overlapping period for the ISCCP and MODIS data is investigated.

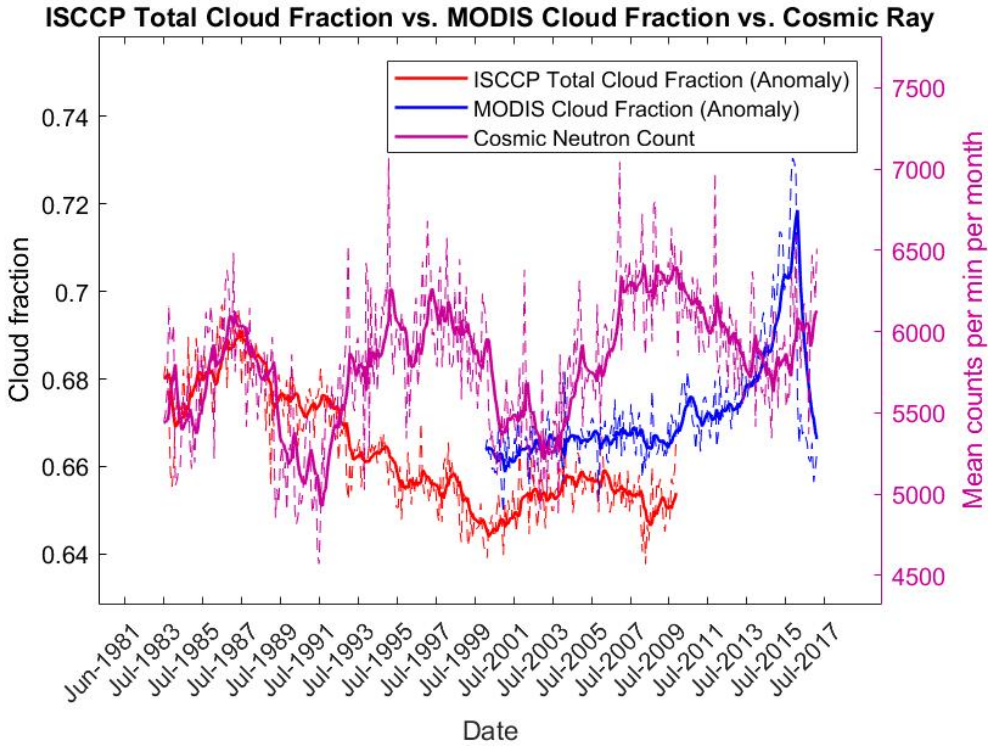


Figure 45: ISCCP total cloud coverage and MODIS total cloud coverage vs cosmic rays. The first part of the ISCCP data follows the cosmic ray data. Over longer time the cloud coverage doesn't seem to follow the cosmic ray data.

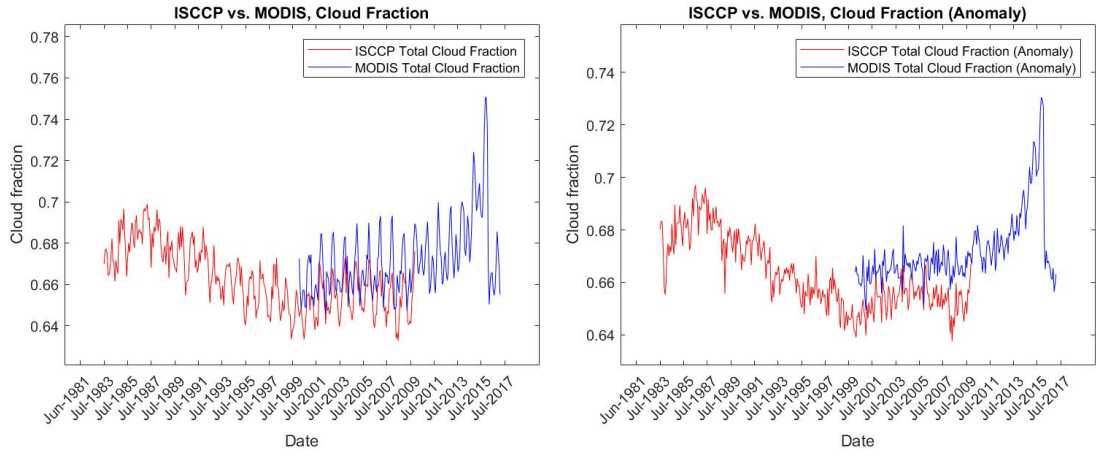
7.5 Comparison of data

In sections 7.1 - 7.3 we've analyzed the 4 types of data individually (ISCPP, MODIS, SSMI 2.5x2.5 & SSMI 0.25x0.25). In this section similarities and differences will be shown, ending up in an overall assessment of how comparable the data is and how much the observations deviate from data set to data set.

MODIS measures the total cloud amount and is therefore directly comparable to the ISCCP total cloud amount.
SSMI measures the liquid water path and should in theory be directly comparable to ISCCP low cloud amount measurements in areas containing solely low level clouds.
Since MODIS doesn't single out the low level cloud amount but rather measures the total cloud amount, a comparison between MODIS and SSMI would not be that useful.

7.5.1 ISCCP vs MODIS

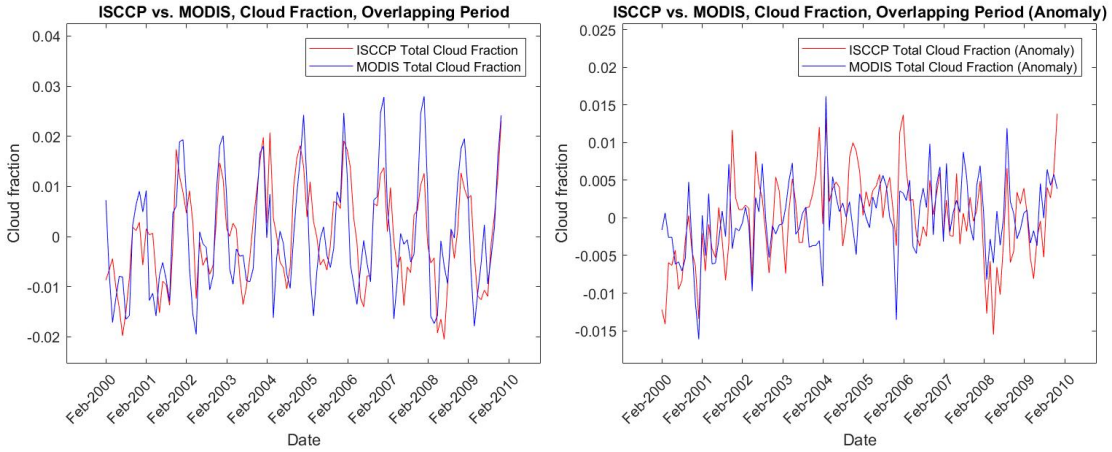
The overall time series for the two sets of data is displayed in figure 46a and with the seasonal variations removed in figure 46b. In the overlapping period the data looks well correlated showing most of the same tendencies. Just after the large spike in the anomaly plot in 2004 we see some differences in the the two graphs. These differences in the period 2004-2007 are $\sim 3\%$. The MODIS data increases where ISCCP decreases, indicating the two types of data might be sensitive to different aspects of the cloud coverage. Further differences are hard to recognize in this plot. In figure 47a only the overlapping data is plotted. Large deviations can be seen just after each yearly spike in both data sets. The MODIS data increases in amount reaching a lower peak at around half the peak of the yearly variations. This indicates that MODIS observes a variation every 6 months in addition to a yearly variation in the cloud coverage. ISCCP doesn't seem to pick up this 6 month seasonal pattern. ISCCP does show a slight increase between most of the seasonal peaks but they are far less pronounced in contrast to MODIS. This missing 6 month variation results in a correlation coefficient of 0.6784 between the two time series. In this plot the differences in minima in each yearly cycle are eye catching. The measurements made by ISCCP in the middle of the yearly cycle (the 6 month seasonal pattern) is almost anti-correlated with the mid-cycle MODIS observations. This is very concerning and results in a low correlation coefficient. Figure 47b shows the anomaly of the two time series, confirming the issues mentioned above, with large deviations in the anomalies measured by the two satellite systems. The correlation of the two time series after the seasonal variations are removed is at 0.5039. Not what we would hope to find should we want to compare future measurements. Overall the data shows the same yearly peak cloud amount, other than this seasonal peak the data shows huge differences making comparisons of detailed events very difficult.



(a) Unprocessed data

(b) Anomaly

Figure 46: (a) ISCCP vs MODIS total cloud coverage. The data look consistent in the overlapping time period with different amplitudes. (b) ISCCP vs MODIS anomaly of the total cloud coverage. Again the same trends are seen in the two data set.



(a) Unprocessed data (b) Anomaly

Figure 47: (a) ISCCP vs MODIS overlapping data. The differences in seasonal variations are eye catching. ISCCP shows a very weak (if any) half year seasonal variation compared to the half year variation observed by MODIS. The correlation between the two plots are at 0.6784. (b) ISCCP vs MODIS overlapping data, seasonal variations removed. In this anomaly plot the differences becomes very clear. The two satellite systems' anomaly are in many places directly opposite. The correlation of the anomaly data in the overlapping period is 0.5039 meaning that a feature in the blue line is only with 50% probability explained by a variation in the red line (or the other way round). This indicates that they measure different events in the cloud coverage of Earth.

7.5.2 ISCCP vs SSMI

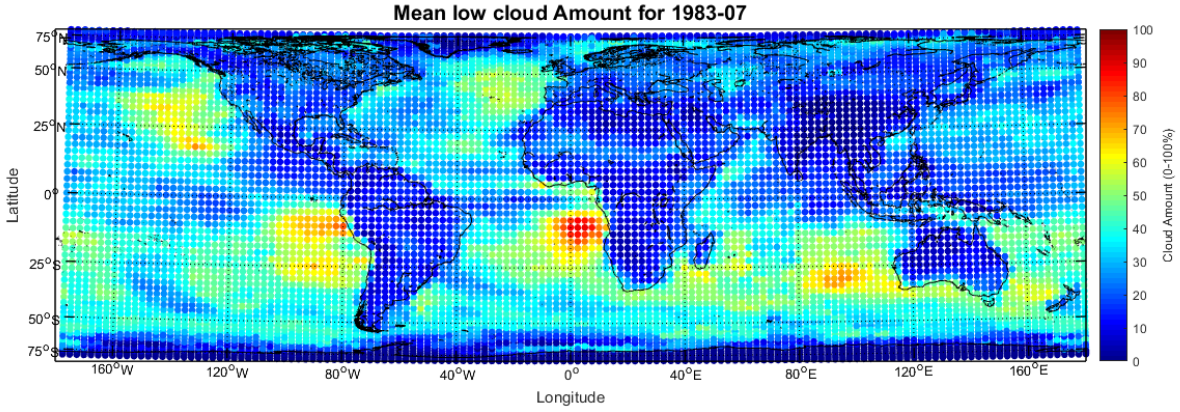


Figure 48: Illustration of the ISCCP low cloud amount in July 1983. A significant percentage of low level clouds are observed west of latin america and west of Angola, Africa. These two areas are chosen to compare ISCCP low clouds vs SSMI.

In this section we zoom in to two areas known for having clouds located almost solely at low altitudes. Using such areas should make a comparison between ISCCP [mean low cloud amount] and SSMI [liquid water path] possible. Because the temperatures in this lower altitude provides for liquid water, every cloud measured by the ISCCP should be directly comparable to the SSMI readings.

7.5.2.1 Coast of Angola

In figure 48, west of Angola, Africa, shows a very large amount of low clouds. A 20 x 20 degree area is chosen with coordinates ranging from 25°S – 5°S, 10°W–10°E. This area is chosen because it is one of the most low cloud populated areas in the world, reaching low clouds amounts upwards of 70%, reaching its peak in October measured by ISCCP.

Figure 50 shows the ISCCP low cloud amount data and the SSMI 2.5 x 2.5 degree data for this specific area off the coast of Angola. In this area we see data well correlated, showing the same seasonalities with some differences in amplitudes. The overall correlation for the whole period is 0.8286 for the original data. In the correlation calculations, the ISCCP data has been removed for the 18 months period where the SSMI observations are missing. The reason the correlation is not higher than 0.8286 could be because the area might also contain some middle- and high level clouds.

Figure 51 shows the sum of the mid- and high level cloud amounts in the period. From this plot it is clear that the chosen area contains other types of clouds than low level clouds.

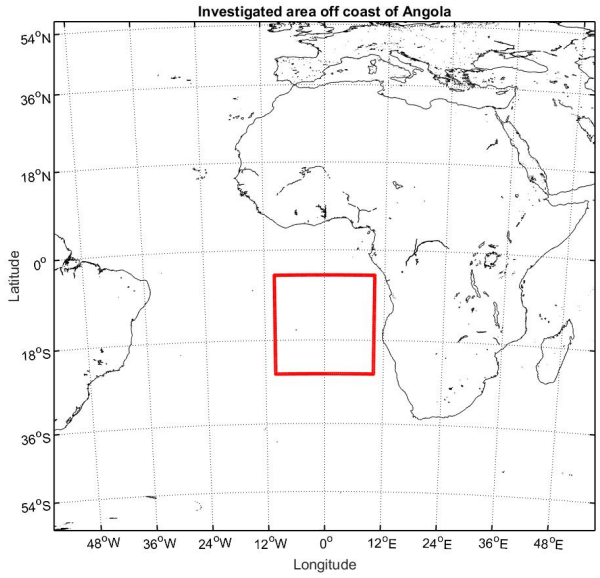


Figure 49: Area off the coast of Angola chosen for further studies.

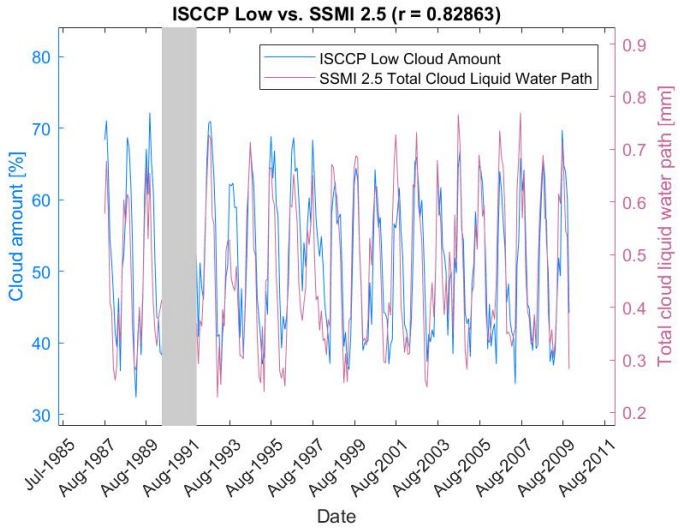


Figure 50: ISCCP low cloud amount vs SSMI 2.5 x 2.5 degree data. The two sets of data shows clear correlated patterns resulting in a overall correlation of 0.83. The correlation might be even better if the area comprised of low clouds, and absent of mid- and high level clouds.

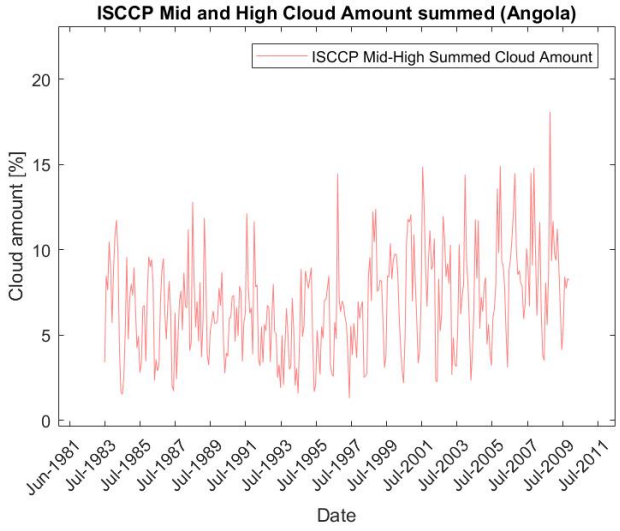


Figure 51: Sum of mid- and high level clouds. The plot shows that the mid- and high level cloud amounts upwards of 7% averaged. The sum of the mid- and high level cloud seasonalities are anti-correlated with the low cloud amount.

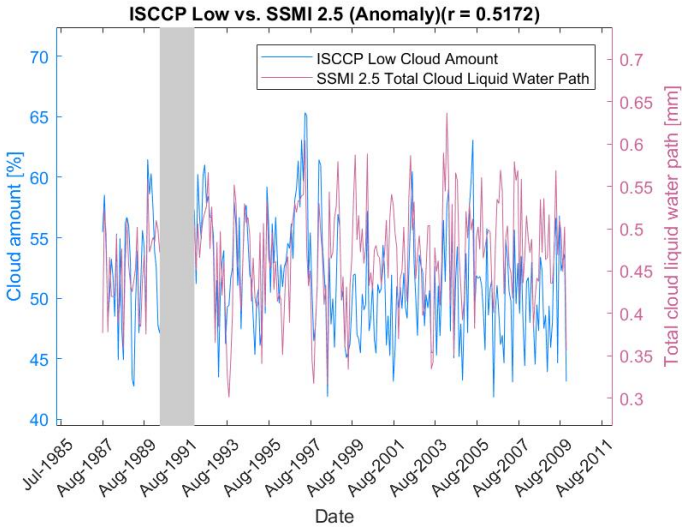


Figure 52: ISCCP low cloud amount vs SSMI 2.5 x 2.5 degree, anomaly plot for the Africa Zone. We see large differences in the measured anomalies in the two data set, especially in the last half of the period where the two time series deviate from each other. This result in a correlation of 0.52. The differences in anomalies might be less precise in the last half of the period because the sum of the mid and high level cloud amount is larger in this period compared with the first half of the period.

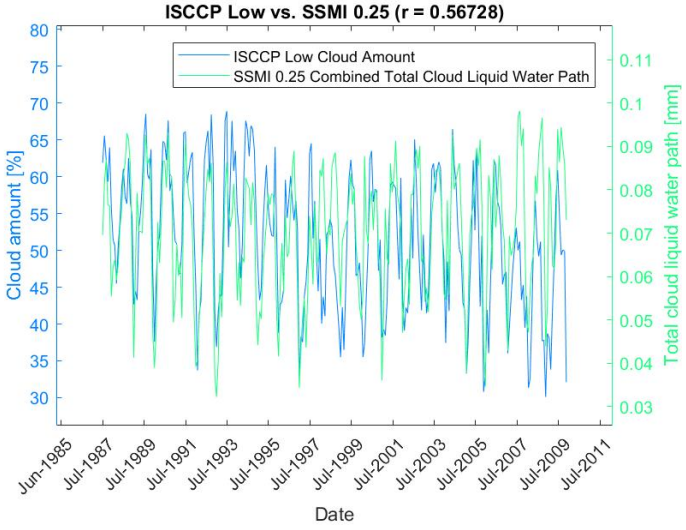


Figure 53: ISCCP low cloud amount vs SSMI 0.25 x 0.25 degree data for the 6 satellites combined. This plot shows a correlation of 0.88 between the two sets of data. This correlation is higher than the SSMI 2.5 data (correlation 0.83).

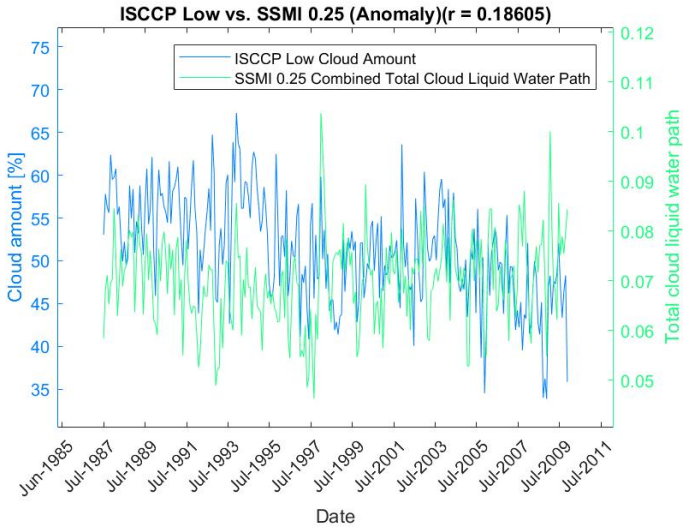


Figure 54: Anomaly plot of ISCCP low cloud amount vs SSMI 0.25 x 0.25 degree data for the 6 satellites combined. The anomalies of the two data sets deviate especially at the end of the observation period just as for the 2.5 x 2.5 degree data. The correlation of 0.59 is slightly higher than the 2.5 x 2.5 degree data (0.52) but this correlation coefficient is still weak.

7.5.2.2 Coast of Peru

The area off the coast of Peru 25°S – 5°S, 100°W-80°W contains a smaller portion of low level clouds seen in figure 48 (and less low level clouds than the area off the coast of Angola.). The analysis is done for both the SSMI 2.5 x 2.5 degree data and the SSMI 0.25 x 0.25 degree data. The 2.5 x 2.5 degree data is in this area better correlated with the ISCCP low cloud amount in contrast to the Angola area where the combined graph for the 6 SSMI 0.25 x 0.25 satellites showed the best correlation.
The combination of the 6 satellites showed in this area a correlation of 0.5673 (large uncertainties of the data), the anomaly of the 0.25 x 0.25 data has a correlation factor of 0.1861.
This further indicates that a comparison between ISCCP and SSMI are tough as the percentage of low level clouds in the area of investigation drops. There might only place on Earth that a comparison between the ISCCP low cloud amount and the SSMI water path is possible. This area is the area off the coast of Angola where most low level clouds are found.

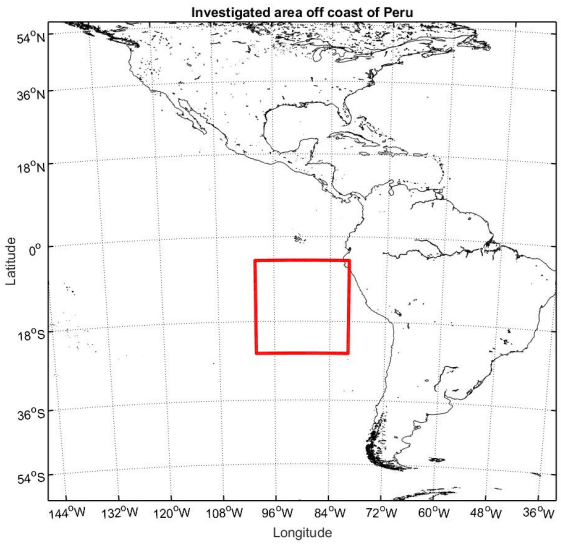


Figure 55: Area off the coast of Peru chosen for further studies.

In figure 56 the summed mid- and high cloud amounts are showed. These amounts are increased significantly compared with the Angola area.

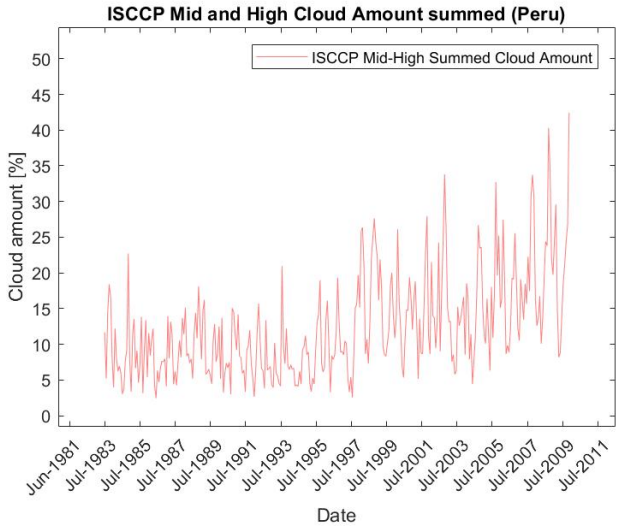


Figure 56: Sum of mid- and high level clouds. The plot shows that the mid- and high level cloud amounts up to of 40% in contrast to the African region where the similiar number was 15%. In the area off the coast of Peru we find a lot more mid- and high level clouds than off the coast of Angola. This increase in mid & high clouds is expected to worsen the comparison between ISCCP low cloud amount and SSMI 2.5 x 2.5 & SSMI 0.25 x 0.25

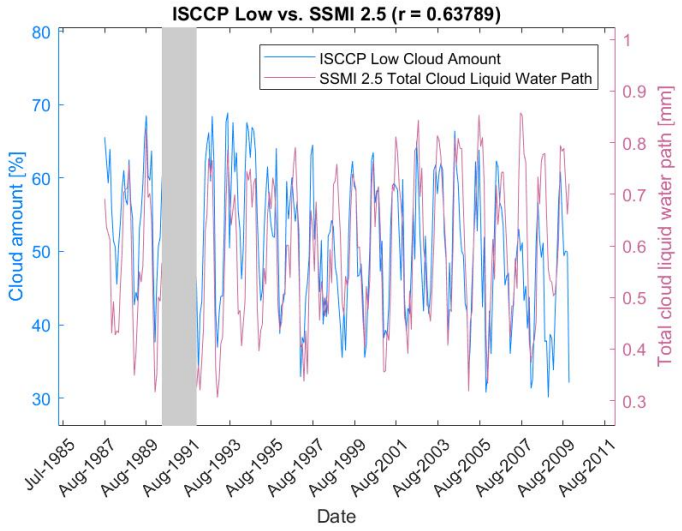


Figure 57: ISCCP low cloud amount vs SSMI 2.5 x 2.5 degree data Peru area. Compared to the Angola area the correlation is worsened significantly. It is in the Peru area 0.6379 (0.8286 in the Angola area). This indicates that that the ISCCP low cloud amount and the SSMI data differs significantly as the mid and high cloud amounts increase.

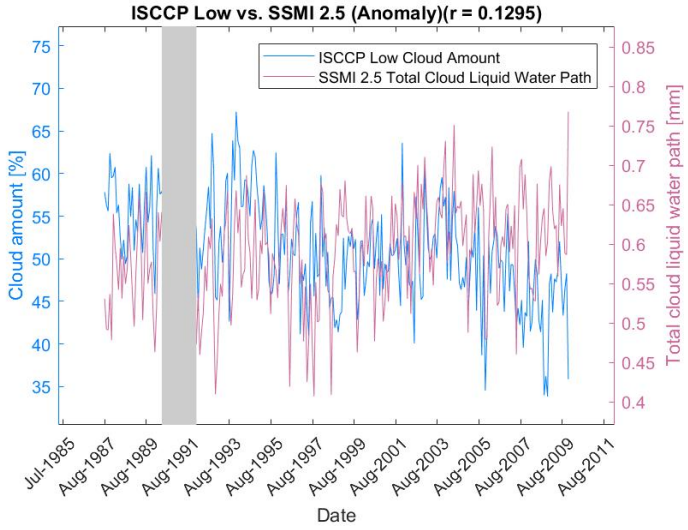


Figure 58: ISCCP low cloud amount vs SSMI 2.5 x 2.5 degree, anomaly plot. The anomalies in the area off the coast of Peru are more or less non-correlated with a correlation coefficient of only 0.1295.

8 Summary and Discussion

This section provides a general overview and discussion of the satellite systems based on the general theory and the analysis of these systems.

Inconsistent data reoccur throughout all of the data and there is a clear tendency of low correlation coefficients when comparing the data sets to each other. Calibration of the satellites could be causing these differences in data. Various other factors including the difference in instruments, viewing angles, cloud layering and analysis algorithms will also contribute towards a lower correlation coefficient between the data sets.
Obvious reasons are the spectrometry methods of the individual systems. ISCCP which is working at a wavelength of 0.6 and 11 microns is measuring in both the VIS and IR spectrum, but is limited to these wavelengths. This may result in some of the results having a different interpretation or even a misinterpretation compared to results from the other systems. MODIS is working in 36 different spectral bands from $\lambda = 0.4$ to 14.4 microns. By covering the whole IR spectrum with various bandwidths we would expect MODIS to interpret data with a better accuracy than ISCCP and therefore supply imagery which is more detailed than VIS and IR images at only 2 wavelengths. Data from these instruments are in good agreement and the period where the ISCCP and MODIS data overlap. As ISCCP satellittes are both geostationary and polar orbiting and are supernumerary compared to the two MODIS satellites, coverage should be better with ISCCP. As these systems are operating on VIS and rely heavily on the visible optical thickness of the observed clouds, viewing angle could be another reason for uncertainties. Geostationary satellites have the same view of the Earth at all times, which causes a constant viewing angle on locations. A higher viewing angle leads to more cloud observed as the path from the satellite to the surface is longer. A consequence of this could be that the readings at points with bigger viewing angles are higher.
As the measurements are made vertically or with a slight angle, another problem causing uncertainties could be cloud layering or cloud clumping. As seen in section 7.1.2 the low cloud amount decreases as mid cloud amount experiences a slight increase, this could be caused by impairing the vision of the low clouds due to the mid clouds reducing their visibility. If low cloud readings experience a decrease following an increase of mid level clouds, this would mean loss of information on low clouds and therefore a lower cloud amount.
Differences in algorithms will make a difference in results. As ISCCP and MODIS use different algorithms to calculate cloud fractions and amounts, different factors in the algorithms also provides an uncertainty when comparing.For example; different accepted values for thresholds when classifying pixels as cloudy or clear could mean that some pixels classified as cloudy in one system, could be marginally cloudy or clear in the other system. The threshold differences could be included to get more clear results.

When comparing the VIS and IR systems to SSMI and SSMIS, we are not looking at a direct comparison in terms of cloud fraction and cloud amount, but instead liquid water path or cloud liquid water. This means different readings in the data sets, but is used to give a general description of the tendencies of the cloud cover. As the SSMI readings are made above water and not close to land (no measurements if land or sea ice is in 25-50 km range of the location for which measurements are made), a total cloud amount would not be comparable to that from ISCCP and MODIS, which is why the SSMI comparisons and analysis are made for areas above oceans. SSMI's algorithm is not made for calculating the cloud water amount when rain is present which also adds uncertainties when comparing.

Additionally, the spatial resolutions of the systems can provide an uncertainty. Optimally a

satellite system should have both polar orbiters and geostationary satellites. This is only the matter for ISCCP. MODIS and SSMI consist of only polar orbiters. The polar orbiter systems provide full coverage of the Earth, but their high spatial resolutions compared to their low temporal resolution could lead to missing areas or data points. Although ISCCP is the system with the best coverage from the satellite systems compared in this thesis, ISCCP does not fulfill the demands for being an optimal system. More geostationary ISCCP satellites on the southern and northern hemisphere might lead to better measurements followed by the better coverage.

In the analysis of the data (section 7) we have seen how inconsistent measurements of different satellites are compared to each other.

6 SSMI satellites were investigated in section 7.3.2 (from the high resolution SSMI data). Each of these satellites were expected to show a very high correlation between one another as they measure the same thing. This was rather surprisingly not the case. In table 5 the correlation coefficients of the satellites in the overlapping periods were found. 3 of the overlapping satellite periods showed a correlation coefficient of 0.89 or higher, which is lower than we would expect. As the satellites were measuring the same parameters we would have epected (and hoped for) correlations coefficients close to 1 (close to perfectly correlated). The other 5 overlapping satellite periods showed correlation coefficients of 0.86 or lower, with 2 of these periods significantly lower at 0.70 (12 months of overlap) and 0.76 (93 months of overlap). Correlation coefficients this low (under 0.8) raises questions to how successful the calibration of the satellites have been. It is definitely worth noting that all of these correlation coefficients suggest that each individual satellite observes each event differently, increasing the uncertainties with the corresponding measurements significantly.
The low resolution data also showed large variations with fluctuations up to 17% from minimum to maximum. This is very large fluctuations, and they make a comparison to another satellite system very difficult as the combined uncertainty between these system will increase quadratically (if the errors are uncorrelated).

In section 7.5.2 the low and high resolution data from SSMI were compared to the ISCCP low cloud amount in two well-considered areas chosen for their large amount of low level clouds (compared to mid- and high level clouds). The area showing the best results were found off the coast of Angola. In this area we saw a correlation coefficient of 0.88 for the high resolution SSMI data and the ISCCP low cloud amount. This is the area on Earth best suited for a comparison between SSMI and ISCCP and even in this area we found the two data sets deviating significantly from each other. The low resolution SSMI data showed a slightly worse correlation, 0.83. In the other area (off the coast of Peru) the low resolution data rather surprisingly (as this was not the case in the Angola area) showed a correlation of only 0.64. This was the area found second best for a possible comparison between ISCCP and SSMI. To find a correlation coefficient this low in the second best suited area is rather problematic as it indicates that this method of comparison might only be valid for one 20° x 20 ° square on the Earths surface.

In section 7.4 we found a possible weak relationship between the galactic cosmic rays and the ISCCP low cloud amount data after having done a regressional correction to the low cloud data. If the downwards trend in the ISCCP low cloud amount is measured incorrectly, this data with the correction added, could indicate how the low cloud amount in the period 1983-2009 really have been, if the cloud coverage had been measured correctly. This is

speculation though, with the data available at the time, we do not know whether or not this downwards trend (in low clouds) is true. A long term cycle (longer than the 26 years of ISCCP data already available) altering the relationship between low and mid level clouds is very possible but remains unseen. Future observations will most likely answer this question.

In section 7.5.1 the total cloud fractions for ISCCP and MODIS were compared. In the overlapping 9 year period of the two sets of data we saw indications of MODIS observing a semi-strong 6 months- along with a strong 12 months seasonal variations. ISCCP showed only a strong 12 month seasonal variation. This resulted in a correlation coefficient of only 0.6784 between the two data set. This is rather surprising as this comparison was thought to be the most obvious of all the satellite systems. This furthermore substantiates the claim that the measurement of the Earths cloud coverage is connected with very large uncertainties at the present time, and that the methods of measuring these data might be in need of better calibrations from satellite to satellite, or even in need of a whole new method of defining how the instruments determines that a cloud is present in its line of sight.

9 Conclusion

Throughout the analysis we have experienced some major differences in the measurements from the different satellite systems. In one specific case, where we zoomed in on the most low cloud dense region in world, the ISCCP low cloud amount and the SSMI (both 2.5 x 2.5 % the 0.25 x 0.25 degree resolution) data showed signs of consistency. This resulted in an overall correlation of 0.88 for the 0.25 x 0.25 data. This correlation is one of the higher found throughout the project. In the second most dense region found (in amount of low clouds), this correlation was significantly lower showing a correlation of 0.64 for the 2.5 x 2.5 data.
These low correlation coefficients found outside the most low cloud dense area is very troubling. They indicate, that this method of comparison (ISCCP low cloud amount vs SSMI liquid water path) is only well-comparable in one region on Earth (because mid- and high level clouds needs to be kept at a very low level, for this type of comparison to show good results).
In future studies other variables measured by SSMI could show better result than obtained throughout this report. Here we chose the "CLOUD" variable [liquid water path] as it was expected to be the most likely variable found for a comparison.

The ISCCP vs MODIS comparison showed an overall correlation coefficient of 0.68 in the period of overlapping data (2000-2009). This comparison between ISCCP total cloud amount and MODIS cloud fraction was expected to be the two best comparable data sets. This low correlation makes the comparison rather inconclusive.

The main goal of this work was to compare these satellite system, to hopefully conclude whether or not the trend change for the ISCCP low & mid cloud amounts were likely to be caused by different/faulty measurements, or if, these trend changes were likely to be caused by a long term seasonal variation in the Earths cloud coverage.
Due to the large variations from satellite system to satellite system, and the low correlation coefficients found when comparable these systems, it is hard to answer whether or not this downward trend for the low cloud amount & upwards trend for the mid level cloud amount is caused by a yet unknown trend in the cloud coverage of the Earth.

The regression found and applied to the low/mid level cloud amounts showed indications of a possible weak correlation between the galactic cosmic rays and the amount of low clouds on Earth. With the regression applied we found a correlation of 0.34. A correlation of 0.34 is definitely not high enough to conclude a relationship present, but the possibility of a connection is there.
Using the unprocessed ISCCP total cloud amount data we found no clear indications of a correlation. In the first couple of years of observations the trend lines of the total cloud amount and the cosmic rays looked somewhat similar, in the latter 20-25 years of observations no real indications of a relationship was discovered. This result was surprising as we would expect the ionization, creating "seeds" for clouds to form, to take place in the whole atmosphere and not only in the lower part of the atmosphere.
The MODIS data showed no signs of a correlation with the galactic cosmic rays.

In future work to figure out whether this downwards trend in the low cloud amount is true, other satellite system could be included in the comparisons. From what we've shown here, data from different satellite systems show significant differences when compared to each other. It would definitely be interesting to see if more satellite systems show these deviations. If that should be the case a rather extensive upgrade to the satellite systems might be required in order to get more usable (and comparable) data.

Furthermore, it would be very interesting to compare the ISCCP low cloud amount to another satellite measuring the low level clouds worldwide. SSMI was in this project intended as a possible low cloud comparison system, but the SSMI data simply showed to be too unreliable for a larger scale comparison.

ISCCP is still active by doing observations. In this project data from 1983-2009 were used but newer data exist although not (yet) published by ISCCP. When this data is released it will certainly be interesting to see if the low cloud amount rises to the earlier "stable" level we saw from 1983-1998 at around 27-28% low clouds.
If the low cloud amount arises to this level again this might be an indication that a long term cycle in the Earths cloud coverage exist.
Furthermore, a more detailed comparison between MODIS and ISCCP will be possible when this data is released, and it will be interesting to see if ISCCP also shows the same massive increase in total cloud amount as MODIS shows during the El Niño in the period 2014-2016.

10 References

[1] London, J. and T. Sasamori
Radiative energy budget of the atmosphere
`Space Research XI 639-649 (1970)`

[2] Hansen, J., Rossow, W. and Fung, I.
Long-Term Monitoring of global Climate Forcings and Feedback
`NASA Conference Publication 3234`

[3] Zhang, Y.-C., Rossow, W. B.m Lacis, A. A.
Calculation of Surface and Top-of-Atmosphere Radiative Fluxes from Physical Quantities Based on ISCCP Datasets, Part I: Method and Sensitivity to Input Data Uncertainties
`J. Geophys. Res. 100, 1149-1165`

[4] Stephen H. Schneider:
Cloudiness as a Global Climatic Feedback Mechanism: The Effects on the Radiation Balance ans Surface Temperature of Variations in Cloudiness
`Institute for Space Studies, Goddard Space Flight Center, NASA.`

[5] ISCCP overview
NASA
`https://isccp.giss.nasa.gov/overview.html`

[6] ISCCP Project description
NOAA
`https://www.ncdc.noaa.gov/isccp`

[7] Satellite Meteorology: An Introduction
Stanley Q. Kidder, Thomas H. Vonder Haar, Stanley H. Vonder Haar
`Academic Press, 1995, p. 173-177`

[8] ISCCP New algorithm
NASA
`https://isccp.giss.nasa.gov/newalg.html`

[9] Mesoscale Variability of Cloud Optical Thickness
NASA
`https://isccp.giss.nasa.gov/climanal6.html`

[10] Mesoscale Variability of Cloud Optical Thickness
NASA
`https://isccp.giss.nasa.gov/climanal7.html`

[11] Hahn, C.J., W.B. Rossow and S.G. Warren, 2001:
ISCCP cloud properties associated with standard cloud types identified in individual surface observations.
`J. Climate, 14, 11-28`

[12] National Oceanic And Atmospheric Administration, National Environmental Satellite Data And information Service, National Climatic Data Center, Satellite Data Services Division (SDSD)
International Satellite Cloud Climatology Project (ISCCP), Catalog of Data and Products
`Central Archive`

[13] Ted Funk
Cloud Classifications and Characteristics
www.weather.gov/media/lmk/soo/cloudchart.pdf

[14] Remote Sensing Systems
SSMI/SSMIS mission description
http://www.remss.com/missions/ssmi

[15] Frank J. Wentz
A Well Calibrated Ocean Algorithm for SSM/I
Remote Sensing Systems

[16] F. T. Ulaby. R. K. Moore and A. K. Fung.
"Microwave Remote Sensing—Active and Passive". "
(Reading, Massachusetts: Addison-Wesley, 1981 and 1982.) Volume I: Microwave Remote Sensing Fundamentals and Radiometry.

[17] M. Peichl, S.Dill, M. Jirousek and H. Süß
"Microwave Radiometry - Imaging Techologies and Application"
Proceedings of WFMN07, Chemnitz, Germany

[18] Christopher L. Brest & William B. Rossow (1992)
"Radiometric calibration and monitoring of NOAA AVHRR data for ISCCP"
International Journal of Remote Sensing, Volume 13, 1992 - Issue 2, pages 235-273

[19] Frank J. Wentz (2013), Remote Sensing Systems, Santa Rosa, CA
"SSM/I Version-7 Calibration Report", RSS Technical Report 011012, January 2013

[20] Rossow, W. B, and Cairns, B.
Monitoring Changes of Clouds
NASA Goddard Institute for Space Studies

[21] Ward Cheney and David Kincaid (2013), University of Texas at Austin
Numerical Mathematics and Computing, Seventh Edition
Brooks/Cole, Cengage Learning

[22] Jan Null, CCM,
"El Niño and La Niña Years and Intensities"
http://ggweather.com/enso/oni.htm

[23] G.A. Bazilevskaya, I.G. Usoskin, E.O. Flückiger, R.G. Harrison, L. Desorgher, R. Bütikofer, M.B. Krainev, V.S. Makhmutov, Y.I. Stozhkov, A.K. Svirzhevskaya, N.S. Svirzhevsky, G.A. Kovaltsov 22 March 2008, Springer
"Cosmic Ray Induced Ion Production in the Atmosphere"
Science+Business Media B.V

[24] Ilya G. USOSKIN, Laurent DESORGHER, Peter VELINOV, Marisa STORINI, Erwin O. FLÜCKIGER, Rolf BÜTIKOFER, and Gennady A. KOVALTSOV
"Ionization of the Earth's Atmosphere by Solar and Galactic Cosmic Rays"
2009 Institute of Geophysics, Polish Academy of Sciences

[25] Jan Houseman and Alan Fehr
The Inuvik Neutron Monitor
http://neutronm.bartol.udel.edu/listen/, 19/6 - 2015

[26] Pierre Auger Observatory
A Detector Thirty Times the Size of Paris
https://www.auger.org/index.php/cosmic-rays/detection, 15/6 - 2015

[27] Cosmic Ray Station of the University of Oulu / Sodankyla Geophysical Observatory
https://cosmicrays.oulu.fi/, 15/6 - 2015

[28] Royal observatory of Belgium
Sunspot Index and Long-term Solar Observations
http://www.sidc.be/silso/datafiles, 15/6 - 2015

[29] Nigel Marsh and Henrik Svensmark
"Galactic cosmic ray and El Niño–Southern Oscillation trends in International Satellite Cloud Climatology Project D2 low-cloud properties"
Danish Space Research Institute, Copenhagen, Denmark
JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 108, NO. D6, 4195, 2003

Appendices

A: MATLAB functions

Following is a list of the essential function written and used throughout to generate and visualize the data. Since the actual compiling of the data takes place in larger files with repetitive commands, these are excluded from the appendices. These data preparation scripts will be included along with the data. The overall data is extracted and saved for later use in the `Data_Generator.m`-script and includes: Coordinates for data, date list for individual satellites (both as strings as well as a number code hereof), import and arrangement of data and time series (raw data and anomaly). In the `Peru_Analysis.m`- and `Angola_Analysis.m`-scripts, this data is filtered according to the area, saved as well and eventually visualized. At last, the sunspot activity as well as cosmic radiation is loaded and visualized in the `SunSpot_CosmicRays_Analysis.m`. In order to generate dates, import data, generate time series, get the anomalies filter the data etc. the following functions are used, many of them, whose math has been explained in section 6.

CloudImport_XXXXX.m

This is an example of an import function. Since we have 4 individual types of data to be imported, 4 versions in different variations exist: `CloudImport_ISCCP_D2.m`, `CloudImport_SSMI_25.m`, `CloudImport_SSMI_025.m` and `CloudImport_MODIS.m` (the one presented in the example). Since files differ in format from data set to data set, the individual `CloudImport_XXXXX.m`-function will have different methods of loading, but the procedure is basically the same. Takes input in the form of a string with the name of the file containing the monthly data, import of data into a matrix based off the name string, and hereafter the arranging and sorting explained in section 5.2.

```
function output = CloudImport_MODIS(file_name)

data = zeros(1,360*180);

monthdata = csvread(file_name);
monthdata = Area_Mean_MODIS(monthdata);

assigncount = 0;
for j = 1:180
    for k = 1:360
        assigncount = assigncount+1;
        data(assigncount) = monthdata(j,k);
    end
end

output = data;
```

Date_Gen.m

Takes the year and month, where the data collecting started as well as when it ended, and creates a list with every month of the format YYYY-MM

```
function output = Date_Gen(startyear,startmonth,endyear,endmonth)
% startyear  = year first data was recorded
% startmonth = month of year first data was recorded
% endyear    = year last data was recorded
% endmonth   = month of year last data was recorded

monthcount = 0;
for i = startyear:endyear
    for j = 1:12
        monthcount = monthcount+1;
        if j < 10
            Date(monthcount,:) = ['',num2str(i),'-0',num2str(j),''];
        else
            Date(monthcount,:) = ['',num2str(i),'-',num2str(j),''];
        end
    end
end

Date(1:(startmonth-1),:)=[];
if endmonth < 12
    Date(end-(11-endmonth):end,:)=[];
end

output = Date;
```

Example:

```
Date_Gen(1999,10,2000,02)

ans =

1999-10
1999-11
1999-12
2000-01
2000-02
```

Area_Mean_MODIS.m

Takes as input the raw data loaded from the .csv files of size 3600x1800 and takes the weighted mean for each area of 10x10 in order to create an 360x180 matrix as output. This is unique for MODIS, and is done in a similar way as making the time series.

```
function output = Area_Mean_MODIS(Input_Data)
% Input_Data = raw 3600x1800 MODIS data extracted from .csv file

theta_n = @(n) n*pi/180;

Ol_Data = Input_Data;
New_Data = zeros(180,360);

long = 0.05:0.1:359.95;
lat = 0.05:0.1:179.95;

y_count = 1;
for n = 1:180
    y_int = y_count:y_count+9;
    lil_lat  = lat(y_int);
    x_count = 1;

    for m = 1:360
        x_int = x_count:x_count+9;
        lil_long = long(x_int);

        lil_area = Ol_Data(y_int,x_int);

        sum_1 = 0;
        sum_2 = 0;
        for i = 1:length(lil_lat)
            for j = 1:length(lil_long)
                if lil_area(j,i) >= 0 && lil_area(j,i) <= 1
                    sum_1 = sum_1 + lil_area(j,i)*cos(theta_n(lil_lat(i)));
                    sum_2 = sum_2+cos(theta_n(lil_lat(i)));
                end
            end
        end

        New_Data(n,m) = sum_1/sum_2;

        x_count = x_count + 10;
    end
    y_count = y_count + 10;
end

output = New_Data;
```

GenTimeSeries.m

One of the most important functions used throughout data processing. Generates the time series used for analysis.

```
function TS = GenTimeSeries(data,lowlim,upplim,long,lati,longlatistep)
% data = raw sorted data for each month
% lowlim = defined lower limit for data
% upplim = defined upper limit for data
% long = corrosponding longitude array
% lati = corrosponding latitude array
% longlatistep = degree stepsize in related long/lati arrays (res)
%
% TS   = Time series generated through weighted mean

phi_n = @(n) n*pi/180;

% Interval of degrees
nint = min(lati):longlatistep:max(lati); %Latitude
mint = min(long):longlatistep:max(long); %Longitude
mcount = length(mint);

N = size(data,1);
TS = zeros(N,1);

for i = 1:N
    sum   = 0;
    sum_2 = 0;
    count = 0;
    for n = nint
        for m = 1:mcount
            count = count + 1;
            % The data set contains extreme values, we simply skip these
            if data(i,count) >= lowlim && data(i,count) <= upplim && isnumeric(data(i,count))
                sum = sum + data(i,count)*cos(phi_n(n));
                sum_2 = sum_2+cos(phi_n(n));
            end
        end
    end
    TS(i) = sum/sum_2;
end
```

Anomaly.m

```
function ano_TS = Anomaly(data_TS,date)
% data_TS = time series created using the GenTimeSeries.m function
% date    = list of date strings created using the Date_Gen.m function

N = length(date);

monthindex = str2num(date(:,6:7));

Y  = zeros(12,1);
NY = zeros(12,1);

for i = 1:N
    Y(monthindex(i))  = Y(monthindex(i)) + data_TS(i);
    NY(monthindex(i)) = NY(monthindex(i)) + 1;
end

Y = Y./NY;

for i = 1:N
    data_TS(i) = data_TS(i) - Y(monthindex(i)) + mean(Y);
end

ano_TS = data_TS;
```

PlotLimits.m

When analyzing the areas off the coast of Peru and Angola, we filter out all data outside the defined limits, which leaves us with new data from which we can create area oriented time series.

```
function [lim_val,lim_lon,lim_lat] = PlotLimits(val,lon,lat,lowlon,uplon,lowlat,uplat)
% lim_val = data filtered with limits
% lim_lon = filtered longitude vector
% lim_lat = filtered latitude vector
%
% val = original raw data
% lon = vector with the corresponding longitude
% lat = vector with the corresponding latitude
% lowlon = lower limit for longitude
% uplon  = upper limit
% lowlat = lower limit for latitude
% uplat  = upper limit

% val, lon and lat, must be column vectors so if not we transpose
if size(val,2) > 1
    val = val';
end

if size(lon,2) > 1
    lon = lon';
end

if size(lat,2) > 1
    lat = lat';
end

Data = [val,lon,lat];

% Finds data that exceeds the chosen range and filters them out
% Longitude
lonlim = find(Data(:,2)>uplon | Data(:,2)<lowlon);
Data(lonlim,:)=[];

% Latitude
latlim = find(Data(:,3)>uplat | Data(:,3)<lowlat);
Data(latlim,:)=[];


lim_val = Data(:,1)';
lim_lon = Data(:,2)';
lim_lat = Data(:,3)';
end
```

DataOverlap.m

In the case where we have 2 sets of data that overlap for a limited period, like ISCCP and MODIS, this function takes in the date list and time series of both and cuts out all data from outside the interval in which they overlap. This is useful in order to determine the correlation of the data in the period they both collected data.

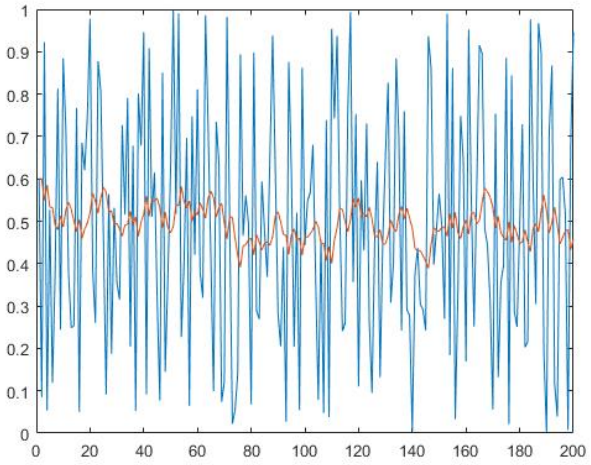
```
function [out_date1,out_TS1,out_date2,out_TS2] = DataOverlap(in_date1,in_TS1,in_date2,in_TS2)
% in_date1 and 2 = date list of the 2 data sets
% in_TS1 and 2   = time series for the 2 data sets
%
% out_date1 and 2 = new filtered date list of the 2 data sets
% out_TS1 and 2   = new filtered time series for the 2 data sets

start_index = find(in_date1==in_date2(1));
end_index   = find(in_date2==in_date1(end));

out_date1 = in_date1(start_index:end);
out_date2 = in_date2(1:end_index);

out_TS1   = in_TS1(start_index:end);
out_TS2   = in_TS2(1:end_index);
```

ExpSmooth.m

This function performs the exponential smoothing used in analysis of cosmic rays and sunspot activity in section 7.4.

```
function s_t = ExpSmooth(x_t,alpha)
% x_t   = original time series
% alpha = smoothing factor
%
% s_t   = smoothed time series


N = length(x_t);

s_t = zeros(N,1);

s_t(1) = x_t(1);

for i = 2:N
    s_t(i) = alpha*x_t(i-1)+(1-alpha)*s_t(i-1);
end
```

```
Example:
N = 200;
x = 1:N;
y = rand(N,1);
yy = ExpSmooth(y,0.1);

plot(x,[y,yy])

ylim([-0.2,1.2])

legend('Raw data','Exponentially smoothed data')
```

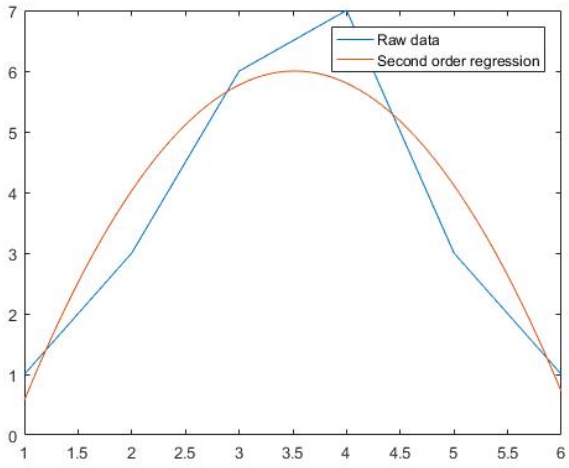


PolyRegression.m

Following takes in the data along with the time vector, and produces a regression of any order.

```
function output_data = PolyRegression(input_x,input_y,new_x,order)
% input_x = x data on which the regression is done
% input_y = y data on which the regression is done
% new_x   = if the regression should be plotted on a timeline
%           with more points, this can be used. If input_x
%           should remain the x axis of choice, set new_x = 0
% order   = order of regression

x = input_x;
y = input_y;

if new_x == 0
    new_x = x;
end

% input_x, input_y and new_x must be column vectors so if not we transpose
if size(x,2) > 1
    x = x';
end

if size(y,2) > 1
    y = y';
end

if size(new_x,2) > 1
    new_x = new_x';
end

N = length(x);

A = zeros(N,order+1);

for i = 1:order+1
    A(:,i) = x.^(i-1);
end

c = (A'*A)\(A'*y);

output_data = zeros(length(new_x),1);

for i = 1:order+1
    output_data = output_data + c(i)*new_x.^(i-1);
end
end
```

```
Example:
input_x = 1:6;
input_y = [1;3;6;7;3;1];
order = 2;
xx = linspace(1,6,200);
yy = PolyRegression(input_x,input_y,xx,order);

plot(input_x,input_y)
hold on
plot(xx,yy)

legend('Raw data','Second order regression')
```



DTU Space
National Space Institute
Technical University of Denmark

Elektrovej, building 327
DK - 2800 Kgs. Lyngby
Tel (+45) 4525 9500
Fax (+45) 4525 9575

www.space.dtu.dk