Technische Universität München

Master's Thesis

Cloud Statistics from Calipso Lidar Data for the Performance Assessment of a Methane Space Lidar

Author:
Nico Trebbin

Supervisor:
Dr. Christoph Kiemle

A thesis submitted in fulfillment of the requirements
for the degree of Master of Science

in the

Earth Oriented Space Science and Technology

Master's Program

October 2013

Declaration of Authorship

I, Nico Trebbin, declare that this thesis titled, 'Cloud Statistics from Calipso Lidar Data for the Performance Assessment of a Methane Space Lidar' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date of Submission: _____

"Science never solves a problem without creating ten more."

George Bernard Shaw

TECHNISCHE UNIVERSITÄT MÜNCHEN

Abstract

Ingenieurfakultät Bau Geo Umwelt

Master of Science

Cloud Statistics from Calipso Lidar Data for the Performance Assessment of a Methane Space Lidar

by Nico Trebbin

In this thesis a performance assessment for the future German-French climate monitoring initiative, Methane Remote Sensing Lidar Mission (MERLIN), proposed by DLR and CNES in 2010 was undertaken. A general space lidar performance issue is the obstruction by optically dense clouds. For this purpose cloud free statistics, the global cloud top flatness and global cloud top distributions were derived from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) level 2, 333 m and 5 km lidar cloud-layer products between 01 January 2007 and 01 January 2008. Merging both data sets together thereby allowed the best possible simulation of near global and seasonal real world atmospheric conditions that a spaceborne Integrated Path Differential Absorption (IPDA) lidar like MERLIN will encounter. With 40.5 % overall global cloud free fraction, a cloud gap distribution which is following a power-law distribution with exponent $\alpha = 1.51 \pm 0.01$ together with a mean cloud gap length of 7.41 km and about 200 daily global cloud top flatness events, the analysis reveals a dominance of small cloud gaps which is confirmed by a low median cloud gap length of only 1 km. While the cloud free fraction results were compared and confirmed with Aqua Moderate Resolution Imaging Spectroradiometer (MODIS) seasonal and annual cloud fraction data, the power-law distribution of cloud gaps was confirmed by an extensive statistical analysis using maximum likelihood estimation, Kolmogorov-Smirnov statistics and likelihood ratio tests. Taking 605 x 10^8 individual CALIPSO measurements of the year 2007 with a horizontal resolution of 333 m and computing cloud gap and cloud free statistics for 2 ° x 2 ° latitude/longitude grid points thereby identified regional and seasonal changes in the probability of spaceborne lidar surface detection. The analysis reveals that MERLIN will be able to perform near global methane mixing ratio column retrievals.

Acknowledgements

I would like to thank ICARE for providing me the CALIPSO data through their FTP server, Dr. Christoph Kiemle for his superior supervision, Dr. Axel Amediek, Dr. Mathieu Quatrevalet and Dr. Stephan Kox for the consultation in preparation of the thesis and the developers of Python, NumPy, Matplotlib, Basemap, PyHDF, h5Py, powerlaw and PyQt for providing me with an excellent software basis for developing the software used within this thesis.

Contents

List of Figures

List of Tables

Abbreviations

ABL	**A**tmospheric **b**oundary **l**ayer
ATBD	**A**lgorithm **T**heoretical **B**asis **D**ocument
BSP	**B**ack**s**catter **p**rofiles
CALIOP	**C**loud-**A**erosol **L**idar w**i**th **O**rthogonal **P**olarization
CALIPSO	**C**loud-**A**erosol **L**idar and **I**nfrared **P**athfinder **S**atellite **O**bservations
CCDF	**C**omplementary **c**umulative **d**istribution **f**unction
CDF	**C**umulative **d**istribution **f**unction
CNES	**C**entre **N**ational d'**E**tudes **S**patiales
CPR	**C**loud **P**rofiling **R**adar
DLR	**D**eutsches Zentrum für **L**uft- und **R**aumfahrt
EarthCARE	**E**arth **C**louds, **A**erosols and **R**adiation **E**xplorer
ECMWF	**E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts
ENVISAT	**Env**iromental **Sat**ellite
EOS	**E**arth **O**bserving **S**ystem
ESA	**E**uropean **S**pace **A**gency
FOV	**F**ield **o**f **v**iew
GAW	**G**lobal **A**tmosphere **W**atch
GLAS	**G**eoscience **L**aser **A**ltimeter **S**ystem
GOSAT	**G**reenhous Gases **O**bserving **Sat**ellite
ICARE	Cloud-Aerosol-Water-Radiation Interactions
IPDA	**I**ntegrated **P**ath **D**ifferential **A**bsorption
IIR	**I**nfrared **I**maging **R**adiometer
IPA	**I**nstitut für **P**hysik der **A**tmosphäre
ITCZ	**I**nter**t**ropical **c**onvergence **z**one
LITE	**L**IDAR **I**n-Space **T**echnology **E**xperiment

LaRC	**La**ngley **R**esearch **C**enter
MERLIN	**Me**thane **R**emote Sensing **Li**dar Missio**n**
MODIS	**Mod**erate Resolution **I**maging **S**pectroradiometer
NASA	**N**ational **A**eronautics and **S**pace **A**dministration
PCFLOS	**P**robability of **c**loud **f**ree **l**ine-**o**f-**s**ight
NOAA	**N**ational **O**ceanic and **A**tmospheric **A**dministration
PDF	**P**robability **d**ensity **f**unction
PSC	**P**olar **s**tratospheric **c**loud
SNR	**S**ignal-to-**n**oise **r**atio
TOA	**T**op **o**f the **a**tmosphere
TTL	**T**ropical **T**ropopause **L**ayer
UTC	**U**niversal **T**ime **C**onversion
VFM	**V**ertical **F**eature **M**ask
WDCGG	**W**orld **D**ata **C**entre for **G**reenhause **G**ases
WFC	**W**ide **F**ield **C**amera
WMO	**W**orld **M**eteorological **O**rganization

SI Units and Abbreviations

deg	degree
$^{\circ}C$	degree Celsius
J	joule
K	kelvin
km	kilometer
m	meter
ms	millisecond
nm	nanometer
per, %	percent
s, sec	second
sr	steradian
V	volt
W	watt
μm	micron, micrometer

float32	floating point, 32 bits or 4 bytes
float64	floating point, 64 bits or 8 bytes
GB	Gbyte, gigabyte, 10^9 byte
int8	integer, 8 bits or 1 byte
int16	integer, 16 bits or 2 bytes
int32	integer, 32 bits or 4 bytes
MB	Mbyte, megabyte, 10^6 byte
uint8	unsigned integer, 8 bits or 1 byte
uint16	unsigned integer, 16 bits or 2 bytes
uint32	unsigned integer, 32 bits or 4 bytes

Symbols

c_{gap}	cloud gap	km
C_{Lay}	cloud layer	No Unit
f_c	cloud fraction	No Unit
f_{cf}	cloud free fraction	No Unit
z_t	cloud top height	km
z_b	cloud base height	km
α	power law exponent	No Unit
β	backscatter coefficient	$km^{-1}sr^{-1}$
λ	wavelength	nm
σ	significant error	No Unit
τ	optical thickness	No Unit

Dedicated to the ones I love and to the only planet inhabiting seven billion people that someday might change their mind and stop wrecking themselves and the nature they are living in.

Chapter 1

Introduction and motivation

1.1 The MERLIN mission

1.1.1 Methane as an important trace-gas

Methane (CH_4), which occurs at about 1800 ppb (parts per billion which $\widehat{=}$ nmol mol^{-1}, dry air mixing ratio) by volume in the atmosphere of Earth (`http://www.esrl.noaa.gov/gmd/aggi/aggi.html`), is considered the second most important anthropogenic greenhouse gas after CO_2. According to the US National Oceanic and Atmospheric Administration (NOAA), methane is 21 times more effective at trapping heat in the atmosphere than CO_2 by weight, over a 100-year period. Present estimates of 2012 are that methane accounts for 0.507 W m^{-2} radiative forcing which equals about 17.5% of the total forcing potential of all long-lived greenhouse gases. Although atmospheric methane levels were almost constant from 1999 to 2006, the CH_4 globally averaged concentration in the atmosphere started rising again. Causes for this recent increase include an Arctic warm period in 2007 and increased precipitation in the tropics for a period from 2007 to 2008. Fortunately, near-zero growth in the Arctic during 2008 suggests that climate feedbacks as a result of melting permafrost and nascent CH_4 hydrates have not been activated yet [Dlugokencky et al., 2009]. Those CH_4 hydrates are on the one hand stored on the seabed along continental margins where they are stabilized by high pressure and low temperatures and on the other hand in soils trapped by the overlying ice sheet of the permafrost. Implying a global warming could lead to a hydrate instability in the oceans due to ocean warming and a melting of permafrost regions, which consequently would result in an additional release of methane into the atmosphere [Forster et al., 2007].

As the global abundance of atmospheric methane increased by a factor of 2.5 since 1750 [Dlugokencky et al., 2011] one reason for scientists to conduct methane analysis is to

assess methane emissions in response to large-scale climate change events in the past. While palaeorecords can help to identify the global long term trend and can project what to expect in the future, present-day analysis is important to identify natural and anthropogenic sources of methane emissions such as wetlands and CH_4 hydrates, fossil fuel exploitation, waste management and biomass burning. Implying a constant increase in the world population and living standards could also lead to additional anthropogenic methane emissions in the future. Besides that, the predicted global warming may increase natural methane emissions due to a positive feedback on the climate system.

As the regions that contribute generally to natural and anthropogenic atmospheric methane have been identified and the existing sampling coverage is sufficient to capture large-scale zonally averaged features, precise regional assessments are not possible yet. Current climate models use inverse calculations [Zhao et al., 2009] and extrapolation from a relatively small number of direct measurements from ground networks or from satellites [Bergamaschi et al., 2013] to estimate methane emissions. This data has obvious value, but due to the loss of ENVISAT in April 2012 and to the last remaining CH_4 observing satellite GOSAT (`http://www.jaxa.jp/projects/sat/gosat/index_e.html`), it could be improved significantly with widespread additional accurate measurements. While present measurements of methane from satellites rely on passive remote sensing with spectrometers which are dependent on sunlight conditions and generally suffer from low measurement sensitivity in the lower troposphere close to the methane sources, additional measurements from active remote sensing satellites will help reduce uncertainty in climate models and lead to a more comprehensive understanding of the CH_4 sources [Kiemle et al., 2011].

In addition to the general scientific interest, methane monitoring is and will be an important component of a greenhouse gas regulatory framework, as agreed in the Kyoto protocol which was extended until 2020. To achieve the goals of the protocol, a continuous global high-resolution monitoring of all greenhouse gases is required and the identification of discrepancies and fraudulent activities of countries who ratified the protocol is, besides the scientific aspect, one of the most important reasons of an extension of monitoring capabilities. The German-French climate initiative [Stephan et al., 2011] with the Methane Remote Sensing Lidar Mission (MERLIN) is intended to help reduce monitoring uncertainties. It will for the first time utilize a space-based active differential absorption lidar, which is particularly sensitive near the methane sources, while being independent of sunlight, offering high and quantifiable accuracy and providing the community with near-global measurements.

1.1.2 The global methane budget

Both greenhouse gases CO_2 and CH_4 play roles in the natural and anthropogenic cycle of carbon, involving continuous flows of large amounts of carbon among the ocean, the terrestrial biosphere and the atmosphere. By photosynthesis, plants and the terrestrial biosphere convert carbon (especially CO_2) into biomass and plant growth and thereby decrease the atmospheric concentration of CO_2. Additionally, carbon dioxide dissolves directly from the atmosphere by forming carbonic acid during the reaction with water bodies (e.g. oceans, lakes, etc.) which then can be absorbed by rocks through weathering or simply being washed into the ocean. The major removal mechanism of CH_4 from the atmosphere is its reaction with the hydroxyl radical OH in the troposphere and stratosphere. However, plant, soil and animal respiration as well as decomposition of dead biomass, return carbon to the atmosphere as CO_2 or as CH_4 by microorganisms under anaerobic conditions. While the cycle of plant photosynthesis, biomass production and decay of dead organic matter was stable until 1750, the injection of additional CO_2 and CH_4 gases into the atmosphere by land use, deforestation and fossil fuel burning disturbed the stability of the cycle [Forster et al., 2007]. Taking into account the relatively short lifetime of CH_4 in the atmosphere of approximately nine years, the accurate detection of methane sources could lead to a mitigation of at least further anthropogenic emissions. While emissions of CH_4 from wetlands vary significantly depending on meteorological conditions [Dlugokencky et al., 2011], and other natural sources such as termites, oceans, wild fires, and wild ruminant animals remain in balance according to pre-industrial atmospheric CH_4 abundance, mainly anthropogenic sources contribute to the rise in atmospheric concentration. These man-made methane sources are generated by crushing, transmission, distribution and processing from coal, oil and gas extraction, as well as agricultural sources such as rice cultivation, domesticated animals and biomass burning. Additionally leaky waste disposals account for a total of approximately 230-467 Tg $CH_4 yr^{-1}$ (1 Tg = 10^{12}g) emitted by anthropogenic sources. Compared to the 124-275 Tg $CH_4 yr^{-1}$ naturally emitted methane which are expected to remain constant [Forster et al., 2007], it is obvious that the anthropogenic sources have to be detected and somehow limited to stop a further growth of methane accumulation in the atmosphere.

1.1.3 In-situ measurements and remote sensing

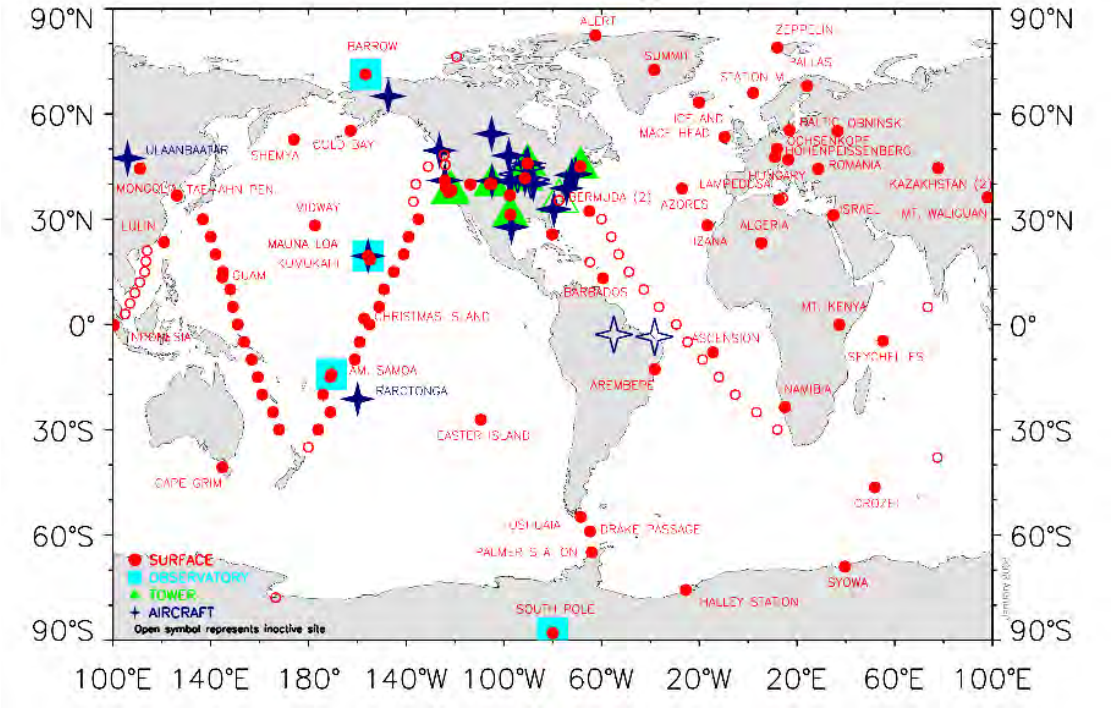
For this purpose NOAA, the World Data Centre for Greenhouse Gases (WDCGG) and others [Dlugokencky et al., 2011, Sasakawa et al., 2011] operate in situ atmospheric CH_4 measurement networks, with surface stations, observatories, towers and aircraft missions that measure directly at the point of interest the atmospheric composition

and the methane content. While some of the sites provide in situ analyzers for quasi-continuous measurements of CH_4 mixing ratios, others are only used to compare and validate their measurements against other stations on an hourly basis to ensure the quality of the measured mixing ratios. As it is highly important to use a uniform measurement strategy and to agree on standards for the use in climate models, NOAA maintains the standard scale used to ensure the World Meteorological Organization (WMO) Global Atmosphere Watch's (GAW) compatibility target of ± 2 ppb or 0.1 % measurement uncertainty [Dlugokencky et al., 2011].

A variety of in situ sampling devices is used such as gas chromatographs with an accuracy better than ± 2 ppb, optical devices using off-axis integrated cavity output spectroscopy or cavity ring down spectroscopy with an accuracy better than ± 0.25 ppb, or flow GC/isotope ratio mass spectrometers with an outstanding accuracy of $\pm$ 0.05 ppb [Dlugokencky et al., 2011]. Most time series of CH_4 mixing ratio and isotopic composition are collected by discrete flask samples collected at remote sites, oil-rigs, buoys or ships. In addition to the continuous marine observation at approximately 15-40 m height above sea level, several tower sites attached to radio and television towers complement the global measurement network with in situ measurements of up to 500 m altitude. As these towers are mostly in the proximity of populated urban areas, they contribute significantly to the understanding of anthropogenic CH_4 emission. These measurements are sometimes complemented by aircraft, capturing the vertical distribution of methane in the troposphere.

Unfortunately, the observation sites are unequally distributed with a high concentration in developed countries and an accumulation in the USA, Japan and Europe (see Figure 1). Therefore important areas for CH_4 emissions, like the tropics, Siberia and South Asia are not covered by the sparse network and its coarse sampling. To complement the missing methane information globally the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) on-board of ENVISAT and the greenhouse gas observation sensor TANSO-FTS on-board of GOSAT are used. While both passive sensors rely on sunlight to retrieve CH_4 column-averaged mixing ratios, the retrieval fails at night, in polar regions and whenever clouds hinder a clear line of sight. Unfortunately, as already mentioned, ENVISAT failed which is why the urgent need for additional ground (to validate satellite data) and space-based measurements has arisen. To circumvent the limitation of the passive spectrometers and to better understand the complex methane cycle, an active integrated path differential absorption (IPDA) lidar was conceived in a French-German initiative to retrieve spatial and temporal gradients of atmospheric methane columns.

(a)



(b)

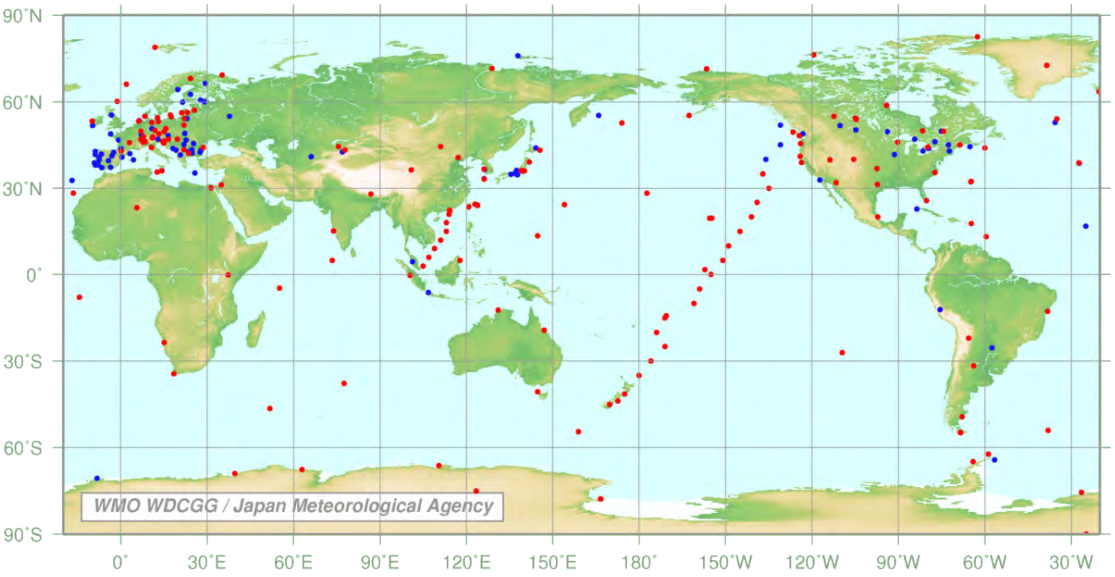


Figure 1: The two main distributors of global methane atmospheric contents NOAA and the WDCGG with their global measurement networks. For (a) the NOAA Earth System Research Laboratory surface stations are indicated by a red dot, observatories by a turquoise square, towers by a green triangle and aircraft sampling sites are indicated by a blue star. Where open, symbols represent inactive sites. For the WDCGG network, the red dots represent surface stations added to the network during the last year and the blue dots indicate sites existing longer than that. Both maps where directly adapted from `http://www.esrl.noaa.gov/gmd/aggi/aggi.html` and `http://ds.data.jma.go.jp/gmd/wdcgg/` respectively.

1.1.4 Scientific challenges and measurement strategy

While the retrieval of methane columns with passive spectrometers is highly limited to certain conditions, the use of IPDA lidar ensures high sensitivity near the surface and extends the observable latitude region to polar areas [Kiemle et al., 2011], which in the case of Siberia could emerge to be a large methane emitter in the next decades. Additionally the lidar technology reduces the aerosol biases passive instruments encounter and which blur the methane retrieval because of the small particle sizes that are buried in the background signal and cannot be clearly distinguished from other constituents. The IPDA technology was also proposed for European Space Agency (ESA) and National Aeronautics and Space Administration (NASA) missions such as Advanced Space Carbon and Climate Observation of Planet Earth (A-SCOPE) and Active Sensing of CO_2 Emissions over Nights, Days and Seasons (ASCENDS). A few recent studies have focused on the expectable performance of such lidars. In general, the technology is based on laser light reflections off the Earth's surface which makes it independent of solar illumination and provides column measurements with an accuracy better than 0.1% and with a precision of around 1% if the single columns are averaged to a 50 km column [Kiemle et al., 2011].

The basic measurement strategy relies on laser light being backscattered from the surface while two wavelengths are emitted by the laser. One wavelength is placed near the center of a methane absorption line and is denoted as on-line, and the second reference off-line wavelength is placed slightly apart from the absorption line. The ratio of the on-line and off-line power of the backscattered laser photons, together with their pulse energy is then used to retrieve information on the methane content. Because of the difference in atmospheric transmission of two co-aligned laser beams with slightly different wavelengths (see Table 1), the off-line wavelength experiences less absorption and therefore results in higher return power detected by MERLIN. As MERLIN will fly on a dawn-dusk polar orbit it will be supplied with enough solar power to operate the laser continuously. While the key parameters of the payload (see Table 1) are fixed once they are finalized, the surface albedo and the atmosphere are highly variable and cloud and aerosol cover could highly influence the performance of MERLIN's methane retrievals. Therefore the precision of the retrievals relies entirely on the mixture of the atmospheric column covered by the laser beam.

Since clouds and thick aerosols are a problem and there is a high demand on prior knowledge and what to expect from a satellite mission like MERLIN, an extensive study on the global cloud distribution and the cloud gap length was performed in this thesis. As MERLIN's single shot footprint is 135 m, several individual consecutive measurements have to be averaged to ensure the necessary SNR to retrieve information on the methane

content in the atmosphere on the desired nominal resolution of 50 km (see Figure 2). Preferably all of these measurements are cloud and aerosol free to ensure the desired measurement precision of 1%. Unfortunately, these conditions are most unlikely and an extensive analysis of real world atmospheric conditions has to be performed.

IPDA lidar payload (MERLIN)	
Laser pulse energy	9.0 mJ
Average output power (on and offline)	0.45 W
Spot diameter	135 m
On-line wavelength	1645.552 nm
Off-line wavelength	1645.846 nm
Telescope diameter	0.55 m
Satellite Platform	
Orbit	polar, sun-synchronous, dawn/dusk
Equator crossing time	1:30 a.m./p.m.
Orbit altitude	506 km
Footprint velocity	7.06 km/s
Along-track resolution	50 km

Table 1: Key parameters proposed during Phase A for MERLIN [Adapted from Kiemle et al., 2011]

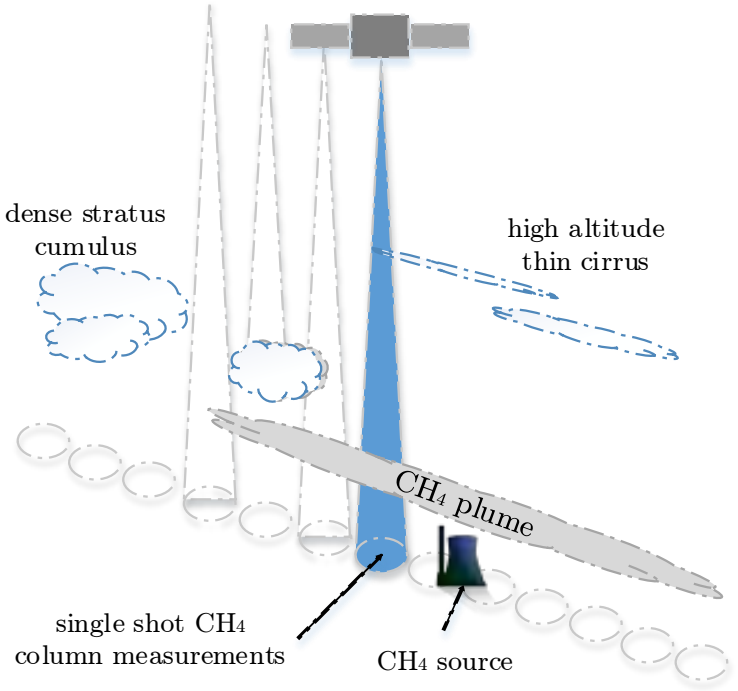


Figure 2: Schematic sketch of MERLIN's atmospheric CH_4 column gradient measurements with indicated different cloud types disturbing either the column retrieval completely for dense clouds attenuating the lidar signal, or partly for cloud edges or thin cirrus. A hypothetical CH_4 point emission source generating a plume with CH_4 levels higher than the background value is included and indicated by the power station.

1.2 The CALIPSO mission

1.2.1 Global cloud coverage by lidar measurements from space

With a measurement geometry, orbit configuration and a two-wavelength polarization-sensitive lidar similar to MERLIN (see Table 2) the NASA Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission is the ideal base for an extensive high resolution cloud distribution analysis. CALIPSO was launched on 28 April 2006 as part of the A-train ("afternoon") constellation of satellites on a 16 day repeat cycle, and offers state of the art insights into aerosol and clouds properties, their various interactions, and their roles in the climate system [Winker et al., 2010b]. As one of the fundamental problems of passive satellite sensors is the inability to retrieve vertical profiles of clouds and aerosols and therefore their optical properties and extent, CALIPSO uses its lidar to distinguish optically thin boundary layer clouds from dense stratus clouds and even from aerosol. Cloud top altitude and extent is observed directly which is why cloudy and cloud free atmospheric columns are clearly distinguished. This information is crucial for MERLIN as the cloud layers and optical properties detected by CALIPSO can be used to retrieve a vertical global distribution of clouds disturbing the surface measurement of MERLIN. Furthermore CALIPSO is the only instrument on-board a satellite which profiles down to the Earth's surface, which is one of the prerequisites of MERLIN's column methane retrieval. In conclusion, CALIPSO's lidar data are optimal to test the performance of MERLIN.

With currently 7 years of global data many studies on cloud and aerosol distribution have been published [Winker et al., 2010a, Wu et al., 2011] and sometimes CALIPSO was used to validate passive and active sensor data [Berthier et al., 2008, Karlsson and Johansson, 2013], or to examine a specific cloud type [Chan and Comiso, 2011, Leahy et al., 2012, Pitts et al., 2009, Schwartz and Mace, 2010, Veglio and Maestri, 2011, Wu et al., 2008, Young, 2012]. With the missions focus on cloud and aerosol detection it is quite comprehensible that the scientific community has published a formidable amount of studies on clouds, their properties and distribution. But only few studies [Emmitt et al., 2012, Reinke and Vonder Haar, T. H., 2011] have been published with the focus on cloud free occurrences and the needs for future lidar missions.

1.2.2 Scientific challenges and limits of CALIPSO

While the general performance of the cloud detection of CALIPSO is good even for the single shot resolution of 333 m, some cloud retrievals remain challenging for a lidar system. Typical cirrus clouds or polar stratospheric clouds [Pitts et al., 2009] sometimes

CALIOP payload	
Laser pulse energy	110 mJ for each wavelength
Polarization	532 nm parallel, perpendicular
Footprint diameter	70 m
Receiver FOV	90 m
Footprint spacing	335 m
Wavelengths	532, 1064 nm
Laser repetition rate	20.16 Hz
Laser pulse length	20 ns
Telescope diameter	1 m
Off-nadir pointing	3.0 ° since Nov. 2007
Satellite Platform	
Orbit	polar, sun-synchronous
Equator crossing time	6:30 a.m./p.m.
Orbit altitude	705 km
Footprint velocity	7.51 km/s
Along-track resolution	335 m
Repeat cycle	16 days

Table 2: Key parameters of the CALIPSO satellite mission

need an extensive averaging of many consecutive lidar measurements to be detected correctly. While the lidar started almost pointing towards nadir, and was therefore able to detect cirrus clouds with horizontally oriented ice crystals, the pointing direction was increased to 3.0 ° in November 2007, which allowed even more accurate determinations of the physical and optical properties of cirrus clouds [Hunt et al., 2009]. A detailed description of the cloud detection algorithm of CALIPSO is given in Chapter 2. Generally, clouds not instantly detected by CALIPSO could also be penetrated by MERLIN and therefore a methane retrieval might still be possible. However it has to be mentioned that CALIPSO not only uses 110 mJ laser pulse energy compared to MERLIN's 9 mJ, but it also has a larger telescope which is indeed counterbalanced by the orbital height and it was specifically designed to retrieve and penetrate multiple cloud layers. Consequently the IPDA lidar is attenuated more quickly by its path trough the atmosphere. [Young, 2012] found out using nighttime measurements, that CALIPSO could penetrate clouds determined to be opaque from around 800 m for broken cumulus, and 970 m for stratocumulus to around 3100 m for cirrus clouds. But CALIPSO sometimes fails to detect a cloud at all as [Chan and Comiso, 2011] found out by comparing MODIS data with CloudSat and CALIPSO. Accordingly undetected clouds are geometrically thin, low-level clouds, originated in the Arctic region and remain below 2.5 km altitude with layer thickness not exceeding 1 km. But using CALIPSO cloud data for the cloud free analysis is still maintainable as CALIPSO is the only satellite which provides up to 30 m vertical and 333 m horizontal resolution on a global scale and uses a similar measurement technique as MERLIN. It might be possible to even obtain methane retrievals from flat low-level clouds and partly broken thin clouds by calculating partial columns

along the satellite track. To analyze the occurrence of such events, the high resolution of CALIPSO is also advantageous.

1.3 Goals

In summary this study uses the single shot CALIPSO cloud lidar data of 2007 to assess the expectable performance of the future methane space-lidar MERLIN. Altough MERLIN's general requirement demands cloud free line-of-sight onto the surface for at least 50 km and preferably over anthropogenic methane sources, this condition is highly unlikely to be fulfilled globally. To detect the regions where an atmospheric methane retrieval is possible, this study focuses on statistics of the cloud free regions detected by CALIPSO. Thereby not only the cloud free fraction and their seasonal variation is analyzed, but also individual cloud gaps contributing to the cloud free fraction are identified. Furthermore the occurrence frequency, the mean and median length of all cloud gaps is statistically analyzed and checked against common mathematical distributions. To account for the possibility of proposed methane retrievals from Partial columns, cloud top flatness events were searched and statistically analyzed. Closing this study and to bring it down to a round figure a global cloud top height statistic is generated and the findings are compared to related studies.

Chapter 2

Background information and methodology of CALIPSO

2.1 The fundamental equations

The fundamental physical principle behind CALIPSO observations is the scattering of radiation by molecules, aerosols and cloud particles. The important equations and principles are introduced in this chapter. Whenever an electromagnetic wave interacts with a particle, it can either be scattered, absorbed, or pass unperturbed. Assuming now, that a beam of radiation is passing through an arbitrarily thin layer of the atmosphere along a specific path with gas molecules and particles, the *intensity* is decreased by the increment:

$$dI_\lambda = -I_\lambda \gamma_\lambda ds \tag{1}$$

where I_λ is the intensity at wavelength λ, γ_λ is the volume extinction coefficient due to the effect of scattering and absorption in m^{-1}, and ds is the path length along the ray path of the incident radiation. By integrating (1) from altitude z_0 to $z > z_0$ which yields an exponential attenuation, one derives:

$$I_\lambda = I_{\lambda,0} e^{-\int_{z_0}^{z} \gamma_\lambda(z')dz'} \tag{2}$$

with I_λ being the monochromatic radiance at z and $I_{\lambda,0}$ being the monochromatic radiance at z_0. By defining the *transmission* T as a fraction of radiation that passes through part of the atmosphere

$$T = e^{-\int_{z_0}^{z} \gamma(z')dz'} \tag{3}$$

(1) can be formulated as [Wallace and Hobbs, 2006]:

$$I_\lambda = I_{\lambda,0} T. \tag{4}$$

Now, either a *elastic* or *inelastic* scattering process can occur. In an elastic scattering process, the wavelength and frequency of the generated wave is not substantially changed, e.g. if the beam is scattered from aerosol particles or atoms and molecules. The representative scattering phenomena are called *Rayleigh scattering* and *Mie scattering*. The first one occurs if the size of the molecules or particles is much smaller than the laser wavelength. On opposite, for spherical particles much larger than the wavelength, Mie scattering is the dominant process. *Raman scattering*, where a small fraction of the scattered light from an atom or molecule is however scattered by an excitation and result in photons having a lower frequency than the frequency of the incident photons, is a representative of the *inelastic* scattering process. Alternatively if the light interacts with density changes of the compressed air, the refraction index changes and the light path consequently bends. This *inelastic* scattering process is called *Brillouin scattering*.

Knowing that in the Earth's atmosphere various particle sizes and shapes, i.e. aerosols, cloud droplets and ice crystals are always present, it is a simplifying approach to consider an ideal spherical particle of homogeneous material with a complex refraction index $m = m_r + im_i$. The real part m_r is the ratio of the speed of light in a vacuum to the speed at which light travels when it is passing through the particle. The scattering, absorption, or extinction efficiency $K\lambda$ can be prescribed as a function of a dimensionless size parameter [Wallace and Hobbs, 2006]:

$$x = \frac{2\pi r}{\lambda} \tag{5}$$

where r is the radius of the particle, and λ is the wavelength of the incident wave.

As seen in Figure 3 the scattering efficiency K_λ is quite low for $x < 1$ which is called the *Rayleigh scattering regime* and where

$$K_\lambda \propto \lambda^{-4}. \tag{6}$$

For $0.1 \leq x \leq 50$ the efficiency follows a damped oscillatory behavior and refers to the *Mie scattering regime*. The *geometric optics regime* starts as $x \geq 50$. Taking into account this idealized representation, and applying Figure 3 to the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) 532 nm channel, Rayleigh scattering can be expected from air molecules, whereas Mie scattering is dominant for aerosol particles of size 11 nm to 3 μm, and geometric optics can be expected from water droplets and ice crystals. For the 1064 nm channel, Rayleigh scattering from molecules should be

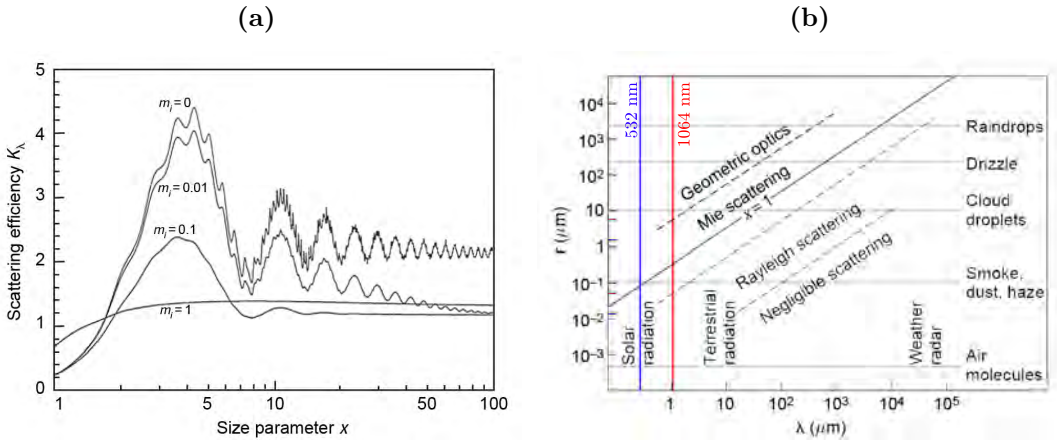


Figure 3: The idealized spherical particle in dependence from the wavelength and its the imaginary part of the refractive index. (a) describes the size parameter x as a function of λ of the incident radiation and particle radius r, while (b) shows the scattering efficiency K_λ as a function of size parameter x, for different refractive indices. The blue line and red line represent the 532 nm channel and the respective 1064 nm channel of CALIPSO [Adapted from Wallace and Hobbs, 2006, p.121-122 and slightly changed].

about 16 times (according to equation 6) weaker compared to the 532 nm channel, Mie scattering can be expected from particles of size 18 nm to 8 μm, and geometric optics can be expected from cloud particles.

As the receiver of CALIPSO measures emitted backscattered laser light from molecules, aerosols and cloud particles, the time span from sending and receiving the pulse is directly correlated to the altitude. Hence, absorption-emission and multiple scattering that might have been induced can not be measured by CALIOP. The measured signal strength P in the so-called lidar equation is directly related to the range-dependent *backscatter coefficient* β by [Winker et al., 2006]:

$$P(r) = \frac{1}{r^2} E_0 G_A C \beta(r) T^2(r) \tag{7}$$

where the two-way attenuation is given by:

$$T^2(r) = e^{-2 \int_{z_{sat}-r}^{z_{sat}} \sigma(z) dz} \tag{8}$$

r = range from the satellite to the sampled volume,
$P(r)$ = measured signal after background subtraction and artifical removal,
E_0 = average laser energy for the single-shot or composite profile,
G_A = amplifier gain or telescope area,
and C = lidar calibration factor,
$\beta(r)$ = volume backscatter coefficient at range r,
$T(r)$ = one-way transmittance from the lidar to the scattering volume at range r,
$\sigma(r)$ = volume extinction coefficient at range r,
z_{sat} = satellite height.

The Level 1 profile products are range-scaled, energy normalized, gain normalized and calibrated versions of the signal (P) and are expressed in terms of the attenuated backscatter coefficient in units of $\mathrm{km}^{-1}\ \mathrm{sr}^{-1}$ [Winker et al., 2009]:

$$\beta'(r) = [\beta_p(r) + \beta_m(r)]T^2(r_c, r) \tag{9}$$

where $\beta_p(r)$ and $\beta_m(r)$ represent the backscatter contributions from particles and from the molecular atmosphere at range r from the satellite. $T^2(r_c, r)$ thereby represents the two-way transmittance between a calibration region at z_c and range z. For all three channels (532 nm parallel and perpendicular, 1064 nm) backscatter coefficients can be split into their molecular and particle scattering constituent. The calibration of the signal consists of converting the measured signal profile $P(r)$, in digitizer counts, to $\beta'(r)$:

$$\beta'(r) = \frac{r^2 P(r)}{C E_0 G_A}. \tag{10}$$

The use of high-altitude nighttime return signals above 40.2 km and of a molecular model is the standard lidar calibration technique for CALIPSO, assuming no particles at that altitude [Winker et al., 2009].

In a next step the Level 1 profile is converted into a profile of attenuated scattering ratio $R'(z)'$,

$$R'(z) = \frac{\beta'(r)}{\beta'_m(r)} \tag{11}$$

in which the profile of $\beta'_m(r)$ is estimated using gridded molecular and ozone number density profile data from the Goddard Earth Observing System Model. The magnitude of $R'(z)$ is then indicative for an identified layer, as enhancements above the expected return signal are directly correlated to cloud or aerosol layers [Winker et al., 2009].

2.2 The CALIOP payload

The CALIPSO satellite consists of three instruments, a Infrared Imaging Radiometer (IIR), a Wide Field Camera (WFC), and the lidar CALIOP. While the IIR and WFC are only supportive passive instruments, the laser transmitter subsystem of CALIOP includes two identical laser transmitters, of which the primary laser failed in March 2009 as it has been predicted by [Hunt et al., 2009]. Each laser is equipped with a beam expander and a beam steering system which ensures alignment between the transmitter and receiver. Each of the Nd:YAG lasers generates 220 mJ of energy, to produce simultaneous pulses with a repetition rate of 20.16 Hz and a pulse length of 20 ns at 1064 nm and 532 nm. Due to the polarization out-coupling a highly polarized output beam is provided and the beam expanders reduce the angular divergence of the transmitted laser beam to produce a footprint of approximately 70 meters at the surface of the Earth.

After a laser pulse has been generated by the CALIOP laser, the receiver subsystem, which consists of a 1-meter telescope, measures light getting backscattered by molecules, aerosols and clouds in the atmosphere. As the different backscattering events span an intensity range of over six orders of magnitude, the on-board detectors need to have a high dynamic range plus high electronic gains to even detect the tiniest backscattering events. Therefore it is very important to prevent any direct solar illumination and stray light of the telescope with a light baffle. To reduce the solar background illumination even further, interference filters are located upstream of the detectors and additionally, to separate the 532 nm parallel and perpendicular returns, a polarization beamsplitter is located upstream of the 532 nm detector [Hunt et al., 2009, Winker et al., 2006]. The parallel and perpendicular channel together with the polarization beamsplitter is used to measure the linear depolarization of the backscattered return, allowing discrimination of the cloud phase and the identification of the presence of non-spherical particles such as aerosols.

Once the signal of one of the three channels hits the detector, several signal processing steps are performed. Before the analog signals are digitized, the background solar radiation is subtracted by an electrical offset. Each channel operates with two 14-bit analog-to-digital converters, set for different gains in order to measure weak and strong backscatter signals at the same time. The receiver electrical bandwidth and the laser pulse repetition rate of 20.16 Hz, determine the fundamental sampling resolution of 30 meters vertically and 333 m horizontally. To save on-board storage and limit the amount of data that has to be transmitted to the ground, CALIOP uses an on-board averaging algorithm. As the atmosphere becomes more spatially uniform with increasing altitude and signals higher in the atmosphere tend to be weaker, CALIOP adapts to this change

by different averaging schemes. For the lower troposphere with its vast spatial variability, it is important to retain the fundamental vertical resolution of 30 m. Whereas for higher altitudes, CALIOP averages 60 m vertically for signals coming from 8.2 km to 20.2 km and even 180 m for values ranging from 20.2 km to 30.1 km to achieve the required SNR and to reduce the data [Winker et al., 2006].

2.3 The Cloud Detection Algorithm of CALIPSO

While the outputs of the Level 1 algorithms are attenuated backscatter coefficient profiles with their respective uncertainties for all three channels, the Level 2 science data products use sophisticated algorithms (see Figure 4) for the feature detection of clouds and aerosols within the measured atmospheric column. These detection algorithms have been improved by Version 3.01 [Liu et al., 2010] and the previously overestimated cloud fraction, due to false classification, has been reduced. In this section the L2 routines and assumptions relevant to the results of this thesis are presented. The main benefits of using the L2 scientific data sets is the availability of the following key parameters:

- layer top and base altitudes with their descriptive properties
- layer type identification and separation into cloud and aerosol layer products
- cloud and aerosol backscatter and extinction coefficients

Having only the calibrated profiles of attenuated backscatter for the three different channels and ancillary geophysical and meteorological data such as surface elevation, temperature and pressure profiles of the atmosphere (provided by NASA's Global Modeling and Assimilation Office), the *Selective Iterated Boundary Locator* (SIBYL) detects layers in the provided Level 1 data. The only assumptions made for the detection algorithm are the correctness of the models and meteorological data provided by GMAO and the digital elevation model GTOP30. Due to the natural backscattering variation that spans several orders of magnitude for various detectable features in the atmosphere, the data is repeatedly scanned for layers with a profile scanner which uses a threshold method. The algorithm starts at the top in a clear air region and is using modeled data to compare them against the measured attenuated backscatter profile. Finally the latter is transformed into attenuated scattering ratios which can easily be processed by a computer. While dense clouds can be detected quite easily because of their strong SNR, thin cirrus or aerosol layers demand extensive averaging of the algorithm to increase the SNR and finally detect even weak layers. With dense clouds always dominating the profile with their strong backscatter signal they have to be removed prior to the

further averaging to ensure that even the faintest layers can be detected. Therefore the averaging process described by the ATBD (Algorithm Theoretical Basis Document) and even the improved version 3 [Liu et al., 2010], removes some highly reflective boundary layer clouds at the original FOV resolution in order to not completely dominate over weaker signals. Nevertheless, it is required to filter strong signals (e.g., stratus and fair weather cumulus) from the weak signals (e.g., thin cirrus clouds or aerosol layers) in the process of cloud and aerosol discrimination. SIBYL scans the profiles for strong features, extracts them with their properties such as transmittance and the layer-integrated signal backscatter and then proceeds with a coarser horizontal averaging scheme of up to 80 km horizontal extent. Those properties are used in the next processing step for the *Scene Classification Algorithms* (SCA).

Within this processing operation, layer descriptors are computed. These layer descriptors provide spatial, temporal and optical characteristics of the feature, as well as a classification of the feature as either being a cloud or an aerosol using the layer-mean (layer top to base) attenuated total color ratio X':

$$X' = \frac{\langle \beta'_{1064} \rangle}{\langle \beta'_{532}* \rangle}. \tag{12}$$

For this purpose the information of both 532 nm channels, the 1064 nm channel, their backscatter intensity and the depolarization information is used to discriminate the features even further (e.g., ice cloud or aerosol type). The process, however, is not entirely based only on the L1 data, but also on ancillary information from models and other observations. After the first classification, the cloud-aerosol discrimination (CAD) algorithm [Liu et al., 2010] is used to distinguish between a variety of layer classes. The cloud, aerosol and clear air lidar ratios are then handed over to the third and last processing step: the *Hybrid Extinction Retrieval Algorithms* (HERA).

Using SIBYL's and SCA's outputs, HERA performs extinction retrievals which produce profiles of particulate backscatter and extinction at both 532nm and 1064nm. Each detected layer is thereby checked for its properties, characteristics and surroundings. Additional parameters like optical depth, assumed particle size and depolarization ratio are computed and assigned to each detected feature (only for 5 km product). Because of the need for extensive averaging of some layers, the spatial resolution for each layer is different, just like the profiles of extinction and backscatter intensities. This effect is compensated by providing spatially uniform Level 2 data sets for the single shot resolution of 333 m, an averaged 1 km product and the 5 km product for which all processing algorithms are performed [Vaughan et al., 2006, Winker et al., 2006].

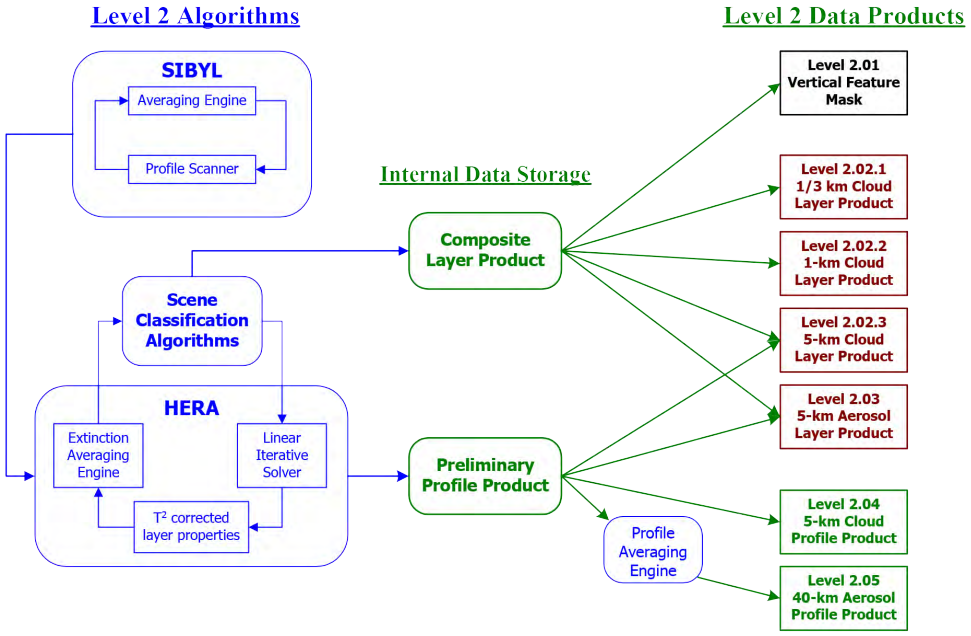


Figure 4: Level 2 Algorithms and their respective Data Products. [Adapted from Winker et al., 2006, p.21]

2.4 CALIPSO data sets

The three Level 2 data products are a *Vertical Feature Mask* (VFM), *Cloud and Aerosol Profiles*, and *Cloud and Aerosol Layer* products (see Figure 4). The VFM thereby includes scene classification data from the SCA and HERA as well as environmental parameters for the entire range up to 30.1 km with the best spatial resolution possible at each region but with only 5 km horizontal resolution. With the VFM a cloud and aerosol type analysis can easily be performed as the respective feature at each altitude is described by a number, which itself describes a cloud or aerosol type. This product was tailored to provide the scientific community with an easy possibility to analyze cloud and aerosol distributions and types on a global scale. However, the poor horizontal resolution prohibits the use for a performance assessment of a future lidar mission. Such an assessment needs more raw and genuine measurements rather than pre-processed data.

The *Cloud and Aerosol Profiles* contain averaged cloud (5 km horizontal resolution) and aerosol (40 km horizontal resolution) profile data as well as ancillary data. While the aerosol profile product uses a 120 m vertical averaging scheme up to 20.2 km, and 360 m up to 30.1 km, the cloud profile product is limited to 60 m vertical resolution and 20.2 km height. Polar stratospheric clouds (PSC) are therefore not considered in the cloud profile product. Therefore the data included in the Aerosol and Cloud Profile sets is

limited to the extinction cross-section, the backscatter cross-section and ancillary data. Again, as the purpose of this thesis is a performance assessment at the highest resolution possible, the data set limits prohibit the use for an analysis.

Instead, the *Cloud and Aerosol Layer* products are produced at three horizontal and vertical resolutions for clouds: 333 m horizontally and 30 m vertically, 1 km horizontally and 60m vertically and 5 km with 180 m vertical resolution. The aerosol layer product is once more limited to 5 km horizontal resolution. The components of such products are column properties, which include position data and viewing geometry of the column of the atmosphere through which a given lidar pulse travels, as well as layer properties like top and base altitudes, integrated attenuated backscatter, and optical depth (only for the 5 km product). The 333 m cloud products, which range from -0.5 km to 8.2 km, can distinguish up to 5 layers, while the 1 km product ranging from -0.5 km to 20.2 km can distinguish up to 10 layers. For each detected cloud layer, an integrated attenuated backscatter signal and the cloud top and base altitude within the column are included. As it is most unlikely that more than 5 individual cloud layers are present at 333 m or 1 km horizontal resolution, the discrimination is sufficient. However, there is an important difference in these three cloud products. While the 333 m and 1 km products are only a representation of the originally detected features at this horizontal and vertical resolution, the 5 km product is highly post-processed and averaged to find even fine structures in the atmosphere. Therefore the 5 km product is the most artificial representation of the atmosphere while providing most information, even on tiny features. It is also the only product for which optical depths, the two-way transmittance, several lidar and depolarization ratios and even cirrus shape parameters are provided. The same holds for the aerosol layer product.

As the VFM and the *Aerosol and Cloud Profile* products can not provide the required information for this analysis, it is necessary to familiarize with the characteristics and properties of the *Cloud Layer* products, as they are the only products applicable. Figure 5 summarizes the vertical and horizontal resolutions each product can provide up to their altitude. As already mentioned, the 333 m and 1 km products are closest to the original measurements performed by the lidar, and the 5 km product is the product with the most additionally provided information.

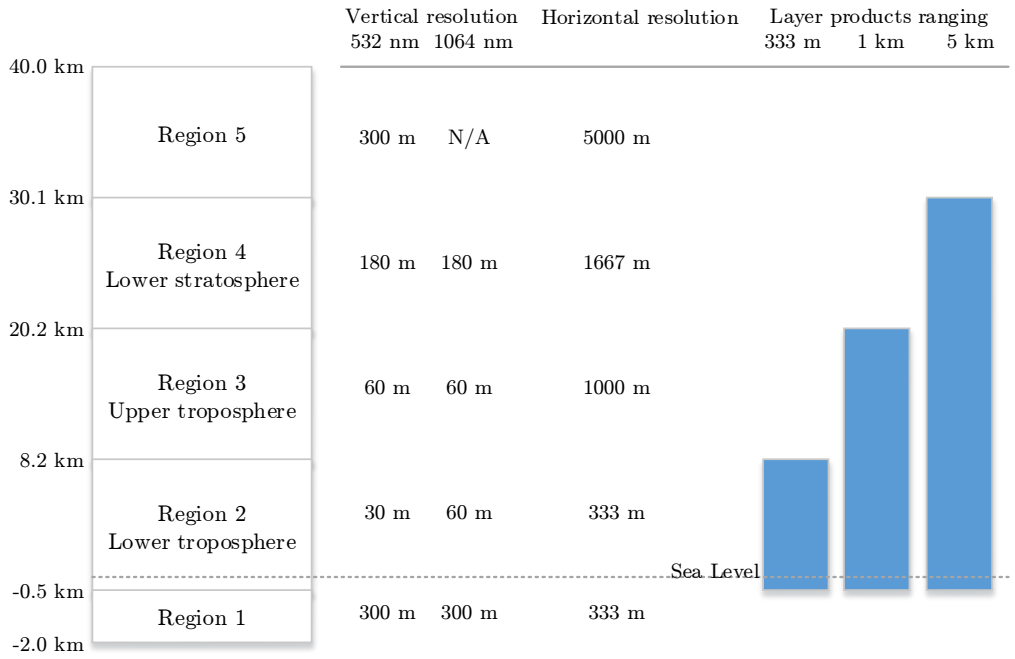


Figure 5: The properties of the L2 Cloud Layer data sets

To properly understand the differences of each product and to decide which product might be best suited for the needs of a performance assessment, data for all products were processed for the period of one month. The results are summarized in Table 3.

While the 333 m product with its 30 m vertical resolution provides only information about clouds in the lower troposphere up to 8.2 km altitude, the probability of detected clouds is very high and currently all layers detected at single shot resolution are automatically classified as clouds [Liu et al., 2010]. This is confirmed by the fact, that 37.6 % of all detected clouds (47.9 % of all measurements) are the only layer detected within the 333 m column penetrated by the laser. Additionally only 5.7 % of the measurements achieve a surface return although they are cloudy. This implies that, whenever a cloud was detected within the single shot column, it was almost completely attenuated by a dense cloud. Therefore almost no (only 7.2 %) additional layer was detected below the first detected cloud. Even though 52.1 % of the measurements were cloud free, only 29.3 % of the laser beams hit the surface and were successfully detected as surface returns. The question why the overall surface return performance is that bad for the single shot profiles, even though the surface spike has a magnitude much greater than the adjacent atmospheric return, can not be answered without further investigation (see Section 2.5).

	333 m	**1 km**	**5 km**
cloud free fraction	0.521	0.324	0.310
1 layer fraction	0.376	0.522	0.391
2 layer fraction	0.072	0.138	0.163
$\geq$ 3 layer fraction	0.031	0.016	0.136
surface return fraction	0.349	0.579	0.612
cloud free and surface return fraction	0.293	0.321	0.307
cloudy and surface return fraction	0.057	0.254	0.305

Table 3: Key features of the 333 m, 1 km, and 5 km product for the analyzed data set of January 2007

In contrast to the 333 m product, the 1 km product with its 60m vertical resolution contains a lot more cloudy measurements (67.6 %) while 52.2 % of them are single layered and 13.8 % are two layered. With a total of 57.9 % detected ground returns, the on-board averaging algorithms detect additional 23.0 % of lidar surface returns w.r.t the single shot data. The fact that on the one hand 32.1 % of all measurements are cloud free and return a surface signal while on the other hand 25.4 % are cloudy and still return a surface signal implies, that the 1 km product also consists of less denser clouds which do not completely attenuate the signal.

The 5 km product with its 180 m vertical resolution is capable of detecting even thinner clouds ranging up to 30.1 km and in the case of January 2007, a cloud in 26.4 km altitude was detected. While only 31.0 % of all measurements remain cloud free, the 1 layer fraction diminishes with only 39.1 % left. Consequently multi-layer detection increases with the 2 layer fraction raising to 16.3 % and 13.8 % of all measurements consist of even more layers. Looking at the surface returns, the on-board averaging of 1667 m only adds 3.3 % more lidar surface returns w.r.t the 1 km data, while the fraction between cloudy and cloud free measurements balances with 30.7 % to 30.5 %.

In summary the 333 m product provides the highest vertical and horizontal resolution with the most reliable information on prevailing dense clouds in lower altitudes that attenuate the lidar signal completely. This information is crucial for a performance assessment of a future lidar mission such as MERLIN. Due to the similar resolutions and characteristics of the mission (see Section 1.1.4), the 333 m product is the best possible representation of MERLIN's single shot behavior. The intermediate product with 1 km horizontal resolution is not appropriate for a detailed analysis, based on its limitation of 20.2 km altitude, 1 km averaging and lots of additional parameters retrieved by several algorithms that are only available for the 5 km product [Vaughan et al., 2006]. On the one hand, the 5 km product is the most artificial and coarse product compared to the 333 m and 1 km counterpart, but on the other hand it includes all cloud information

necessary for an analysis that represents not only the best possible representation for a spaceborne lidar system, but also a realistic composition of the atmosphere. Using the 333 m product with its prevailing low level dense clouds and combining it with selective clouds from its 5 km counterpart is in best accordance with the needs. (see Chapter 3 for further information)

2.5 CALIPSO lidar surface return

To understand the great discrepancy of the varying surface returns of the three products and to derive characteristics from the cloud free occurrence, the global distribution of both parameters has to be analyzed. While the detection of a feature spike within the atmospheric column and lidar profile is quite straight forward, the identification of surface returns is accomplished by using a high-resolution digital elevation map (DEM) and a model of the measured impulse response function of the lidar receiver. Due to the attenuation by overlying features, the intensity of the surface spike can be diminished. Therefore the absolute magnitude of the signal is of less importance whereas the signal shape becomes more evident. To be correctly detected as a surface return, the spike must have a shape characteristic for a surface return (defined by the receiver's instrument response function) and also has to occur in an altitude regime consistent with the DEM. The valid altitude range is thereby location-depended, relies on the uncertainties in satellite position and altitude as well as on the relative error of the DEM [Vaughan et al., 2006]. Taking these considerations into account, Figure 6 confirms the inability of CALIOP to retrieve surface signals from a bulk of the oceans, as well as from highly vegetated regions such as the South American northern- and eastern-, the Mid-African- and all Indonesian rain forests. Otherwise a surface return is almost always existent for icy regions and deserts like the Sahara. The underlying Normalized Difference Vegetation Index (NDVI) world map for January 2007 is a high resolution (0.1° x 0.1° degree) Terra Moderate Resolution Imaging Spectroradiometer (MODIS) representation of consistent, spatial and temporal global vegetation conditions (see `http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD13A2_M_NDVI` for more details). Due to low backscatter signals from the ocean and highly vegetated regions, the surface return detection algorithm might fail for a horizontal resolution of 333 m, but due to extensive averaging in the 5 km product a return can be acquired.

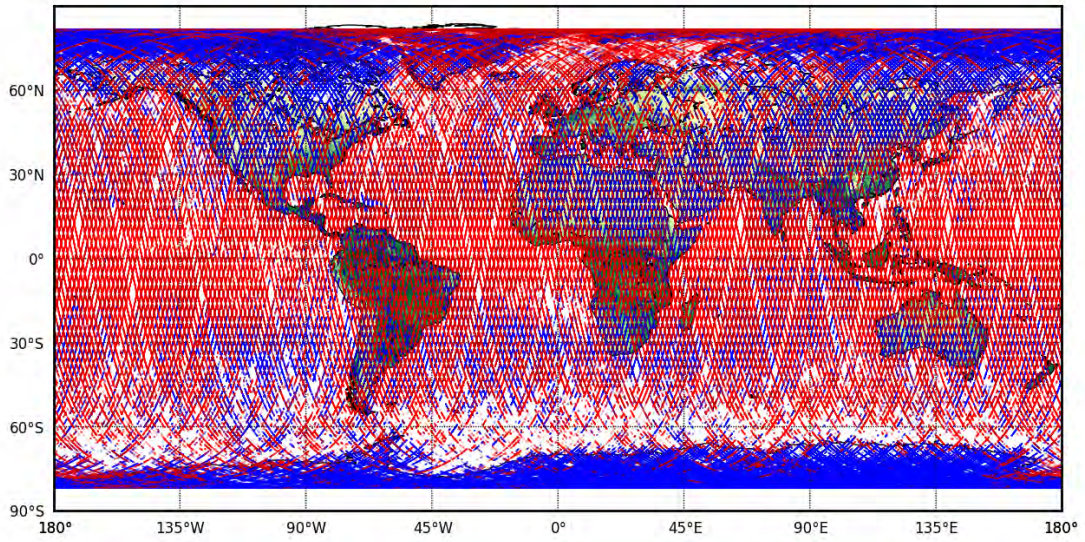


Figure 6: CALIPSO surface return of January 2007. 52.1 % of cloud free measurements with the blue markers indicate valid surface returns and red markers indicate no surface returns. The underlying NDVI vegetation world map of January 2007 was produced with data from the Terra MODIS instrument and is color-coded from yellow with almost no vegetation to green for regions with extensive vegetation.

The variable wave height for the oceans may even complicating the surface return retrieval. This is why the *Lidar Surface Elevation* can not be used in the further analysis. As correct information about surface hits of the laser beam are needed to perform a proper performance assessment and identify regions where MERLIN also has a high probability of surface returns and retrievals of methane, the *Lidar Surface Elevation* is completely neglected and the following assumptions are made:

- If no features are detected within the atmospheric column, a surface return is assumed
- If features are detected in the 333 m product, the probability of complete signal attenuation is very likely, and no surface return is assumed
- If features are detected in the 5 km product, the feature optical depth τ serves as additional discriminator of a surface return

In consequence the selected 333 m and 5 km data sets have to be merged partly which is further described in Chapter 3.

Chapter 3

Methodology behind cloud free and cloud gap statistics

3.1 Introduction

In order to obtain statistically significant data from the CALIPSO mission and to draw concrete conclusions for the aspired MERLIN mission, a software was developed as a part of this thesis. While by now there is no such analysis and statistic tool publicly available for the CALIPSO data set, the need for a fast and easy way to process the data is apparent. For this reason a Python program with a PyQt GUI (Graphical User interface) was developed. If not otherwise stated, the data sets used for the analysis and the development of the algorithms are *Level 2 1/3 km Lidar Cloud Layer v3.01* and *Level 2 5 km Lidar Cloud Layer v3.01* from 2007-01-01 00:22:49 until 2008-01-01 00:37:39. The year 2007 was chosen because of the minor losses of only 5% of data due to orbit maneuvers.

The way the CALIPSO data sets are stored by NASA and CNES does not allow an enhanced analysis for periods larger than a day without storing them locally, reprocessing them and compressing the original half orbit data sets to a new data set for the desired analysis period. With a required disk space of over 700 GB in total for the 333 m and 5 km product for the entire analysis year 2007, the first processing step has to focus on reducing the data set by deleting unnecessary parameters that do not contribute to the cloudless gap analysis (see Section 3.2). A next step is to use a gzip (GNU zip) compression to reduce the total required disk space to less than 20 GB for all monthly data sets of both products for the entire year.

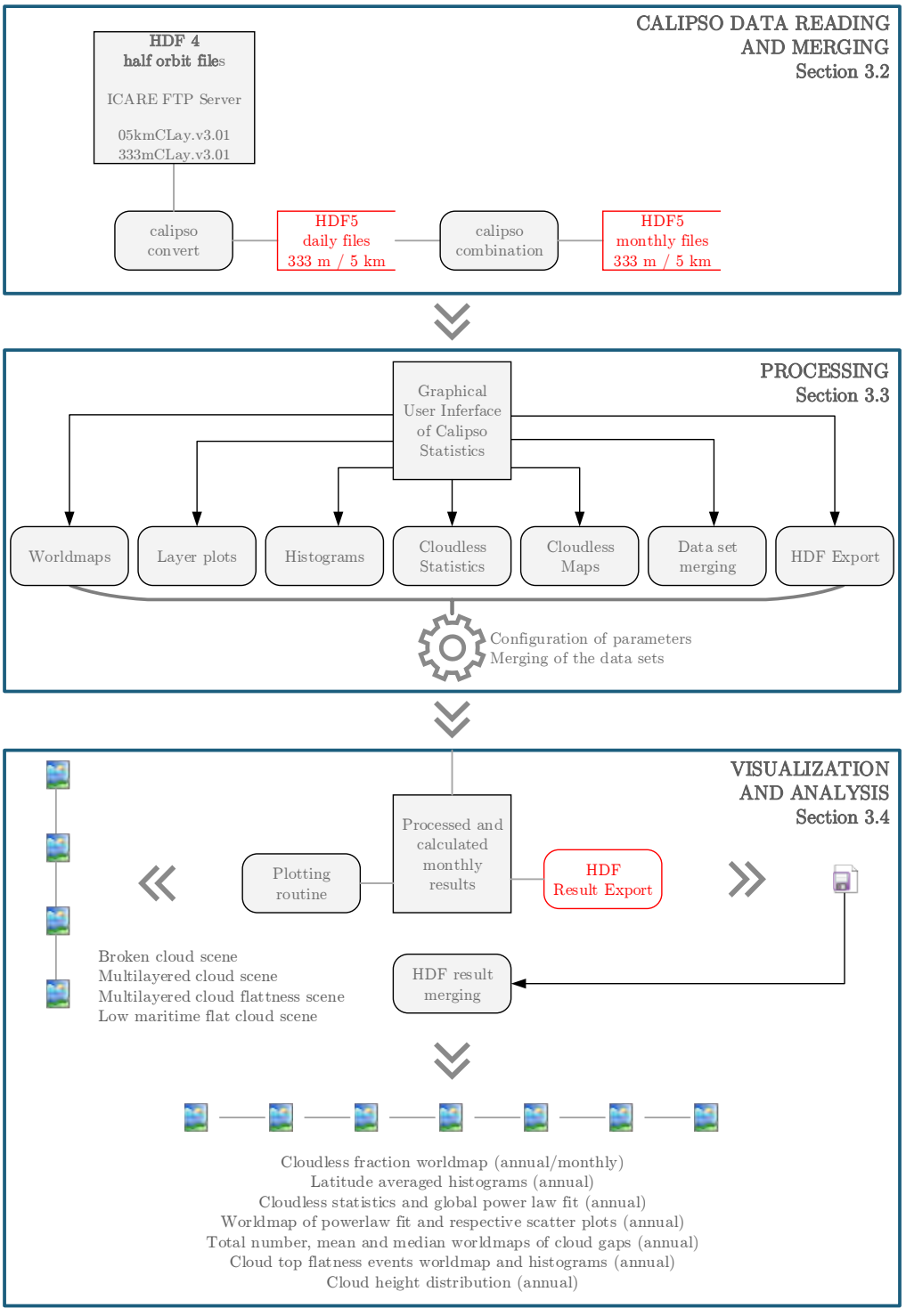


Figure 7: The diagram introduces all components of the developed software and the typical data flow beginning from the original HDF 4 file from NASA/ICARE to the final products presented in this thesis. Thereby red boxes indicate data outputs of the program.

While decoding of the compressed monthly data requires some processing time, the benefit of combined monthly files compensates the processing time needed for reading through all single half orbit files. After the preparation of the two data sets for the 333 m and 5 km resolution, the next step concentrates on combining the relevant cloud information. With the merged data set at hand, the processing can be performed entirely within the developed GUI, where several parameters can be set and 20 different plots and several statistical outputs allow a detailed analysis of all relevant aspects for this thesis. The detailed routines and algorithms are presented in Section 3.3.

In 3.4 the possible plots, their properties and context are discussed and some guidelines are given on how to use the plots and outputs of the program. While the public available pre-processed quick-look images do not allow a valuable analysis, the results of the software are highly optimized to analyze the CALIPSO data set for all relevant aspects of a future lidar satellite mission. While the handling of monthly combined data sets is not that challenging for a single computer, the analysis of the entire year can not be performed without several drawbacks. Therefore an export function was implemented to allow a further processing of the monthly intermediate results, instead of the whole merged data set.

In summary, three major processing steps have to be performed for the analysis of a larger period of CALIPSO data sets. All components of Calipso Statistics and their interaction are shown in Figure 7. The entire program was developed using the following packages and frameworks:

- Python 2.7.4 (`http://www.python.org/`)
- NumPy 1.7.1 and SciPy 0.12.0 (`http://www.numpy.org/`)
- Matplotlib 1.2.1 and Basemap 1.0.6 (`http://matplotlib.org/`)
- pyhdf 0.8.3 (`http://pysclint.sourceforge.net/pyhdf/`)
- h5Py 2.1.3 (`http://www.h5py.org/`)
- powerlaw 1.0 (`http://code.google.com/p/agpy/wiki/PowerLaw`) [Alstott et al., 2013]
- PyQt 4.10 (`http://www.riverbankcomputing.co.uk/software/pyqt/intro`)

3.2 Data Reading and Merging

The HDF4 (Hierachical Data Format version 4) was chosen as the primary file format for scientific and ancillary data for the CALIPSO mission of NASA. As a software library

that provides high-level APIs (Application Programming Interface) and a low-level data interface, it is also a platform-independent binary format with interfaces for FORTRAN-77 and C. Due to third-party libraries, e.g. for IDL or Python, almost any programming language can access the multi-dimensional arrays ('Scientific Data Set' or 'SDS') within the files. NASA decided to split up the data into half orbit files, for day and night orbits, following the name convention of Figure 8 and resulting in approximately 29 half orbits a day.

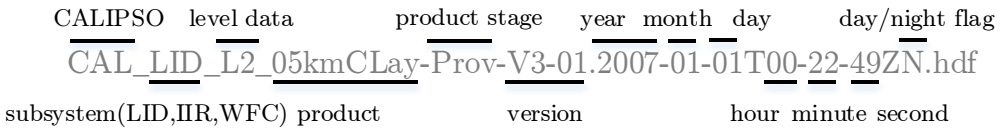


Figure 8: File name convention of the HDF4 files.

Unfortunately, the split sequence of NASA sometimes loses some measurements at the end of each half orbit which results in some dropped measurements approximately every 45 minutes. The system behind this behavior is not fully understood by now, as no official documentation points out this problem. Though it might be related to the problem of the sunlight effect mentioned by [Hunt et al., 2009]. Once the satellite goes into sunlight, both detectors are affected differently by a thermally induced misalignment. As stated, this problem was corrected by re-calibration and the data products were adjusted with the Release v2.01. But looking at the original data from the release v3.01, some measurements are still missing. While analyzing much larger periods (e.g. a month) this phenomenon shows up looking at a normed histogram (see Figure 9) of all measurements. Between $70\,°S$-$50\,°S$ and $50\,°N$-$70\,°N$ latitude, small deviations from an idealized polar orbit can be observed, which originates in missing measurements at the end of some half orbits. These negative spikes of the histogram vary seasonally and sometimes never appear. The consequence of this phenomenon is a slightly biased cloud gap statistic, as a consecutive cloud gap could be interrupted by missing measurements in this region. The peaks at ± 80-82 ° can be explained by the polar orbit of the satellite and the accumulation of measurements as a result of the turning point of the satellite track around the poles. Therefore around 7.5 % of all measurements are located around the poles which could lead to a bias of the statistics in the polar regions. However, the only statistic directly being influenced by this bias is the number of total cloud gaps, as all other statistics are either independent of the geographical position or represent statistics for single grid points.

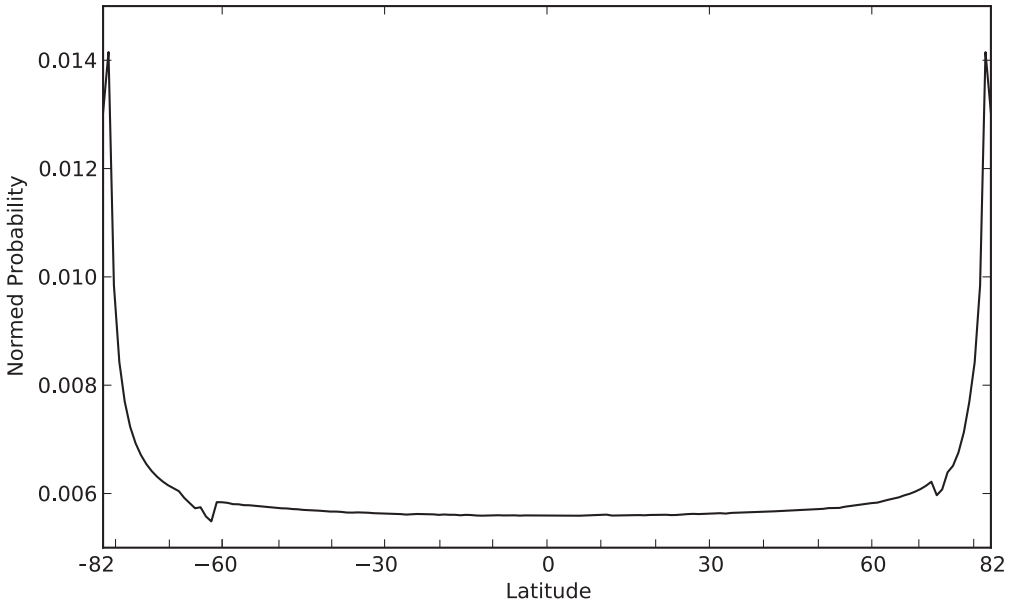


Figure 9: The latitude averaged integral normed histogram of January 2007 shows deviations from an idealized histogram due to the half orbit splitting routine.

The HDF4 single half orbit files are converted into a combined HDF5 daily file, while already reducing the amount of data by selecting only the important parameters (see Table 11 and Table 12 in the Appendix) required. To further reduce the data amount, the built in compression mode gzip of HDF5 and the Python library h5Py are used. For floating point scientific data, utilizing the gzip compression algorithm can significantly improve the compression ratio while it takes insignificant extra processing time to decompress the data. Fortunately h5Py uses straightforward NumPy and Python metaphors while still guaranteeing platform-independence. Thus for one day of the L2 Cloud Layer product the required disk space reduces from 1.6 GB to 51 MB for the 333 m product and from 350 MB to 7 MB respectively for the 5 km product.

The conversion is done by the *calipso convert* routine and has to be performed for all data sets that have to be analyzed further, because the main program Calipso Statistics only supports converted data sets and the HDF5 format. To obtain statistically relevant results it is necessary to have a look at even larger periods and therefore the routine *calipso combination* allows a combination of several daily files to one larger file. This step is necessary because Calipso Statistics is programmed in a way to only allow one file to be loaded at a time. This restriction is necessary to reduce the processing time once the single file was loaded, because Calipso Statistics allocates the required system memory after loading the file. Additionally, due to the combination of the files the

timely order of the single measurement is ensured which is an important prerequisite for the analysis of consecutive clouds and cloud gaps.

While the size of a file is arbitrary, a fast processing of the algorithms used for the analysis is only possible if the limit of combined files was set to one month. With approximately 50 million individual measurements, the software needs almost 4 GB of user addressable memory space on a 64 bit system if all calculations are performed. In case that only a 32 bit system is available to run Calipso Statistics, the file limit is even less. To circumvent this restriction and allow the processing of annual data sets, a HDF export routine was implemented that enables the user to store the intermediate results on hard disk. As a side effect, the acquired results can also be processed in other environments like IDL or MATLAB.

3.3 Processing and algorithm

After having converted and combined all files to be analyzed, the GUI of Calipso Statistics is used to configure and set all parameters for a detailed analysis of the data set. To document all calculations, parameters and results produced by Calipso Statistics, this section introduces all algorithms in detail.

Data set merging algorithm

As the 333 m product is limited to a top altitude of 8.22 km while providing the best vertical and horizontal resolution of 30 m and respectively 333 m, the merged product is always based on the 333 m product. Unfortunately, the 333 m product only includes the densest clouds where the laser signal gets almost fully attenuated. Less dense clouds (found through averaging) and clouds above 8.22 km can only be found in the 5 km product(see Section 2.4 for more details). In order to include additional relevant clouds of the 5 km product, a routine checks every detected cloud layer of the 5 km product (Layer Top Altitude) if its $z_t \geq 8.22$ km and if the corresponding $\tau_{col} \geq 0.5$ (*Column Optical Depth Cloud 532*).

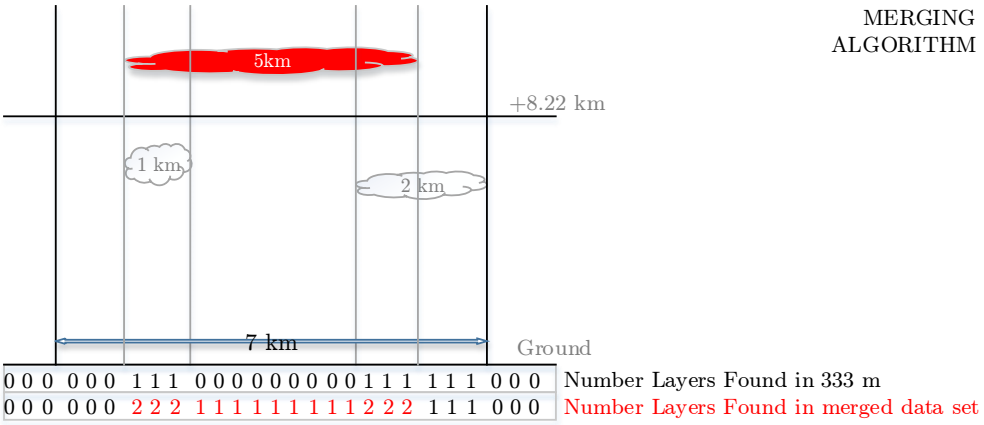


Figure 10: The 333 m Number Layers Found array gets larger if a cloud layer is found with $z_t \geq 8.22$ km and if the corresponding $\tau_{col} \geq 0.5$.

The threshold was chosen in a way to guarantee an acceptable measurement precision of the future MERLIN mission [Kiemle et al., 2011]. In addition, the cloud detection algorithm works only correct for clouds with $\tau \geq 0.45$ [Karlsson and Johansson, 2013]. Therefore, optically thin aerosol layers, even though they can have $\tau > 0.5$ are neglected as their corret detection is questionable for a horizontal resolution of 333 m. However if a single shot feature with $\tau > 0.5$ has been detected it is most likely classified as a cloud and therefore included in the merged data set. The same holds for optically thin clouds below 8.2 km with $0.5 < \tau < 1.0$. If this condition is met, the previously cloud free (*Number Layers Found*) entry in the 333 m product is raised by the amount of layers found and from there on appears as cloudy (see Figure 10). The *Column Optical Depth Cloud 532* parameter is the sum of all detected optical depths of each feature detected in the complete 5 km column. Due to the limited resolution of 5 km horizontally and 180 m vertically as well as to a different cloud retrieval methodology, the merging of both data sets cannot consider any cloud gaps smaller than 5 km for columns where a cloud above 8.2 km was added. This effect can bias the occurrence frequency of cloud gaps as multiples of 5 km seem to appear more often (see Figure 21). But the benefit of the single shot resolution up to an altitude of 8.22 km circumvents this drawback, especially for the cloud flatness analysis and high resolution statistics. Analyzing the merged data set (see Table 4) for a limited test period of one month reveals the composition of the merged product. While the overall cloud free fraction diminishes from 52.1 % to 40.6 %, 11.5 % clouds are added from the 5 km product, and 3.7 % of the 5 km columns previously described as being cloud free are now set to cloudy. The less transmissive layers of the 5 km product occur both over land and ocean, whereas the previously cloud free 5 km profiles occur mainly over land (e.g., Sahara, South China).

Cloud free fraction	Percentage
n_{cf333m} only	52.14 %
n_{cf333m} AND $n_{cfmerged}$	40.64 %
n_{cf333m} OR $n_{cfmerged}$	11.50 %
n_{cf5km} only	31.00 %
n_{cf5km} AND $n_{cfmerged}$	27.34 %
n_{cf5km} OR $n_{cfmerged}$	3.66 %

Table 4: Cloud free fraction part for the validation study of January 2007, with n_{cf333m} as the cloud free measurements in the 333 m product, n_{cf5km} as the cloud free measurements in the 5 km product, $n_{cfmerged}$ as the cloud free measurements after the merging process. The different logical combinations of individual data sets measurements reveal the cloud fraction prior and after the merging process for all 49,974,480 measurements.

The overall performance of the simple threshold merging algorithm is satisfying and the merged product represents all clouds that would have also been detected by MERLIN.

Probability distribution function (PDF):

Within this study a detailed analysis on the occurrence frequency of cloud gap lengths was performed. As this is by definition a discrete probability distribution function with a discrete random variable X and x being only integer values, the PDF is given by:

$$p(x) = Pr(X = x). \tag{13}$$

Cloud free fraction and cloud gap length: (as Figure 16, 23 and 24)

The individual cloud gap length for each consecutive sequence of cloud free measurements is gained through calculating group lengths for all consecutive sequences of uniform layers and checking whether or not a layer was detected for the group length. A group is thereby defined as a consecutive uniform row of measurements. For this purpose *Number Layers Found* is scanned for the occurrence of cloud free groups, and the midpoint in rounded latitude and longitude degree of each found cloud gap, is assigned to its respective grid point. For this study, the grid size was chosen to be 2 ° x 2 ° in latitude and longitude, because a higher resolution of the grid would cause data gaps due to the equatorial spacing of the CALIPSO groundtrack of approximately 172 km at the equator. As only the midpoint of each cloud gap is decisive for the assignment to its corresponding grid point, the algorithm tends to bias the distribution of cloud gaps exceeding the respective grid size. While degrees of latitude are almost parallel and only vary little

due to the Earth's oblateness from 110.57 km at the equator to 111.69 km at the poles, degrees of longitude gradually shrink from 111.12 km at the equator to zero at the poles [Karney and Deakin, 2010]. With 15.46 km, the smallest grid size longitude distance is determined at 82 °N/S. But with approximately only 2.7% of all cloud gaps exceeding 31 km or the 2 ° grid limit at this latitude (see Figure 21 and Table 5) the bias effect is negligible. Consequently, the amount of individual cloud gaps differs for each grid point and the data set to be analyzed has to be large enough to be statistically significant. While generally each cloud gap length can be assigned to any grid point with this method, the visualization of cloud gap lengths on an equally spaced world map with a 2 ° x 2 ° grid is always problematic. In contrast to the midpoint algorithm which is

	1 °	2 °	error at 2 °
at the equator	111.12	222.24	0.3 %
± 30 °	96.23	192.46	0.4 %
± 60 °	55.56	111.12	0.7 %
± 82 °	15.46	30.92	2.7 %

Table 5: Longitudal surface distance in kilometers between two points at selected latitude circles. The worst case error is assumed to be the probability of clouds exceeding the longitudal spacing of the grid point.

searching for consecutive cloud gaps along the ground track, the mean cloud fraction is computed differently. Again a 2 ° x 2 ° grid is chosen and all measurement positions are rounded to their next even integer. With the measurements assigned to the grid point, *Number of Layers* is checked whether the measurement is cloud free or cloudy. The cloud free fraction of each grid point is consequently calculated by dividing all cloud free measurements by the total amount of measurements for this point. Both algorithms are depicted in Figure 11.

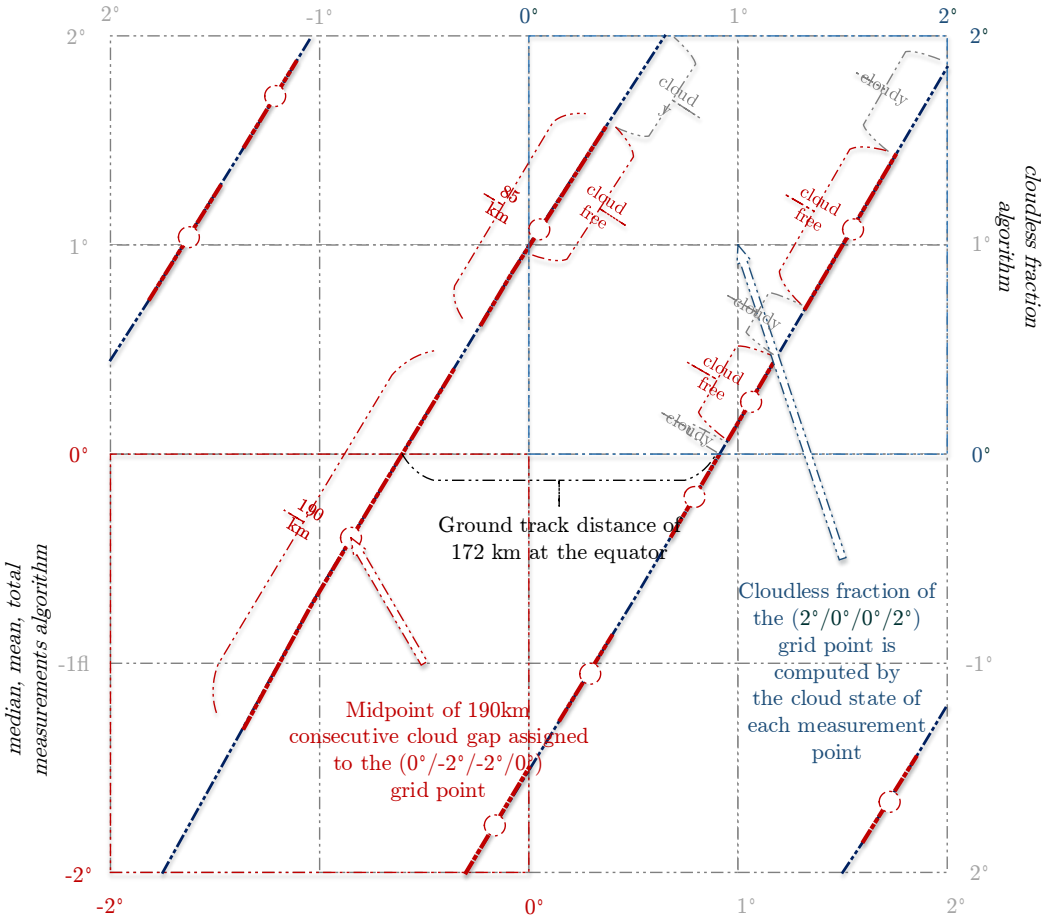


Figure 11: A diagram, which consists of a 2 ° x 2 ° grid in latitude and longitude, some typical ground tracks of the CALIPSO satellite indicated as solid lines with red sections of cloudless measurements, as well as the respective midpoint of one cloud gap highlighted by a red circle. For the cloudless fraction calculation, each independent measurement is checked whether it is cloudy or cloudless, rounded and assigned to its closest grid point (blue grid box). For the cloud gap length calculations, the consecutive length of a cloud gap is calculated, the midpoint is determined and the complete cloud length is assigned to the grid point (red grid box).

3.4 Visualization and Analysis

Generally, the visualization of all graphical outputs is performed within the graphical user interface of Calipso Statistics. Within the program the loaded data set can be analyzed graphically and by the computed and provided statistical output data in the command line. Besides several world map representations, layer plots and their histograms can be used to analyze single scenes of interest (as provided in Chapter 4). It is also possible to scroll through the data set and follow the ground track of the CALIPSO

data while being provided with layered representations of the cloud layers (also with regard to the merged data set), the lidar surface elevation and the digital elevation model. While this function provides the same information as the quick look images of NASA, additional histograms are automatically computed. Furthermore, several parameters can be adapted to the needs of the scientist. Probability density functions and related statistics can be performed in the graphical user interface and important results are provided numerically. The cloud gap distribution with its world map representation for the mean, median and the total number of cloud gaps can also be processed within the program, while the grid size can be adapted according to the desired resolution. The data sets, which are supposed to be exported to an external HDF5 file are also determined graphically and can be set individually. Generally, all plots are scaled to fit the fixed resolution of the program of 1600x900 pixel while the canvas, all plots are drawn to, is fixed to 1180x720 pixel. By default all generated plots can be exported to a PDF file, which keeps the vector data in the same scaling as the program. Again it is worth mentioning that almost any important intermediate result of the performed computations can be exported to an external HDF file and the statistical information within the command line can simply be copied. This can be used to adjust the output to individual preference or for the combination of several intermediate results. In summary, Calipso Statistics provides all necessary features for a detailed analysis of the CALIPSO Level 2 Cloud Layer products, while still being easily expandable as it only uses standard Python libraries.

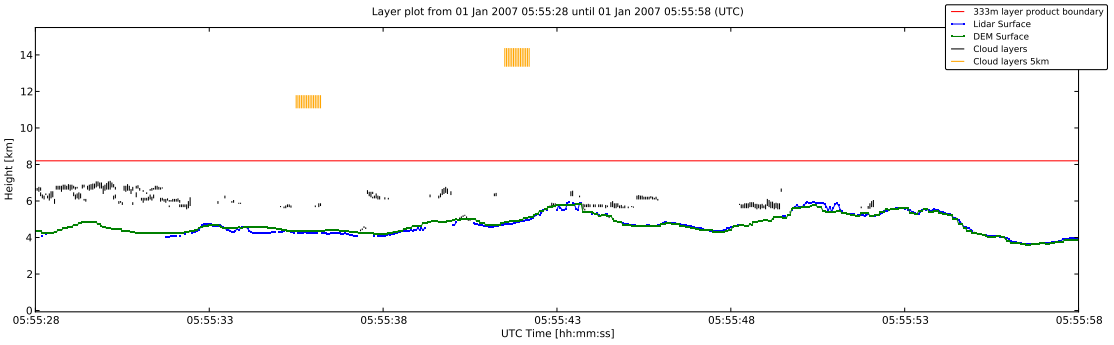
Chapter 4

Case studies

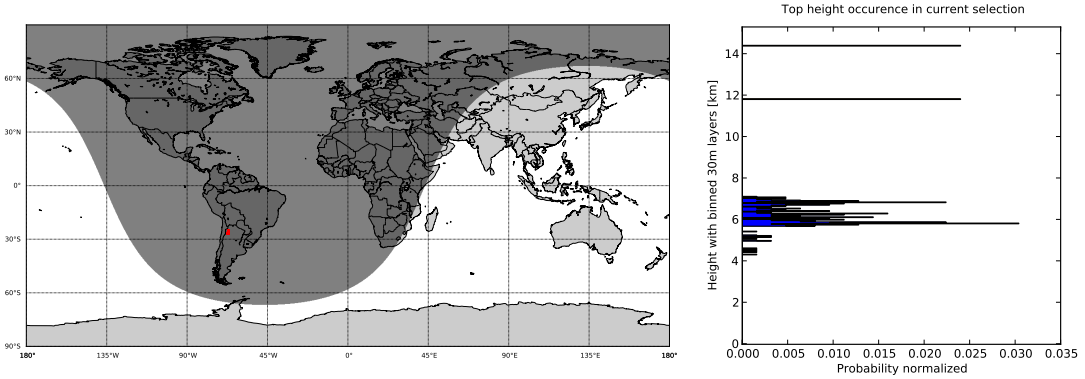
While the results presented in Chapter 5, 6 and 7 are calculated for one year to be statistically significant, the constituents of these statistics, i.e. the single clouds and cloud gaps that contribute to the different results, are blurred. To circumvent this problem and to get a feeling for the different occurrences and results of the merging procedure, typical cloud scenes are presented. Therefore the layered product was used to visualize a 200 km horizontally slice through the atmosphere measured by CALIPSO. As already mentioned in Chapter 1 and 3, the assumption was made that any cloud detected by CALIPSO without additional averaging and post-processing of the data, is most likely optically dense enough to cause a full attenuation of the MERLIN laser signal. Consequently, the cloud CALIPSO detected marks the point where the methane content retrieval of MERLIN would encounter an insufficient SNR.

For simplicity and to visually stand out, clouds found in the single shot product are indicated in black, where the vertical lines represent the vertical extent of detected clouds. The green line, if visible, signifies the underlying DEM, whereas the blue line represents the lidar surface return if actually present. Orange vertical lines represent clouds found in the 5 km product that meet the merging condition of $z_t \geq 8.22km$ and $\tau \geq 0.5$. Due to the detection process, they sometimes seem to dominate the scene, but that is owed to the coarse vertical resolution of 180 m compared to the 30 m of the single shot product.

4.1 Typical broken cloud scene



(a)



(b) **(c)**

Figure 12: A typical broken cloud scene above the Argentinean Andes (b) with small clouds and cloud gaps complemented by small clouds from the 5 km product(a). The top height distribution is scattered (c) and the surface return deviates from the DEM.

The cloud scene described in Figure 12 is a very typical representation of broken cloud occurrence in high altitude mountain ranges like the Andes. With only few cloud gaps exceeding 10 km length, it also represents a good validation of the global cloud gap frequency statistic. Given that the scene was captured during nighttime, when the detection sensitivity is higher, the layers found by CALIOP are most likely to be a real world representation. Although there is no cloud present, the inability of CALIOP's surface return detection is visible. The histogram not only shows that just few clouds are present but also indicates that the top height distribution of the clouds is quite scattered. Due to the merging process, two small clouds with 5 km vertical extent were added to the scene and the cloud free detection algorithm described in Section 3.3 changes the *Number Layers Found* parameter accordingly.

4.2 Multilayered cloud scene

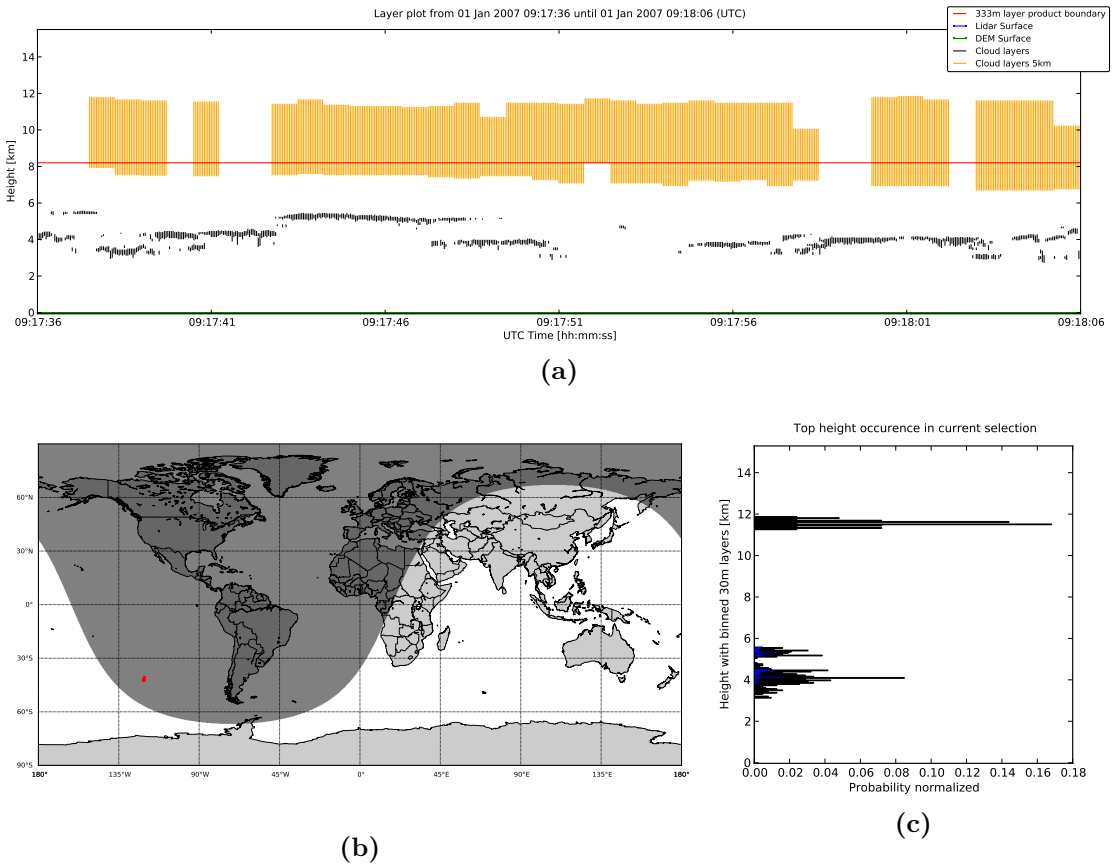


Figure 13: A typical multi-layer situation above the South Pacific Ocean with no surface return (b) and with at least three different distinct layers (a) at 11.8 km, 5.6 km and 4.1 km height (c). The dominance and vertical extent of the cirrus cloud is evident whereas the optical depth is quite low.

The cloud scene described by Figure 13 is a typical representation of a vertically scattered multi layered sky cover combined with a relatively dense cirrus ranging over more than 3 km vertical extent from the 5 km product. The scene was captured at nighttime above the South Pacific. If a threshold of $\tau \leq 0.5$ had been chosen for the merging process, the overlying cirrus cloud would not only have dominated the whole scene, but it would have covered the entire 200 km completely. The fact that a cirrus cloud at that height with an optical depth of $\geq$0.5 and more than 3 km vertical extent is present and CALIOP still can detect two additional separated cloud layers, could be quite challenging for MERLIN. However, looking at the horizontal averaging parameter confirms that the cloud was detected at the maximum averaging step of 80 km and therefore is optically very thin. Nevertheless the entire scene is assumed to be totally cloud covered.

4.3 Multilayered cloud flatness scene

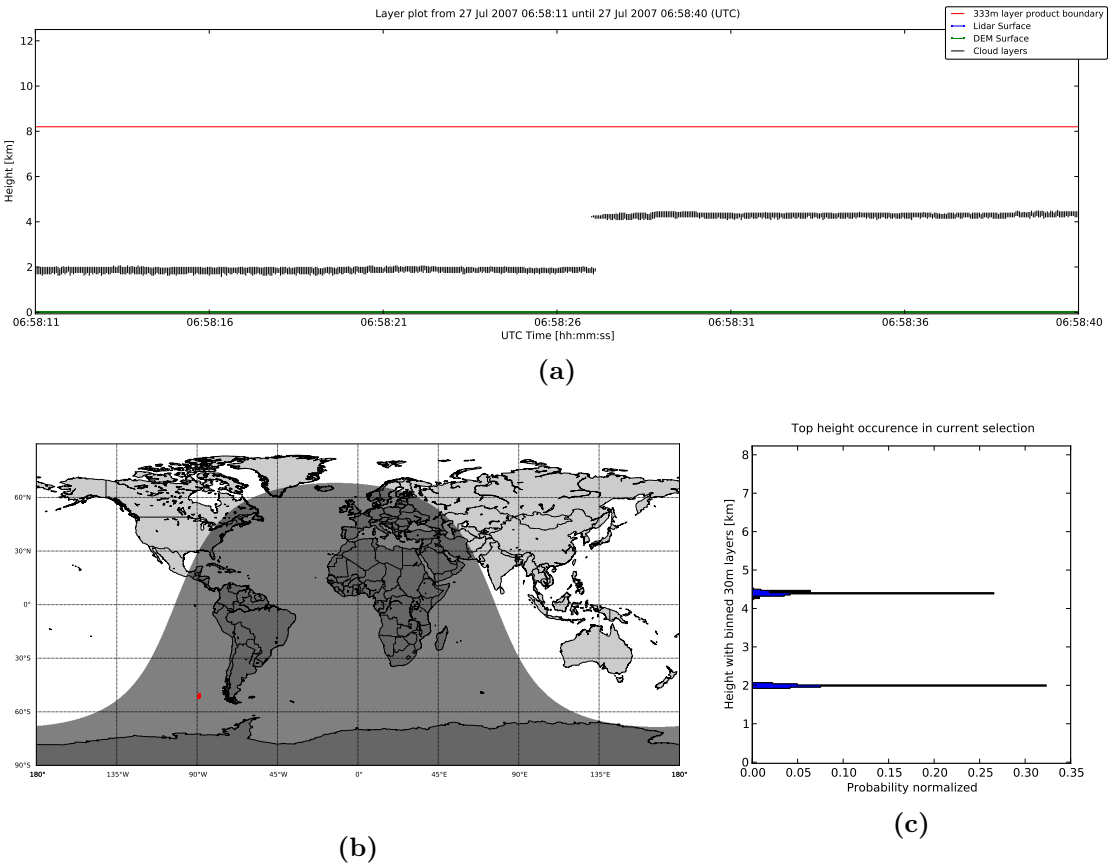
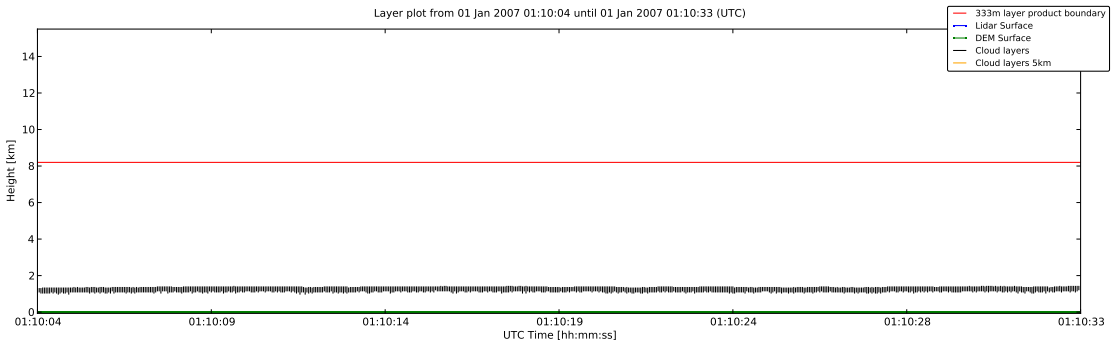


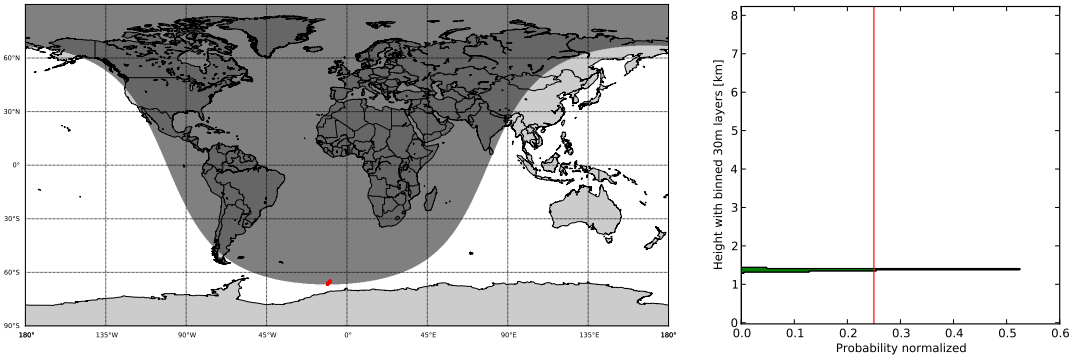
Figure 14: Two well separated 50 km consecutive clouds with a smooth cloud top (a), with cloud top heights at 1.98 km and 4.38 km (c) over the South Pacific Ocean (b). This clear separation of about 2 km is very rare and has only occurred once in 2007.

The scene described by Figure 14 is uncommon and therefore interesting. The analysis in Chapter 6 deals with the cloud top flatness. Here two consecutive cloud layers with only 30 m variation in their respective cloud top altitudes are clearly separated by 2.4 km. Unfortunately the scene takes place over the South Pacific Ocean and therefore it is not that relevant for a future methane retrieval. This specific case with a vertical separation of more than 2 km is unique for the analyzed period, but a few other cases with less vertical separation can still be found in the data set.

4.4 Low maritime flat cloud scene



(a)



(b) **(c)**

Figure 15: A typical low maritime cloud in proximity to the antarctic continent captured at nighttime (b) and without any overlying clouds from the 5 km product. The cloud situated at exactly 1.38 km altitude (c) is representative for all clouds discussed in Chapter 6.

In contrast to the rare scenery of Figure 14, the scene described by Figure 15 is representative for low maritime clouds with a smooth cloud top. Once again, the cloud top does not exceed 30 m vertical variation for over 105 km within a 200 km search window. With a top altitude of about 1.38 km, no surface return detection and its geographical position, the scene corresponds perfectly to clouds discussed in Chapter 6. Furthermore it confirms the ability of CALIOP and the single shot product to detect flat clouds with the highest vertical resolution possible.

Chapter 5

Cloud free statistics

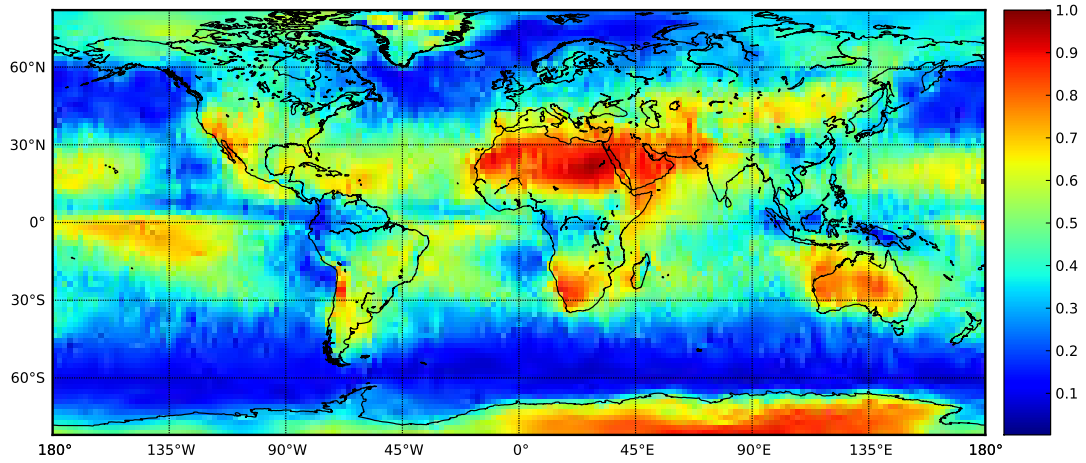
5.1 Global cloud free fraction

A number of statistics have been computed to analyze the atmospheric state observed by CALIPSO during the entire year 2007. All results presented in the following chapters rely on the merged data set and algorithms described in Chapter 3. In particular, the merged data set consists of clouds with a 333 m horizontal resolution below 8.2 km and of clouds with $\tau \geq 0.5$ above. In total, twelve monthly data sets (see Table 6) were combined to derive the annual statistics presented in this chapter.

2007	333m	5km	measurements	start time	end time
Jan	47.85	+11.49	49,974,480	2007-01-01 00:22:49	2007-02-01 00:27:51
Feb	46.89	+11.37	48,468,480	2007-02-01 00:27:51	2007-03-01 00:09:10
Mar	46.07	+11.13	50,118,720	2007-03-01 00:09:15	2007-04-01 00:16:07
Apr	46.44	+11.77	46,537,920	2007-04-01 00:16:10	2007-05-01 00:31:38
May	47.53	+12.15	53,056,560	2007-05-01 00:31:42	2007-06-01 00:42:38
Jun	46.93	+12.66	51,815,280	2007-06-01 00:42:49	2007-07-01 00:02:35
Jul	46.66	+12.58	49,554,000	2007-07-01 00:02:24	2007-08-01 00:10:05
Aug	48.11	+12.19	53,900,160	2007-08-01 00:10:15	2007-09-01 00:10:36
Sep	49.01	+12.10	51,956,160	2007-09-01 00:10:38	2007-10-01 00:19:19
Oct	49.49	+11.95	49,197,840	2007-10-01 00:19:24	2007-10-31 21:07:16
Nov	49.22	+11.65	50,253,840	2007-11-01 02:03:59	2007-12-01 00:34:57
Dec	47.03	+11.71	50,495,520	2007-12-01 00:35:03	2008-01-01 00:37:39
Mean 2007	47.60	+11.90	605,328,960	2007-01-01 00:22:49	2008-01-01 00:37:39

Table 6: The analyzed data set consists of Level 2 Cloud Layer data from 01 January 2007 00:22:49 UTC time to 01 January 2008 00:37:39 UTC time with a total of 605,328,960 measurements. The percentage of the global cloud fraction for the 333 m product only is given under 333 m for each month. Additionally the upper-level global cloud fraction added to the data set can be found in the column titled '5km'.

(a)



(b)

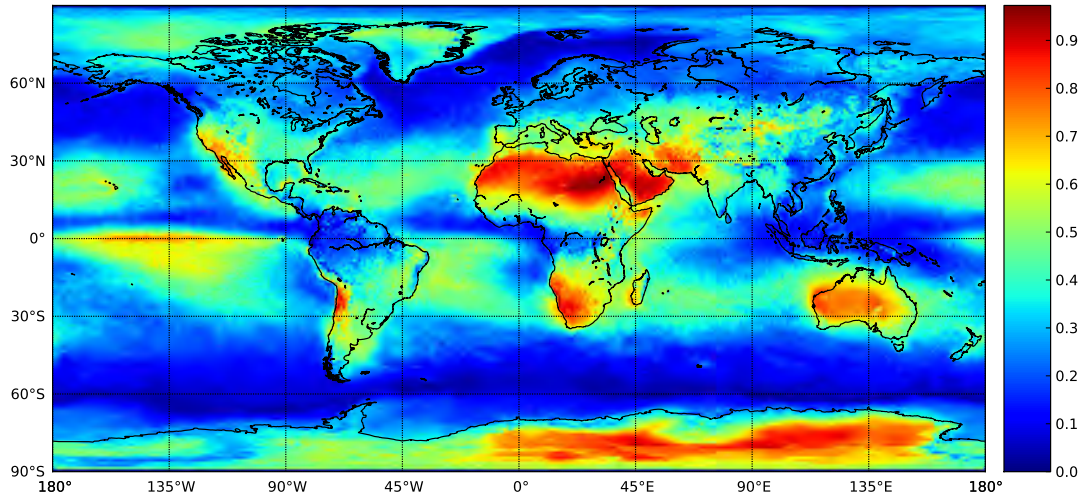


Figure 16: In (a) the global cloud free fraction was computed for 2 ° x 2 ° grid points from January 2007 to December 2007 for 6.05 x 10^8 individual measurements of the merged CALIPSO data set. In (b) the global cloud free fraction was inverted from 1 ° x 1 ° Aqua MODIS cloud data. Typical cloudless regions like the Sahara, the Arabian Peninsula, South Africa, parts of Australia, parts of Antarctica, nearly all high altitude mountain ranges as well as oceans in the tropics and subtropics are mainly cloud free on an annual average. The cloud free fraction is color coded from 0 being totally cloudy and 1 being totally cloud free. Global cloud free fraction averages vary from 40.5 % for CALIPSO to 32.6 % for MODIS.

With about 6.05 x 10^8 individual single shot measurements and a mean cloud free fraction of 40.5 % the data set covers all seasonal and atmospheric conditions a space-based lidar like MERLIN can encounter. The global variability of total cloud coverage of the Earth is limited to only a few percent as can be seen in Table 6. Seasonal and geographical variations have to be assessed differently. This is why the cloud free fraction of 2 ° x 2 ° grid points was computed with the algorithm described in Section 3.3. On the one hand the annual mean cloud free fraction was computed in Figure 16 and on the

other hand the seasonal variation is presented in Figure 17. Additionally, to validate the results of the derived cloud free fraction, 1 ° x 1 ° Aqua MODIS cloud cover data was inverted and compared against the CALIPSO data.

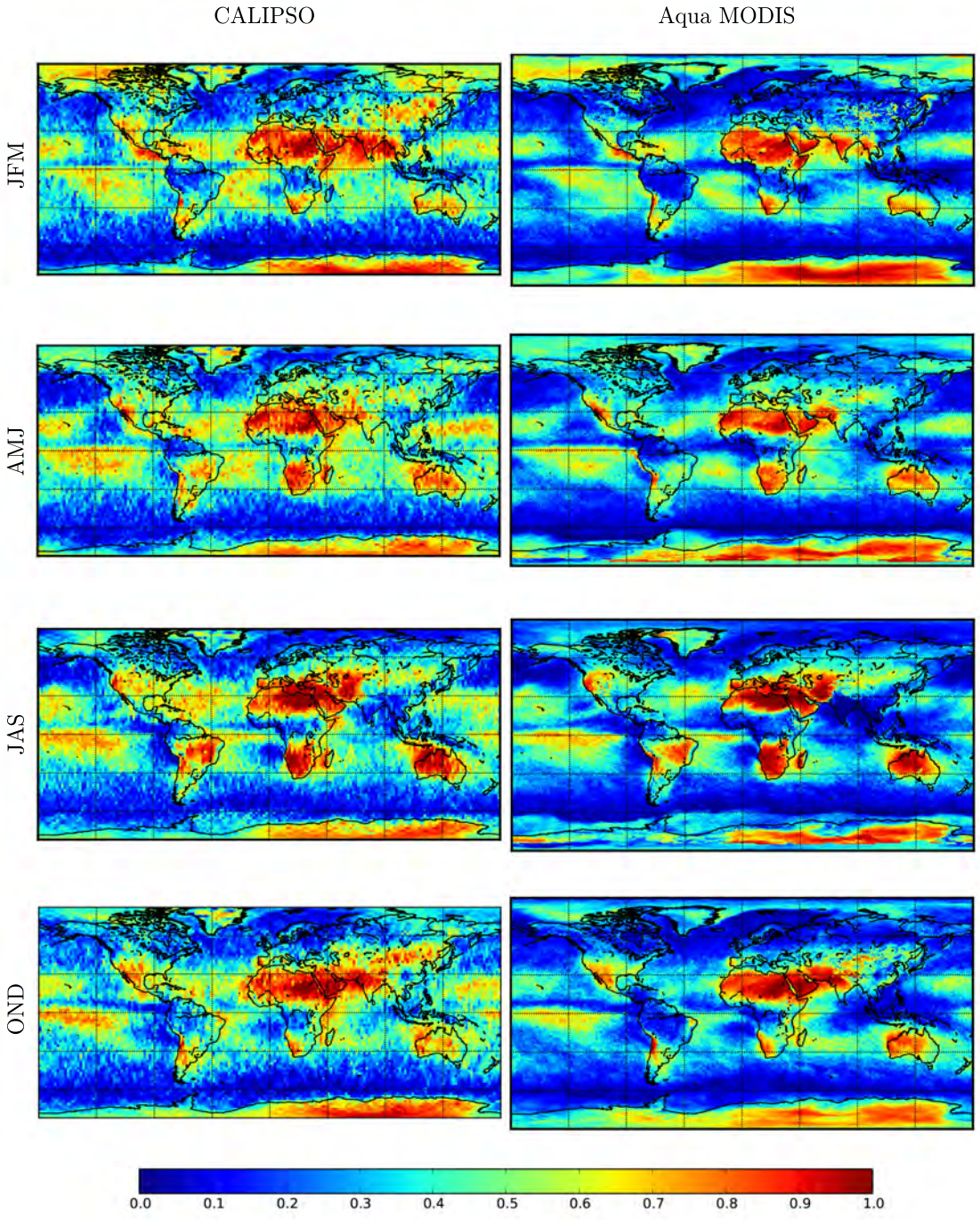


Figure 17: Global cloudless fraction from January 2007 to December 2007, for different seasons JFM, AMJ, JAS and OND. The left side represents the global seasonal cloud free fraction computed for 2 ° x 2 ° grid points and the merged CALIPSO data set whereas the right side represents the global seasonal cloud free fraction computed for 1 ° x 1 ° grid points using the Aqua MODIS data. The cloud free fraction is color coded from 0 being totally cloudy and 1 being totally cloud free.

As part of the A-train constellation, Aqua is only 93 kilometers - about 12.5 seconds - apart of CALIPSO which eased into the train behind Aqua in 2006. While Aqua and CALIPSO pass south to north over the equator in the afternoon, Aqua MODIS views the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands at a horizontal resolution of 250 m. Given that the temporal separation of both instruments on board of the the satellites is minimal, the clouds observed by both satellites are comparable even though the cloud detection is accomplished differently for active and passive instruments. Consequently the major cloud free regions like the Sahara, the Arabian Peninsula, South Africa, parts of Australia, nearly half of Antarctica, as well as all high altitude mountain ranges and oceans in the tropics and subtropics are cloud free in both data sets. Additional local cloud free maxima are found at the west coast of North and South America around the latitudes of 30 °S and 30 °N and Greenland and Antarctica. In accordance with [Wu et al., 2011], the most frequent cloudy areas are the latitude bands in the southern hemisphere between 50 °S and 70 °S and the ocean area in the northern hemisphere between 50 °N and 70 °N. However, the coarser resolution of the CALIPSO data set blurs fine structures (e.g., Indonesia, Siberia, oceans around the equator) visible in the Aqua MODIS data set. Thus the mean global cloud free fraction (in Figure 16) of the CALIPSO data set of 40.5 % is reduced to 32.6 % in the Aqua MODIS data set (see Table 7 for different latitude zones). This deviation can be explained to some extent by the different grid resolutions and by the different cloud retrieval methods used for active (CALIPSO) and passive (MODIS) remote sensing instruments. Due to the difficulties of the passive spectroradiometer to distinguish between aerosols and clouds and the general uncertainty of nighttime measurements, especially under polar night conditions and above the oceans, the cloud fraction of individual grid points can differ significantly. Additionally the spot size of CALIPSO is limited to 70 m, compared to the MODIS swath width of 2330 km. This causes, especially at the equator with a ground track spacing of 172 km after a 16 day repeat cycle, low observation density for CALIPSO.

2007		82 °S-60 °S	60 °S-30 °S	30 °S-0 °	0 °-30 °N	30 °N-60 °N	60 °N-82 °N
JFM	a	38.3	23.5	45.5	58.1	36.5	46.3
	b	32.1	23.4	35.1	46.0	21.8	34.2
AMJ	a	41.7	22.1	52.3	51.9	38.5	36.5
	b	37.0	16.8	42.0	40.1	32.1	31.2
JAS	a	44.8	22.0	51.6	46.0	44.0	28.5
	b	38.3	16.5	45.7	31.8	39.2	24.3
OND	a	43.9	22.5	44.7	54.3	38.9	31.5
	b	34.5	21.0	37.8	40.5	29.7	21.0
Mean	a	42.2	22.5	48.5	52.6	39.5	35.7
	b	35.4	19.2	40.3	39.6	30.7	27.6

Table 7: The cloud free fraction for averaged latitudal zones of the analyzed CALIPSO (a) and Aqua MODIS (b) data set from Figure 16 and 17.

Using only the 5 km cloud layer product would flip this effect and the CALIPSO data set would be more cloudy with a cloud free fraction of only 24.8 % [Wu et al., 2011] compared to the MODIS data set. This significant difference compared to the merged CALIPSO data set is caused by the averaging algorithm of the 5 km product and the absence of cloud gaps being smaller than 5 km. A monthly overview of the cloud free fraction for the CALIPSO merged data set was performed in the Appendix (see Table 13). Although certain variations have to be noticed, the general trend and the major large-scale patterns are congruent.

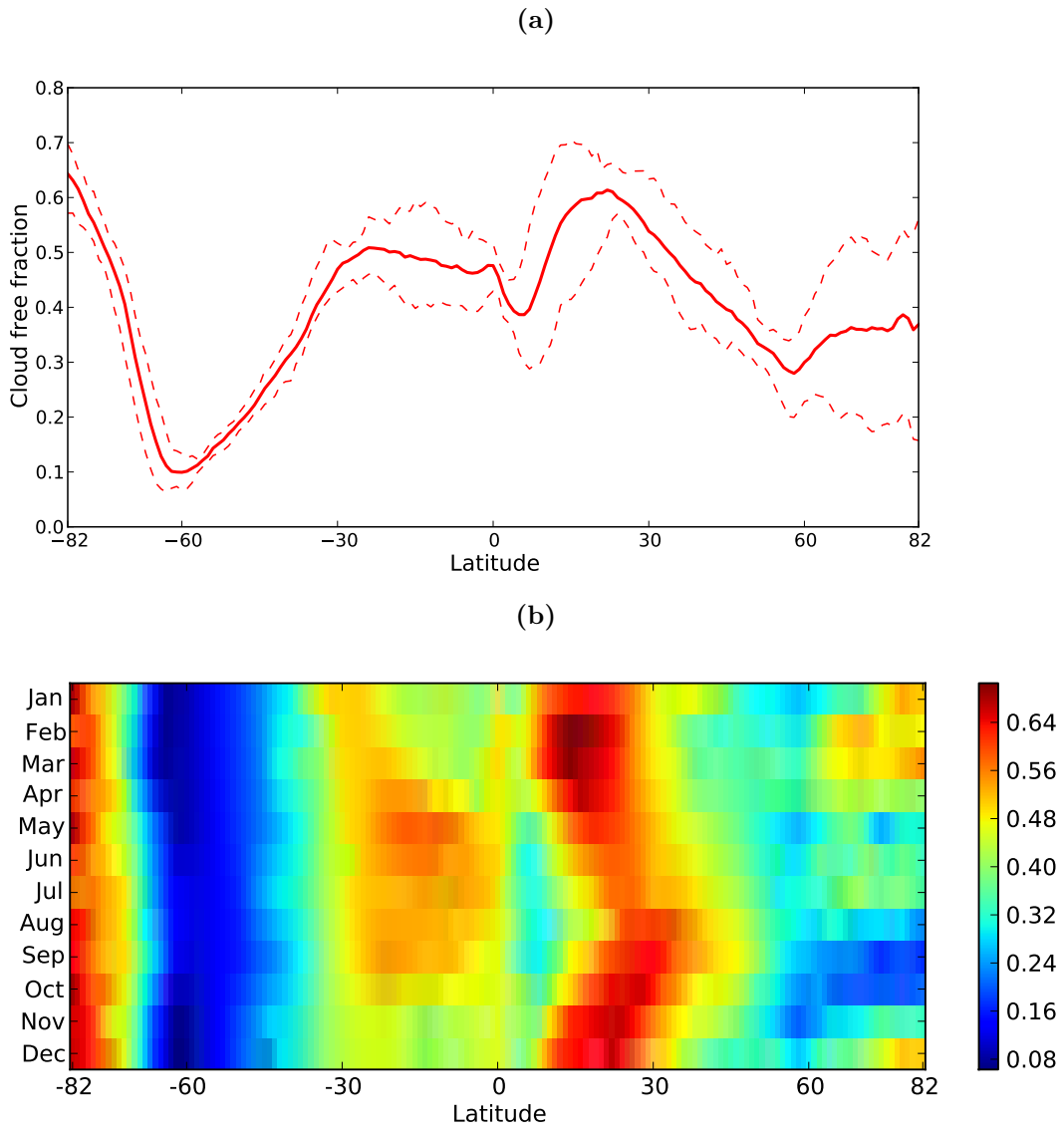


Figure 18: Cloudless fraction summed over each latitude degree from 82 °S-82 °N for the entire data set. For (a) the dashed lines indicate the maximum and minimum values, while the solid line indicates the annual mean for the respective latitude degree. For (b) the colorbar represents the cloudless fraction for each latitude degree for different months from January to December 2007.

The same holds for the analysis of seasonal variations in Figure 17, which presents the same calculations as for Figure 16 but for trimesters ranging from January, February, March (JFM) to October, November, December (OND). Once again, the general trend of cloud free regions remains the same for both data sets, even though, the MODIS data set contains more clouds, especially in the subtropic oceans (see Table 7 for details). In both data sets the Intertropical Convergence Zone (ITCZ) with high cloud cover moves north in the summer and with it the wind circulation that drastically affects the trade winds and thereby the cloud free fraction. As the merged CALIPSO data set is in good qualitative accordance with the MODIS data set, the MODIS data are not used in the further analysis.

To analyze the zonal distribution and the variability of cloud free regions, a histogram (see Figure 18) was computed by rounding all latitude values to its next integer and checking whether or not the corresponding measurement is cloud free (using the data array *Number Layers Found*). Additionally such a histogram was computed for each monthly data set from January to December to identify the temporal relationship of changes. While the Antarctic and the ocean cloud coverage up to 45 °S stays quite stable the entire year, a clear shift of the ITCZ can be seen in the northern subtropics from 10 °N to 20 °N. Here the cloud free band shifts approximately 20 °N with a peak in August. Otherwise the cloud free fraction in the southern subtropics from 0 ° to 30 °S also rises to a peak in the northern summer. The arctic region also shows at least two different extremes with a cloud free peak from December to March and a low from August to October. The previously mentioned cloud band over the oceans between 50 °S and 70 °S remains constant the entire year.

5.2 Global cloud gap frequency

In a next step, the cloud gap length frequency contributing to the cloud free fraction was analyzed. Therefore all individual cloud gaps regardless of their geographical location were counted and illustrated by their probability density function (PDF) on a log-log scale. The evident linear behavior is suspect to be a power law distribution as PDF's of high resolution lidar data like CALIPSO or the airborne NASA LaRC High Spectral Resolution Lidar (HSRL) with combined MODIS cloud data revealed a domain-wide power law fit for global cloud lengths of $\alpha = 2.03 \pm 0.06$ ($\pm 95\%$ confidence interval) [Leahy et al., 2012] and $\alpha = 1.66 \pm 0.04$ respectively [Wood and Field, 2011]. As this power law behavior can also be found for other natural processes [Klaus et al., 2011] and was predicted for cloud gaps by [Wood and Field, 2011], a detailed analysis deals with

the question, if the cloud gap length distribution (see Figure 19a) follows a power law, too.

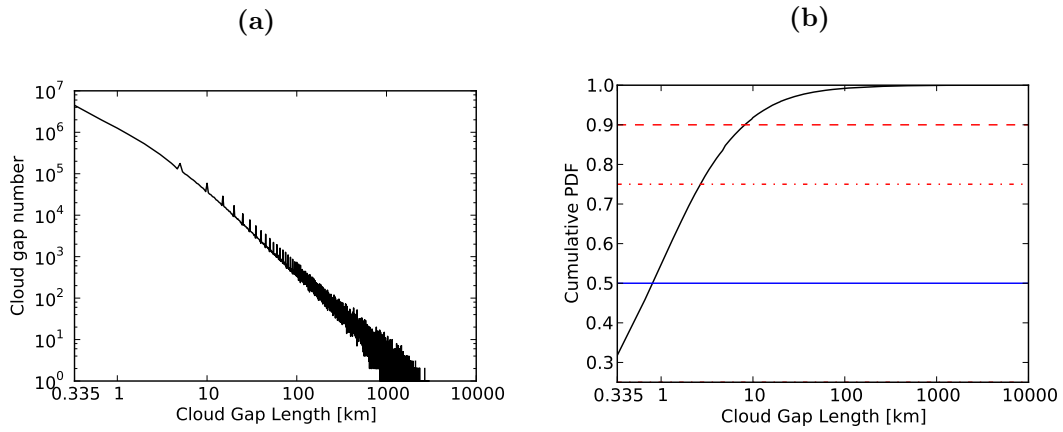


Figure 19: Occurrence frequency of specific cloud gap lengths represented by their respective probability density function for the CALIPSO 2007 data set with 14,092,495 individual cloud gaps and no binning performed (a), where the maximum cloud gap length is 4807 km. (b) is the respective cumulative probability density function, showing that the median lies below 1 km. The mean cloud gap length is 7.41 km.

If the cloud gap length distribution can also be explained by a power law distribution it is possible to derive basic physical rules that can be used for climate models, as well as for comparison with other natural processes following a power law. Due to the mathematical simplicity of such a power law distribution the computational effort for climate model calculations could be reduced. For this purpose methods and equations presented by [Clauset et al., 2009] were used to analyze the PDF.

While commonly used methods for a power law analysis like least-squares fitting can produce substantially inaccurate estimates and also give no indication whether the data obey a power law at all, [Clauset et al., 2009] uses a maximum-likelihood fitting method combined with a goodness-of-fit test based on Kolmogorov-Smirnov statistic and likelihood ratios, to analyze the cloud gap distribution of the data set. Due to the biased cloud gap length distribution at 5 km multiples, caused by the merging process (see spikes in Figure 19a) a logarithmic binning had to be performed prior to the power law analysis (see Appendix for further information of logarithmic binning) to smooth the distribution. The apparent higher probability of 5 km multiples and the break of the linear slope for cloud gaps smaller than 5 km is caused by the fixed size of multiples of 5 km clouds added to the original 333 m data set. Thus, cloud gaps smaller than 5 km that are covered by a cloud from the 5 km product vanish and consequently cause a lower probability of occurrence. The same holds for cloud gaps being larger than 5 km but then being interrupted by 5 km clouds (e.g. a 23 km cloud gap, being interrupted by two 5 km clouds would cause two 5 km cloud gaps and one 3 km cloud gap).

Mathematically, a power law is described by:

$$p(x) \propto x^{-\alpha} \tag{14}$$

where α is the constant parameter known as the power law exponent, and $p(x)$ signifies the probability distribution of a quantity x obeying a power law. For the cloud gap length distribution the quantity x is equivalent to the cloud gap lengths. Therefore all integers x can be treated as multiples of 335 m (e.g., $x_1 = 0.335$ km, $x_2 = 0.670$ km etc.). Although the power law is an idealized description of the distribution of a quantity, in practice the power law applies only for values greater than some minimum x_{min}. To find this tail of the power law, there are several statistical tools at hand. First of all, however, both possible distributions are introduced: the continuous distribution and discrete distributions. While a continuous distribution is described by a probability density $p(x)$ so that:

$$p(x)dx = Pr(x \leq X < x + dx) = Cx^{-\alpha}dx, \tag{15}$$

where X is the observed value and C is the normalization constant. As power laws are undefined for $x = 0$ a lower bound x_{min} has to be found above which the power law distribution holds. With the calculated normalization constant and $\alpha > 1$ the equation can be changed to:

$$p(x) = \frac{\alpha - 1}{x_{min}}(\frac{x}{x_{min}})^{-\alpha}. \tag{16}$$

In the discrete case, where x can only take integer values, the probability density function changes to:

$$p(x) = Pr(X = x) = Cx^{-\alpha} \tag{17}$$

and again after calculating the normalization constant, a lower bound for x_{min} can be found by:

$$p(x) = \frac{x^{-\alpha}}{\sum_{n=0}^{\infty}(n + x_{min})^{-\alpha}}. \tag{18}$$

A discrete power law fit can be approximated by its continuous counterpart by treating integer power law values of x as they were generated from a continuous power law and then rounded to the nearest integer. In this case, the probability mass at x is equal to the sum of the continuous probability between $x_{-0.5}$ through $x_{+0.5}$. This approach does not just give reliable and accurate results, but often is the only possibility to express a power law behavior properly (see Table 8). As the power law follows a straight line in a doubly logarithmic plot, estimating the scaling factor of the power law requires a lower bound x_{min} that can be set or elsewise estimated from the data as well. A maximum likelihood estimator for $\alpha > 1$, such as the Hill estimator, gives accurate parameter

estimates $\hat{\alpha}$ in the limit of large sample sizes n and is expressed by:

$$\hat{\alpha} = 1 + n[\sum_{i=1}^{n} ln\frac{x_i}{x_{min}}]^{-1} \tag{19}$$

for the continuous case, and by:

$$\hat{\alpha} \backsimeq 1 + n[\sum_{i=1}^{n} ln\frac{x_i}{x_{min} - 1/2}]^{-1} \tag{20}$$

for the discrete case, where $x_i, i = 1...n$ are the observed values of x for which $x_i \geq x_{min}$. The standard error of the estimated power law exponent $\hat{\alpha}$, is thereby given by:

$$\sigma = \frac{\hat{\alpha} - 1}{\sqrt{n}} + = O(1/n). \tag{21}$$

Note that all hated parameters are estimates of the respective true parameters. [Clauset et al., 2009, White et al., 2008] performed several analyses and tested the MLE against least-square fitting routines with the finding that the MLE result in the best power law fit in the discrete and continuous case. This is based on the low statistical error of the estimator, which decays as $O(n^{1/2})$. Clauset also found out that the initial x_{min} value is crucial and in the case of an underestimation of $x < x_{min}$ $\hat{\alpha}$ deviates quickly from its true value which can also be seen in Table 8. As the cloud gap data used for this study is a discrete version of the probability distribution it cannot be accurately fitted with its continuous counterpart. Therefore the MLE fit for a discrete power law is found by numerical optimization in which the continuous distribution is summed up to the nearest integer. To ensure the best power law fit possible, an alternative to the maximum likelihood method is the Kolmogorov-Smirnov statistic. It finds the x_{min} that best fits the starting point of the power law, by minimizing the distance between the data and the fit.

Kolmogorov-Smirnov distance:

The Kolmogorov-Smirnov statistic simply calculates the distance between the cumulative distribution function of the estimated data with $\hat{x}_{min}$ and compares it with the cumulative distribution function of the fitted model.

$$D = \max_{x \geq x_{min}} |M(x) - P(x)| \tag{22}$$

where D is the distance, $M(x)$ is the CDF of the observation data with at least one value x_{min} and $P(x)$ is the respective CDF of the power law best fitting the data in the region $x \geq x_{min}$. The optimal $\hat{x}_{min}$ minimizes D.

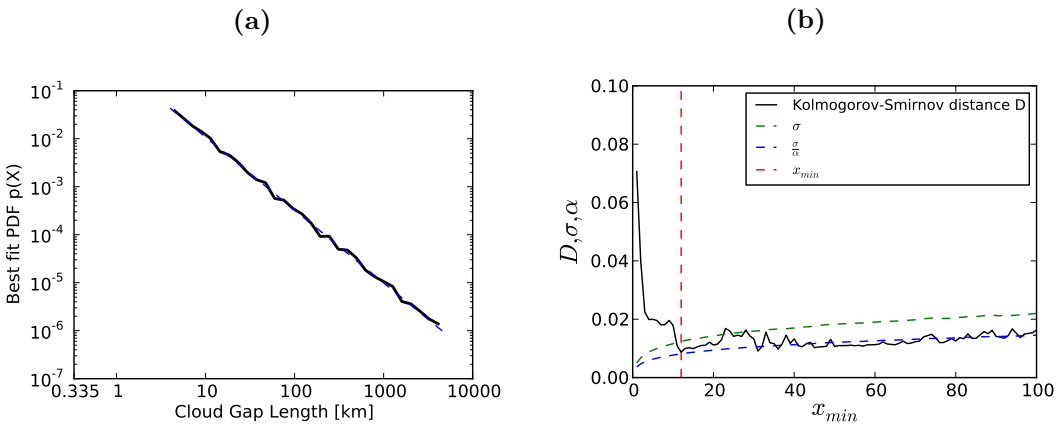


Figure 20: The best fit power law result for the logarithmically binned and integral normed PDF from $x_{min} = 4$ km to $x_{max} = 4807$ km (a), where the power law exponent is $\alpha = 1.518$ and $\sigma = 0.012$. The Kolmogorov-Smirnov distance and the significant error σ in dependence of the corresponding x_{min} are shown in (b).

As it can be seen in Figure 20, the Kolmogorov-Smirnov distance is first minimized for $x = 12$ which corresponds to a cloud gap length of 4 km. Consequently the discrete power law fit starts at 4 km ranging to the maximum cloud gap length of 4807 km with $\alpha = 1.518 \pm 0.012$. The KS method is better than other estimation methods [Clauset et al., 2009] and under the assumption that many observations are available $\hat{\alpha} \to \alpha$ for $n \to \infty$.

Cumulative distribution function (CDF):

The cumulative distribution function (the integral of the probability distribution function for real-values) of a power law distributed variable, is a useful quantity and allows a conclusions about the median and percentiles of the distribution. Generally, the form of a CDF is more robust against the finite sample size fluctuation, at the end of the power law tails [Clauset et al., 2009]. For the continuous case the function $P(x) = Pr(X \geq x)$ equals:

$$P(x) = \int_x^\infty p(x')dx' = (\frac{x}{x_{min}})^{-\alpha+1} \tag{23}$$

and for the discrete case:

$$P(x) = \frac{\sum_{n=0}^{\infty}(n+x)^{-\alpha}}{\sum_{n=0}^{\infty}(n+x_{min})^{-\alpha}} \tag{24}$$

holds.

Taking all cloud gap lengths and their global occurrence frequency (see Figure 19a) into account, different fitting methods (MLE and KS distance) as well as continuous and discrete fits have been performed to obtain the best power law fit possible. To have a smooth distribution the individual cloud gap lengths were logarithmically binned. As already mentioned, the CALIPSO merged data set is a discrete distribution with values

for fixed cloud gap lengths and therefore the fit in between of those cloud gap lengths has to be explained by a continuous distribution. However the continuous distribution is an idealized representation and should be regarded as a model. Consequently the terms discrete and continuous are from now on renamed to CALIPSO data fit and Model data fit. The bold power law estimates in Table 8 are found by estimating x_{min} from the data by the methods described above.

	fitting method	CALIPSO/Model	x_{min}	α	σ	D
$\hat{\alpha_1}$	**KS**	**CALIPSO**	**13**	**1.5287**	**0.0128**	**0.0076**
$\hat{\alpha_2}$	KS	CALIPSO	1	1.3401	0.0123	0.0512
$\hat{\alpha_3}$	**KS**	**Model**	**1**	**1.4378**	**0.0058**	**0.0253**
$\hat{\alpha_4}$	KS	Model	13	1.6482	0.0157	0.3053
$\hat{\alpha_5}$	**likelihood**	**CALIPSO**	**12**	**1.5183**	**0.0123**	**0.0087**
$\hat{\alpha_6}$	likelihood	CALIPSO	1	1.3721	0.0050	0.0706
$\hat{\alpha_7}$	**likelihood**	**Model**	**1**	**1.5169**	**0.0069**	**0.0734**
$\hat{\alpha_8}$	likelihood	Model	12	1.5861	0.0139	0.2941

Table 8: The estimated power law exponents for two fitting methods and the CALIPSO data and Model data with its respective power law exponent α, the significant error σ and its Kolmogorov-Smirnov distance D. The results indicated in bold were found by estimating x_{min} from the data, whereas for the other results x_{min} was set manually.

Consequently the other estimates are computed for manually set x_{min} values. As a result, those fits are quite poor which is reflected by their larger significant error or Kolmogorov-Smirnov distance that rapidly deviates for $x > x_{min}$. With the exception of one outlier ($\hat{\alpha_3}$), the power law exponent is found to be approximately 1.51 ± 0.01 regardless of the fitting method or the kind of distribution. While the cloud gap lengths distribution is discrete, $\hat{\alpha_1}$ and $\hat{\alpha_5}$ are the best possible fits according to the MLE and KS estimators.

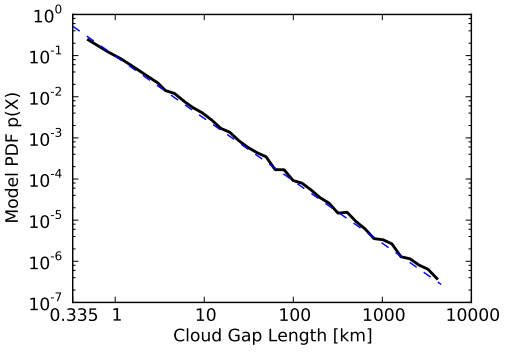


Figure 21: The logarithmically binned integral normed model PDF and its fitted power law with the exponent $\hat{\alpha_7} = 1.517$ and $\sigma = 0.007$, starting from the smallest cloud gap possible.

However, as $\hat{\alpha_7}$ lies within the significant error of both CALIPSO data fits, it represents a good approximation of a cloud gap length model starting from the smallest length of 335 m. Consequently the PDF (see Figure 21a and b) of the model power law fit agree well with the respective logarithmically binned CALIPSO data. However it is worth mentioning that for cloud gap lengths being smaller than 5 km, the model fit is only achieved due to the logarithmic binning which smoothed the distribution.

Power law fitting validation:

Unfortunately, the presented fitting routines for a power law do not give any qualitative measure of the correctness of the distribution itself. Therefore a goodness-of-fit test [Clauset et al., 2009] has to be performed and other distributions must be compared against the best fit power law. Again, the KS statistic serve as an indicator which distribution represents the data best. For this purpose the cloud gap data is checked against two other distributions which tend to be satisfactory alternatives. This check is performed by a likelihood ratio test with two probability density functions $p_1(x)$ and $p_2(x)$ and their corresponding log-likelihood ratio:

$$R = \sum_{i=1}^{n} [lnp_1(x_i) - lnp_2(x_i)] = \sum_{i=1}^{n} [l_i^1 - l_i^2] \tag{25}$$

where $l_i^j = lnp_j(x_i)$ is the log-likelihood ratio for a single value x_i within the distribution j. A fundamental assumption for the log-likelihood ratio is, that all individual data points are independent, [Alstott et al., 2013] and by the central limit theorem R therefore becomes normally distributed [Rice, 1995] as $n \to \infty$ with its approximated variance σ^2:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} [(l_i^{(1)} - l_i^{(2)}) - (\bar{l}_i^{(1)} - \bar{l}_i^{(2)})]^2 \tag{26}$$

with:

$$\begin{aligned} \bar{l}^{(1)} &= \frac{1}{n} \sum_{i=1}^{n} l_i^{(1)}, \\ \bar{l}^{(2)} &= \frac{1}{n} \sum_{i=1}^{n} l_i^{(2)}. \end{aligned} \tag{27}$$

The probability of the log-likelihood ratio, and therefore an indicator for the ratio is given by:

$$p = \frac{1}{\sqrt{2\pi n\sigma^2}} [\int_{-\infty}^{-\|R\|} e^{-t^2/2n\sigma^2} dt + \int_{\|R\|}^{\infty} e^{-t^2/2n\sigma^2} dt]. \tag{28}$$

While the ratio R is either positive, and thus tending to the first distribution, or negative and consequently tending to the second distribution, the probability p is an indicator

for the log-likelihood tests itself. For small p (i.e., $p \leq 0.1$) it is very likely that the ratio is an indicator for the better fit and contrary for large p, the ratio is not unambiguous. As already mentioned and demonstrated, the best possible power law fits are $\hat{\alpha_1}$, $\hat{\alpha_3}$, $\hat{\alpha_5}$ and $\hat{\alpha_7}$. To check the correctness of the assumption of a power law fit, the estimated distributions are compared against the log-normal and truncated power law distribution with a log-likelihood ratio test (see Table 9).

	power law vs log-normal		power law vs trunc. power law		log-normal vs trunc. power law	
	$\|R\|$	p	$\|R\|$	p	$\|R\|$	p
α_1	**1.1909**	**0.2336**	**0.6392**	**0.3548**	-1.2078	0.2270
α_3	-0.1459	0.8839	2.8826	0	3.6977	0.0002
α_5	**0.4121**	**0.6802**	**0.1232**	**0.9602**	-0.4048	0.6856
α_7	-1.1804	0.2378	-8.4586	$4.44*10^{-16}$	-10.9631	$5.74*10^{-28}$

Table 9: The results of the log-likelihood ratio test for selected power law cases and three possible distributions with its log-likelihood ratio R and the respective probability p.

While the model distributions $\hat{\alpha_3}$ and $\hat{\alpha_7}$ are failing the test and are revealing the likelihood of a different distribution, both CALIPSO data fits $\hat{\alpha_1}$ and $\hat{\alpha_5}$ are in favor of a power law distribution. Therefore the log-likelihood ratio test finally confirms the correctness of a power law distribution with $\alpha = 1.51$ and $\sigma = 0.01$.

5.3 Global cloud gap distribution

Taking into account the various geographical variations in the cloud free fraction led to the question whether $\alpha = 1.51 \pm 0.01$ is valid for each point on the Earth. For this purpose the discrete power law fit with an threshold of $\sigma \leq 0.1$ was computed for the cloud gap lengths distribution that was assigned to each 2 ° x 2 ° grid point (see Figure 22) by the algorithm described in Chapter 3. The global distribution of individual cloud gaps is not equally distributed (see Figure 23) as most cloud gaps occur over the oceans from 45 °S to 45 °N, despite the fact, that there are more measurements in polar regions (see Figure 9). The high concentration of individual cloud gaps combined with the above average cloud free fraction present over subtropic oceans increases the likelihood of larger cloud gaps dominating the grid point which is confirmed by the low power law exponent seen in Figure 22. However, with 82.2 % of all grid points having at least 500 individual cloud gaps (see Figure 23), a power law fit can still be performed.

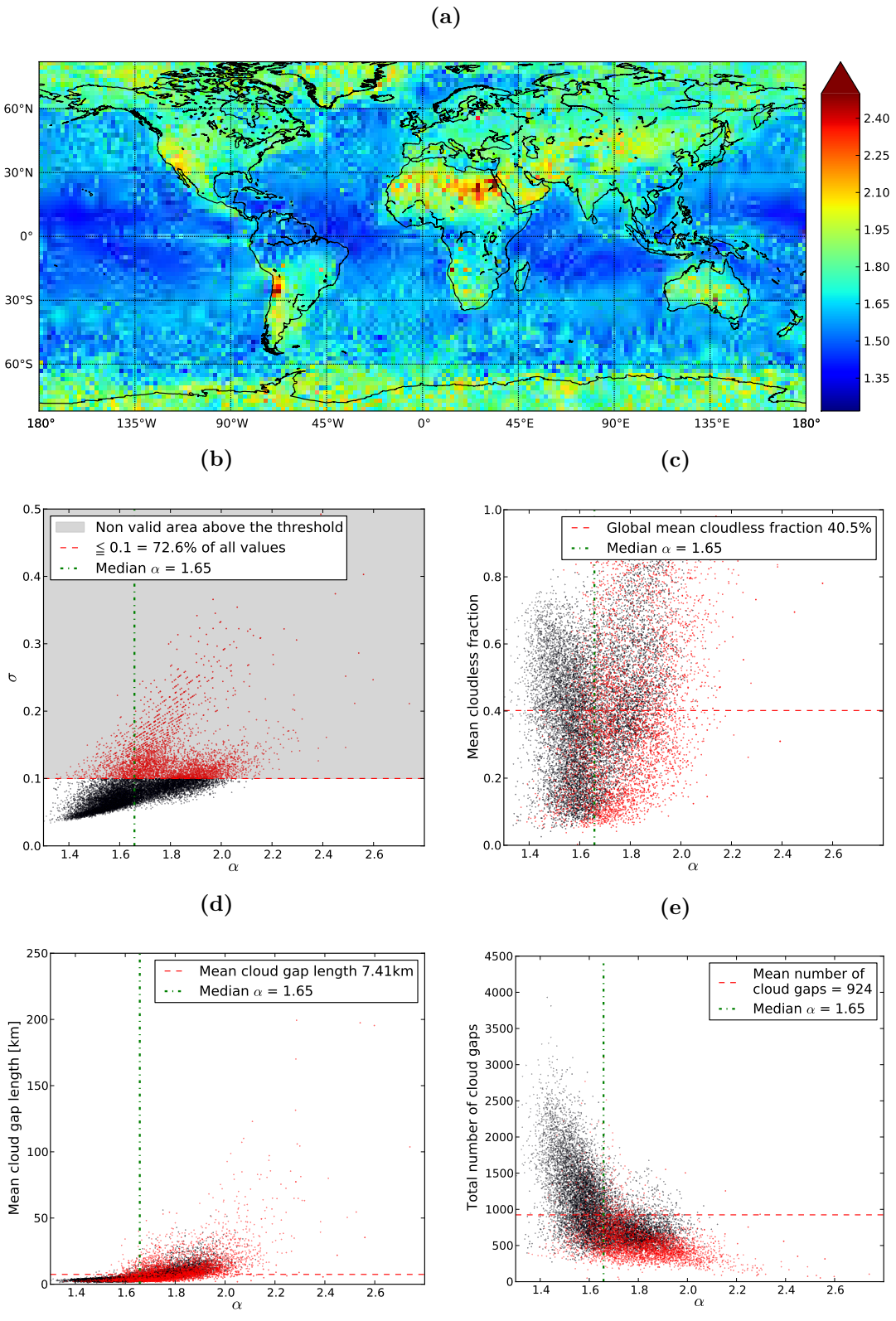


Figure 22: (a) is the best fit power law exponent α computed for each 2° x 2° grid point (complete 2007 data set) and its significant error σ (b), where all power law exponents that exceed a threshold of $\sigma = 0.1$ are marked as red. Non valid power law fits can mainly be seen in regions with high cloud cover, while the relationship of the cloud free fraction and the power law exponent α is spread over the whole domain as indicated in (c). Only for regions where the mean cloud gap length is above average the power law exponents exceeds $\alpha > 2$ as indicated in (d). The significant error most likely exceeds the threshold (deviation from power law) for grid points where the total measurement points for the fit lay below average, as seen in (e).

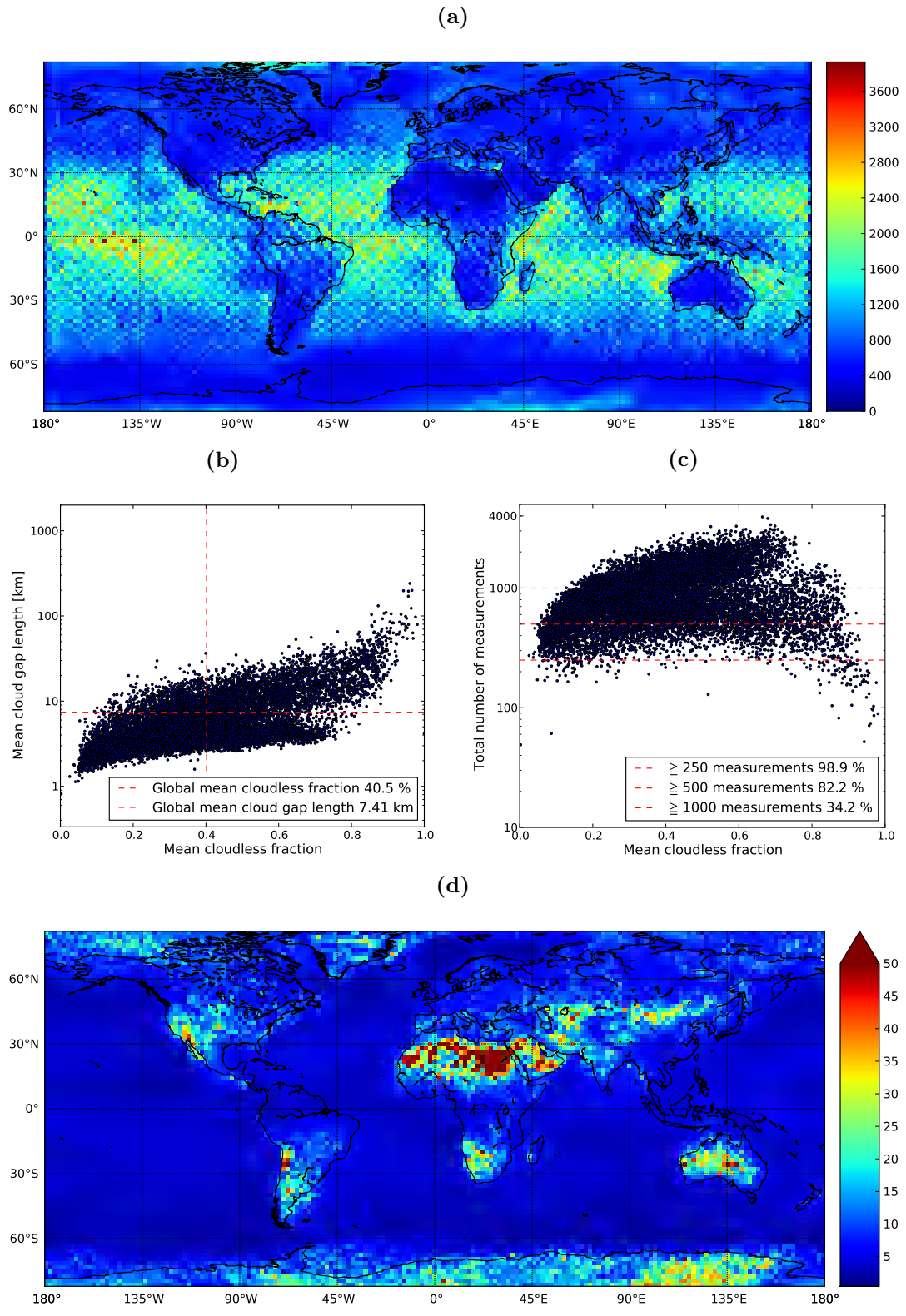


Figure 23: Total number (a) and their respective mean length (b) of cloud gaps for the complete data set of 2007 on a 2 ° x 2 ° grid. As shown in (a) the amount of pure cloud gaps dominates over the tropic and subtropic ocean regions, while the mean cloud gap length in km (d) has its maximum above the global deserts and high mountain ranges. To support this statement, (b), on the one hand shows that for a high cloudless fraction, the mean cloud gap length rises, but on the other hand very few measurements are available for regions with a high cloudless fraction (c).

The anomalies indicated by high power law exponents above the Sahara, at the west coast of North and South America around the latitudes of 30 °S and 30 °N, at the Arabian Peninsula and above the Chinese High Plateau are characterized by a high cloud free fraction and large cloud gaps (see Figure 22c and d). In general all power law fits exceeding $\sigma \geq 0.1$ (27.4 % of all 15023 grid points) have a low amount of individual cloud gaps and deviate from a power law fit (see Appendix for the corresponding σ and x_{min} for each grid point). Consequently the closest approach to the global power law fit is reached only for grid points having more than 1000 individual cloud gaps (34.2 % of all grid points), with a mean cloud gap length not exceeding 15 km and preferably a high cloud free fraction. Regardless of the threshold of $\sigma \geq 0.1$, the median power law exponent for all grid points is 1.65. This is based on the irregular global distribution of individual cloud gaps which causes a strong deviation from a power law fit.

With a global mean cloud free fraction of 40.5 % and a global mean cloud gap length of only 7.41 km, the dominance of small cloud gaps becomes evident. Only for the probable cloud free regions detected in Section 5.1 and especially for the Sahara, the mean cloud gap length increases up to 230 km. The median cloud gap length (see Figure 24) is related to the mean cloud free fraction. The median is the numerical value separating the higher half of the data from the lower half, and with only a few grid points exceeding 5 km the global dominance of small cloud gaps is confirmed. Interestingly, three latitude bands can be found above the oceans with rising medians to the equator. The first band from ±82 ° to ±50 ° with a median of 666 m, is replaced by a band from ±50 ° to ±20 ° and a median of 1 km. The equatorial band from ±20 ° to 0 ° is dominated by medians up to 2 km.

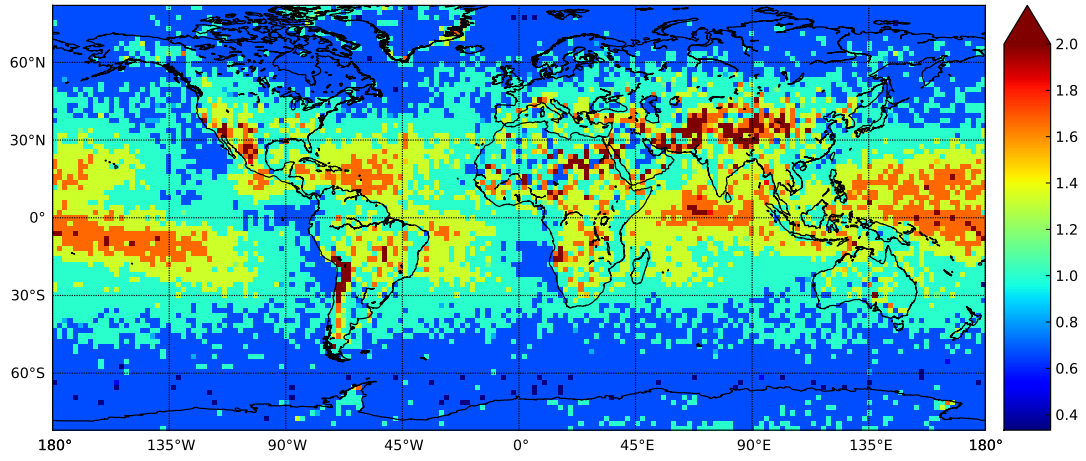


Figure 24: Median of cloud gap lengths for the entire CALIPSO data set of 2007 computed for 2 ° x 2 ° grid points.

5.4 Conclusion

With a global cloud gap length and distribution statistic at hand, a first estimate on the performance of a future space-based lidar system can be performed. While approximately only 2 % of all cloud gaps exceed 50 km length (see Figure 19) a perfect global retrieval of methane from MERLIN is highly unlikely. However, as the cloud gap length statistics reveal a dominance of small cloud gaps, the assumption is made, that at least in probable cloud free regions the consecutive cloud free retrieval is only interrupted by small broken clouds. This assumption is supported by the mean cloud gap length being higher for cloud free regions and the total amount of cloud gaps being lower especially over continents. As [Kiemle et al., 2011] stated, a methane retrieval is also possible for cloudy regions even though the integration length has to be increased. The next chapter will deal with this.

Chapter 6

Cloud flatness occurrence

6.1 Search criteria and algorithm

As [Kiemle et al., 2011] stated, a methane retrieval might also be possible for flat low altitude cloud tops as they describe a clear boundary for an automatized integration routine. Therefore a search method was implemented that searches the lower troposphere for flat cloud tops with at least 50 km horizontal extent and a maximum of 30 m vertical variation. The 50 km horizontal extent is based on the desired nominal resolution of MERLIN [Kiemle et al., 2011], while the 30 m vertical variation is the minimum possible step size of the underlying 333 m product with 30 m vertical resolution. For this analysis only the CALIPSO single shot data ranging up to 8.2 km was used. To find consecutive cloud top heights with equal altitude, a search window of 200 km was introduced to step through the whole data set. This search window is necessary because the search algorithm is based on a histogram approach which uses a threshold to identify the peak height within the 200 km. One drawback of such a fixed search window irrespective of the geographic position is that individual clouds could be arbitrarily cut. Thus, if a 50 km cloud with a flat cloud top is in between of two sequenced search windows, the threshold will not be met for neither of the search window histograms. To circumvent this problem, an overlap of 100 km (half-overlapping search window of 200 km size) was introduced prior and past of the actual search window (see Figure 25). The detection threshold d_{thres} of the histogram is given by the fraction:

$$d_{thres} = \frac{c_{len}}{w_{len}} \tag{29}$$

with c_{len} being the cloud length with uniform top height and w_{len} being the size of the search window.

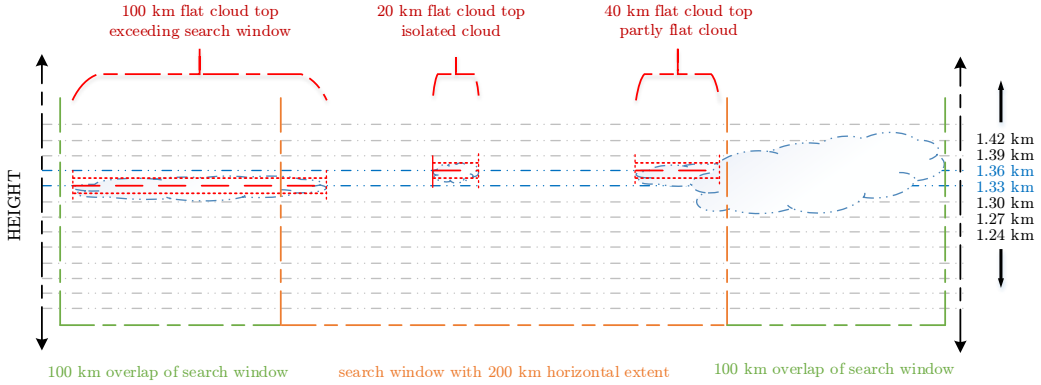


Figure 25: The cloud flatness search algorithm uses a histogram approach with half-overlapping 200 km search windows to detect flat cloud tops.

Thus, the detection threshold for a 50 km cloud with a flat cloud top in a 200 km search window is 0.25. As a next step the algorithm computes a cloud top height histogram with 30 m vertical binning of the clouds within the search window. If a cloud height bin exceeds the detection threshold, a cloud flatness event will be registered within the search window. In case there are multiple bins that exceed the threshold, an additional event is assigned. For the example seen in Figure 25, two events would have been detected because the height bin of 1.33 km and 1.36 km would have exceeded the threshold of 0.25. By using the histogram approach, not only consecutive clouds are found, but also broken cloud scenes are taken into account (see Chapter 4 for examples). This is possible because cloud free columns are treated apart. The major drawback of the histogram approach is the inability of the algorithm to distinguish whether the histogram bins belong to one or more clouds. As a consequence, cloud flatness could also be associated to one large cloud with a slightly rough cloud top instead of multiple individual clouds. To exclude the detection of neighboring height bins, an additional vertical layer threshold can be used. However, such a threshold would also exclude multiple individual flat cloud tops with similar heights such as in Figure 25. Therefore two analyses have been performed, one with no additional vertical threshold and one with a minimum vertical spacing of 120 m, which accounts for slightly rough cloud tops of large clouds. To assess the global distribution of such cloud flatness events, the rounded latitude and longitude coordinates of the search window midpoint as well as all event top heights are stored. All individual measurements that contribute to the cloud flatness event are not retained and a determination of different cloud layers and their exact position is not forseen. However, for a basic estimation of the global flatness distribution in location and height, this approach is sufficient.

6.2 Global cloud flatness occurrence

Once again, the algorithm was applied to all monthly data sets and in total 76163 events (about 200 a day) were detected in 2007 with cloud tops being flat (± 15 m) at least over 50 km. Therefore the cloud flatness events account for approximately 1.3 % of all clouds being detected in the lower troposphere up to 8.2 km for the entire year. As already mentioned the minimum vertical threshold for the layer separation was either set to zero to account for any flatness detected in the data set, or it was set to a minimum of 120 m to account for two or more vertically separated cloud layers. Consequently, the additional layers for two and more layers without any vertical layer threshold sum up to 17140, whereas only 224 additional layers were detected when setting the threshold to 120 m. As already mentioned in Section 4.3 multilayered cloud flatness events are very rare.

2007	1 layer	2 layers	3 layers
Jan	5818	1155/9	77/0
Feb	5450	1008/15	66/0
Mar	5534	1079/10	69/0
Apr	4993	1001/10	69/0
May	7012	1524/21	104/0
Jun	5924	1304/14	79/0
Jul	5831	1339/20	86/0
Aug	6488	1372/19	92/0
Sep	7011	1475/18	103/0
Oct	7312	1549/24	100/0
Nov	7579	1633/39	131/0
Dec	7201	1614/25	111/0
Total	76163	16053/224	1087

Table 10: Monthly total number of single- and multilayered cloud flatness events with a minimum size of 50km within a 200km search window, a maximum allowed vertical variation of ±15 m with either no minimum vertical threshold between two layers or 120 m vertical threshold. In total 76163 single layered cloud flatness events were registered.

Looking at Table 10 and the total number of cloud flatness events reveals a slightly increasing trend to the end of the year. However, the analysis time span is too short and the trend cannot be confirmed by further samples. Nevertheless, the total number of cloud flatness events proves the assumption of [Kiemle et al., 2011] that flat cloud tops should be considered in a future methane retrieval algorithm. In a next step the geographical location and the height of such events is assessed (see Figure 26). For this purpose all 76163 single layered cloud flatness events were analyzed but only the highest cloud flatness event for each 1 ° x 1 ° grid point was plotted. This was done because of the dominance of such events up to an altitude of 2.5 km with a peak at around 1.4 km as indicated in the top height histogram in Figure 26b. It is obvious that cloud flatness

events are located mainly above oceans and especially in front of the western coast of all continents (e.g. West Coast of North America, South America, Europe, Africa and Australia). The general maritime dominance is related to the typically smooth shape of the maritime boundary layer.

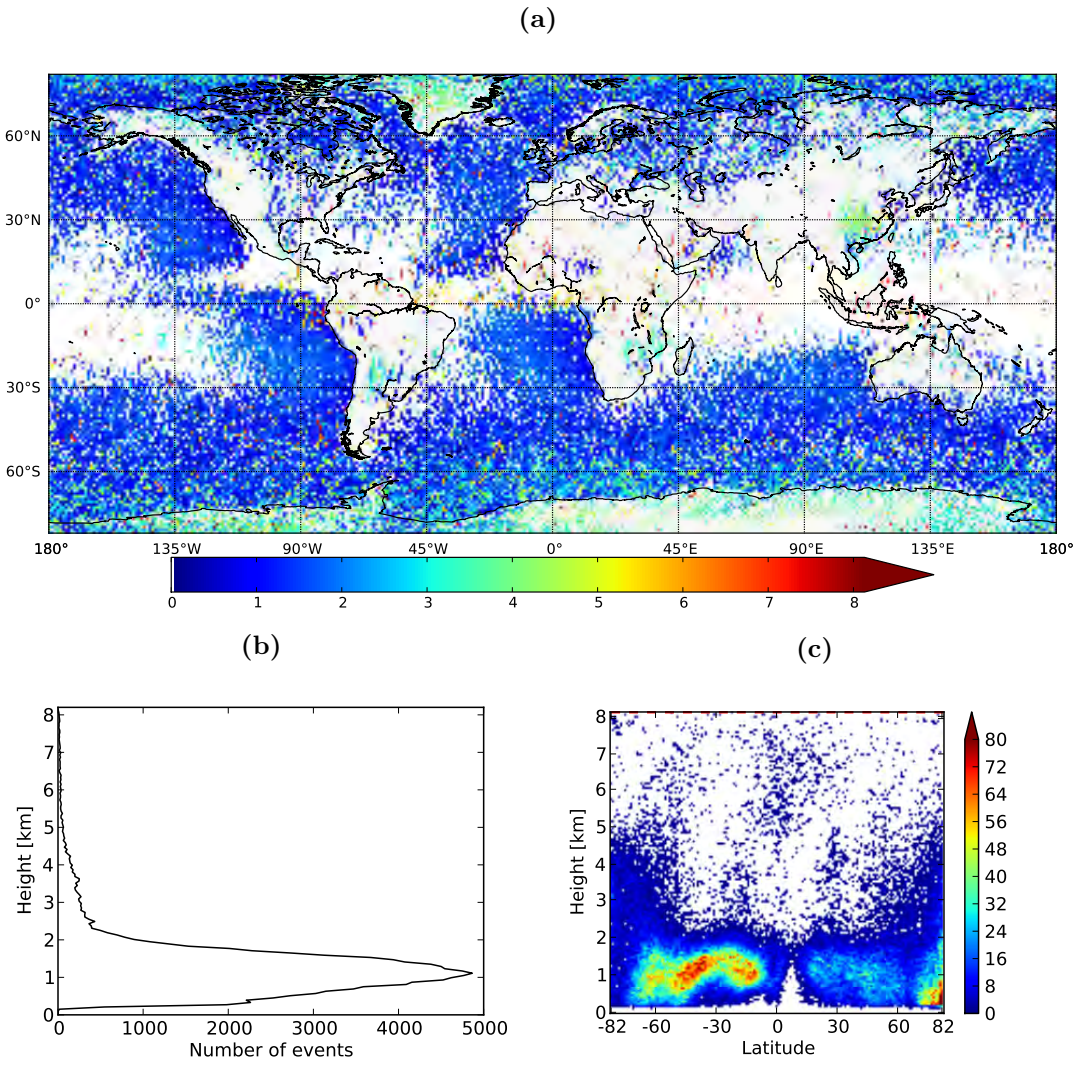


Figure 26: Global cloud top flatness events computed for 2007 and a partly overlapping search window of 200 km and a 50 km consecutive or partly broken cloud within the search window, while the vertical variation does not exceed ± 15 m. The global distribution of flat cloud tops and their respective maximum height is shown on a 1 ° x 1 ° grid (a). The height of such events is color-coded. (b) represents the top height histogram of all events, while (c) shows the latitude averaged cloud top flatness height distribution in number of events. Once again, the oversampling at high latitudes is evident.

Generally, the top height of cloud flatness occurrence is related to cloud cover which is discussed further in Chapter 7. Not surprisingly the accumulation of events in the northern hemisphere between 50 °N and 70 °N, as well as in the southern hemisphere between 50 °S and 70 °S is based on the overall cloudiness in these latitude bands as

discussed in Section 5.1. The lack of events in the equatorial latitude band from 20 °N to 20 °S, especially over the ocean areas, arise not only from the general cloud free fraction, but also from the general circulation of rising air masses due to the Hadley cell circulation. Consequently several cloud flatness events located at the equator occur at altitudes above 6 km. The few events in the Gulf of Mexico, the Caribbean Sea, the Mediterranean Sea, the Indian Ocean and the equatorial part of the Pacific Ocean imply a highly inhomogeneous cloud height distribution. Opposite to the elevated events at the equator the accumulation of events at around 4 km above Antarctica and Greenland result from the topographic elevation of approximately 2 km of the underlying land masses. Whereas, the accumulation of events at around 4 km altitude above East China might be originated in air masses being trapped and elevated due to the Chinese High Plateau in the west. To derive seasonal variations (see Figure 27) the differences of all four trimesters results were computed and analyzed (see Appendix A for detailed results). Interestingly, seasonal differences only occur for events located above land, while maritime events remain unaffected by seasonal effects.

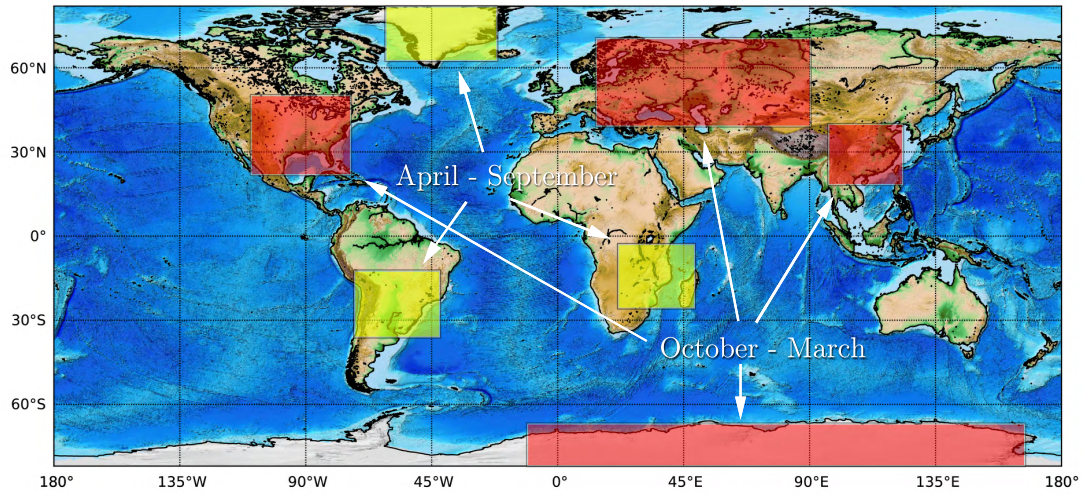


Figure 27: A global relief world map with data from NOAA (`http://www.ngdc.noaa.gov/mgg/global/`) and indicated seasonal cloud flatness events. Yellow boxes indicate events occuring mainly between April to September, and red boxes indicate events occuring mainly between October to March.

6.3 Low maritime clouds

As already mentioned, the fundamental cloud forming process over ocean regions differs from cloud forming processes over continents. In general four cloud forming processes can be distinguished. While the main process of cloud formation is always originated in saturated air masses that are cooled to their dew or frost point this process can be

achieved by different mechanisms. The cloud formation e.g. along the west coast of North America and South America is based on the orographic uplift due to the elevated mountain ranges which cause the rising of the air parcel and a cooling due to adiabatic expansion at a rate of approximately 10 ° Celsius per kilometer until saturation. Another process, responsible for the formation of cumulus clouds and cumulonimbus clouds near the equator or over warm land surfaces, that cause air masses to rise, expand and finally cool and saturate, is called convectional uplift. When sufficient cooling has taken place, saturation causes the formation of clouds. This process is causing most of the cloud formation located in the interior of continents. Additionally radiative cooling that occurs at night or under exclusion of solar radiance while emitting energy in the form of longwave radiation causes the formation of surface fog or near surface clouds. This process is powered by the loss of energy which consequently forces the overlying air to cool and finally saturate. Finally the last cloud forming process takes place when two masses of air with different temperature and moisture characteristics come together, which is called convergence or frontal lifting. One of the air masses is usually cold and dry while the other is warm and moisty. Thereby the cold air mass acts as a ramp causing the warm moisty air to be lifted and cooled due to expansion which again results in saturation. This process is the major mechanism in the mid-latitudes and near the equator where the trade winds meet the ITCZ [Wallace and Hobbs, 2006].

The dominating process over large areas of the oceans that mainly cause low altitude stratocumulus clouds is based on a strong temperature inversion at a height of 0.5 - 1.5 km, which marks the top of the marine boundary layer [Wu et al., 2008] (see Section 4.4 for a typical low maritime flat cloud scene). While the base of such clouds is warmed by longwave radiation from the ocean surface, the tops are cooled by longwave radiation into space. Because of the homogeneity of the ocean surface, there are no strong horizontal temperature gradients above 1.5 km. Consequently this causes a homogeneous MBL and therefore also homogeneous cloud tops. While the total amount of vapor converted to droplets is generally small and would only cause fair-weather cumulus and short lasting stratocumulus, the stronger vertical motion due to convection in the tropics forces the collision and formation of numerous larger droplets up to the upper troposphere. There, the colliding cloud droplets freeze which causes a mixing of ice and liquid droplets which favor a further accumulation of colliding droplets (see Figure 28). The initial cloud forming process is additionally boosted by particles such as dust, sea salt, bits of organic matter or chemical aerosol particles like dimethylsulfat (DMS) or sulfur dioxid (SO_2) which serve as a nucleus for condensating water droplets. With an exception in the tropics, the marine boundary layer limits the occurrence of marine clouds above 2.5 km, which also explains the dominance of marine cloud flatness events with a peak around 1.38 km mainly located over the oceans. The rough location of

those events, agrees well with the major optically thick low maritime clouds analyzed by [Leahy et al., 2012]. Contrary, optically thin low maritime clouds (with an optical depth $\tau \leq 3.0$) were found in the equatorial latitude band from 20 °N to 20 °S at the locations where Figure 26a denotes no cloud flatness returns. Taking into account the underlying 333 m data set used for the cloud flatness analysis, and the fact that optically thin clouds ($\tau \leq 0.45$ [Karlsson and Johansson, 2013] cannot be initially detected without the averaging process of the 5 km product, confirms the correctness of this analysis.

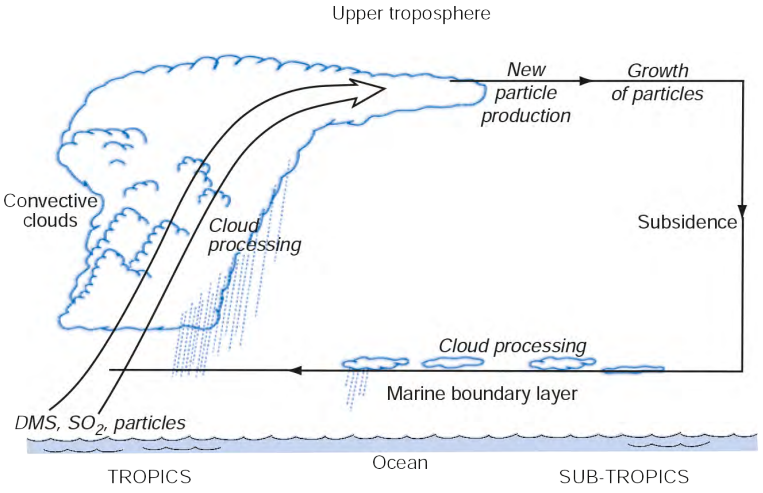


Figure 28: The Hadley cell circulation as a major transport mechanism of aerosol exchange between the tropics and the subtropics and the indicated marine boundary layer responsible for low maritime clouds [Adapted from Wallace and Hobbs, 2006, p. 263].

6.4 Conclusion

While maritime cloud flatness events located in the open ocean are not that important for a future methane lidar, CH_4 sources stored in seabeds along continental shelves could benefit from a retrieval process taking also into account the possibility of low flat cloud tops. Despite a strong accumulation of dense maritime clouds in front of the continents at altitudes ranging from 300 m to 1.5 km most of the methane sources can still be measured because the MBL and therefore also the methane content ranges up to 2.5 km. Various large, seasonal dependent cloud flatness regions above the continents can also be used to retrieve methane column mixing ratios, as they occur at altitudes ranging from 500 m to about 4 km, which corresponds to 300 m to 2.5 km above ground. In summary, the dominance of maritime cloud flatness events is based on the characteristics of the maritime boundary layer and the cloud forming processes involved, and most of the events occurring above the continents are the consequence of various seasonally changing processes.

Chapter 7

Cloud top height distribution

7.1 Global cloud top height distribution

To set the results of Chapter 6 into relation to the underlying merged data set a small analysis on the global cloud top height statistics for 2007 was performed. With many studies focusing on cloud properties, their global distribution, cloud types etc., a lot of effort was put into the research of cloud characteristics by using the space lidar technology on-board of CALIPSO. In this chapter the data set is checked for its consistency with recent cloud analyses such as [Berthier et al., 2008, Pitts et al., 2009, Schwartz and Mace, 2010, Veglio and Maestri, 2011, Wu et al., 2011]. For this purpose a global cloud top height histogram was computed with a homogeneous vertical bin size of 120 m (see Figure 29). Although the original vertical resolution below 8.2 km of 30 m is much better than 120 m, this step was necessary to harmonize the 333 m product with the 5 km product. Consequently all cloud top heights contributing to the histogram are either directly from the 333 m product or due to merging with the 5 km product, whereby the merging criteria includes clouds above 8.22 km and $\tau \geq 0.5$. In total 76.9 % of all cloud top heights were detected up to 8.22 km, while 23.1 % have been detected in the 5 km product. The strong dominance of optically dense low level clouds up to 3.48 km, which amounts to 50.3 % (roughly median) and the very pronounced peak at 1.32 km corresponds well with the results of the space-lidar comparison of [Berthier et al., 2008]. In addition, the second peak at around 10 km was also detected in the study and taking into account the results of [Veglio and Maestri, 2011] this peak is originated in the accumulation of cirrus clouds in the mid latitudes. The negative spike at 8.22 km is owed to the transition zone between the two data sets where clouds theoretically detected at the 333 m product but exceeding 8.2 km top altitude are removed.

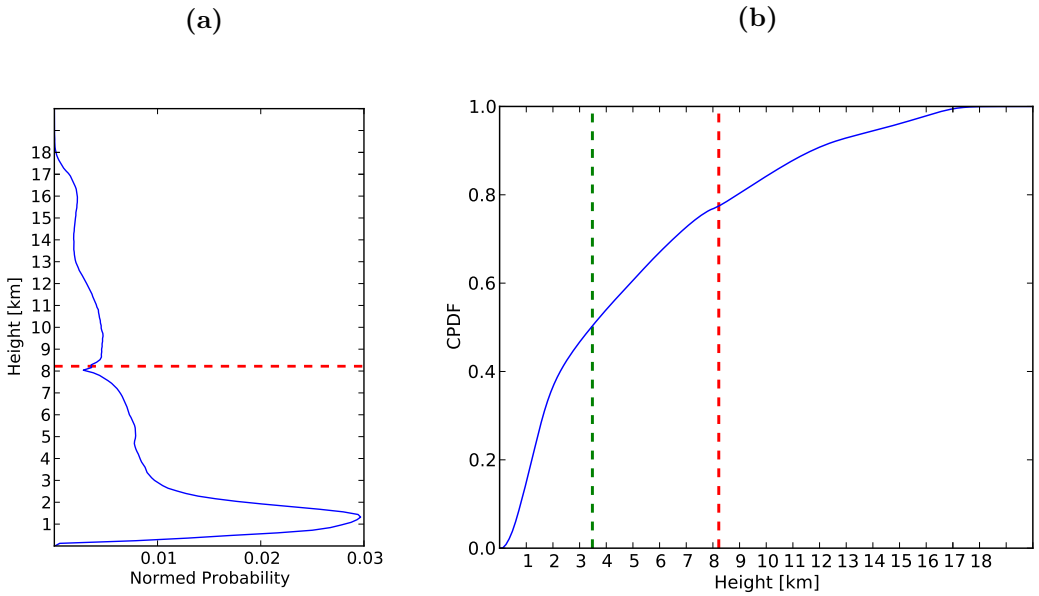
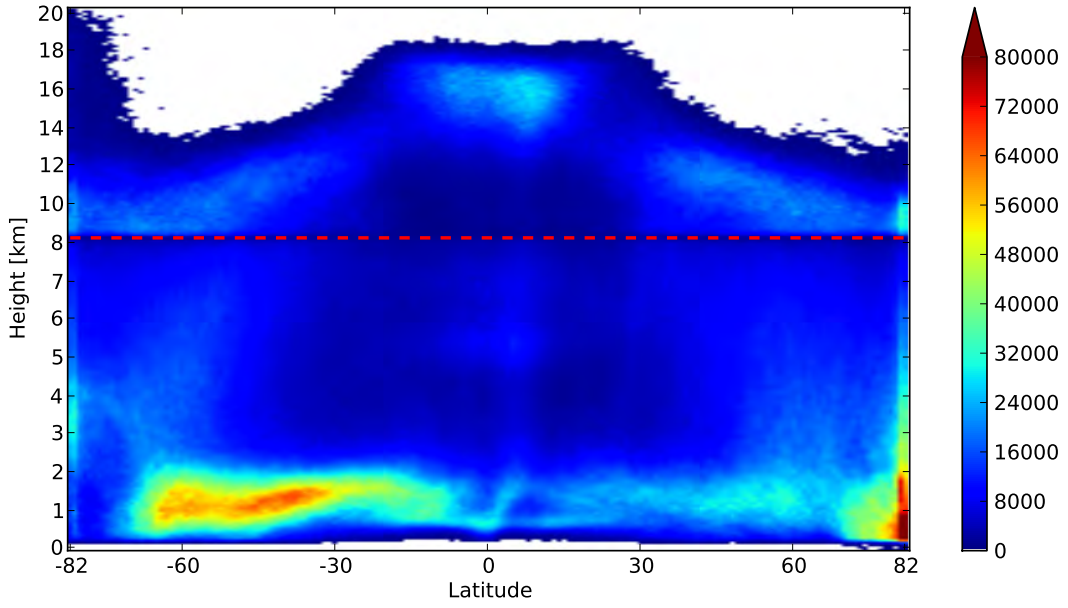


Figure 29: The global cloud top height distribution with 120 m height bins for 374,285,708 cloudy lidar profiles. In (a) the PDF (for 120 m height bins and integral normed) is shown which reveals a maximum peak at 1.32 km. In (b) the cumulative PDF is shown with the median indicated by a green dashed line. The red dashed lines represent the boundary height where the 333 m product migrates into the 5 km product.

As the global cloud height histogram does not reveal the global variability, a two dimensional histogram was calculated. Again the latitude was rounded to its next integer. While the previous vertical binning of 120 m was necessary to harmonize the data set and retrieve a correct probability, the exact determination of the cloud top heights of both data set constituents is most important. For this purpose the vertical binning of the lower troposphere up to 8.22 km was reduced to 60 m, while the upper part remained at the vertical resolution of 120 m. Consequently, a scale-break was introduced at 8.22 km (see Figure 30). As expected, the polar orbit of CALIPSO (see Section 3.2) in combination with the high cloud fraction, reveals a strong accumulation of low level clouds in the arctic region ranging from 75 °N to 82 °N. Additionally, the maritime latitude band from 70 °S to 50 °S with altitudes ranging from 0.5 km to 2 km and dominating cloud fraction throughout the year, can easily be detected (see Figure 30a). Interestingly, the elevated cloud top occurrence caused by the Hadley cell with a maximum above the equator can be seen. The rising air masses, resulting from convection in the tropics, force the uplift of the Tropical Tropopause Layer (TTL) and thus shifting of the top cloud height to greater altitudes. As a result a cloud top occurrence minimum is visible from 45 °S to 45 °N in an altitude range from 2.5 km up to 12 km. Below the dominance of low level clouds in the Atmospheric Boundary Layer (ABL) is preserved. To assess the cloud top distribution statistic further and to account for the overestimation of high latitude

cloud tops, the histogram was normed over its respective latitude degree (Figure 30b) to identify the most likely cloud top heights for one latitude degree.

(a)



(b)

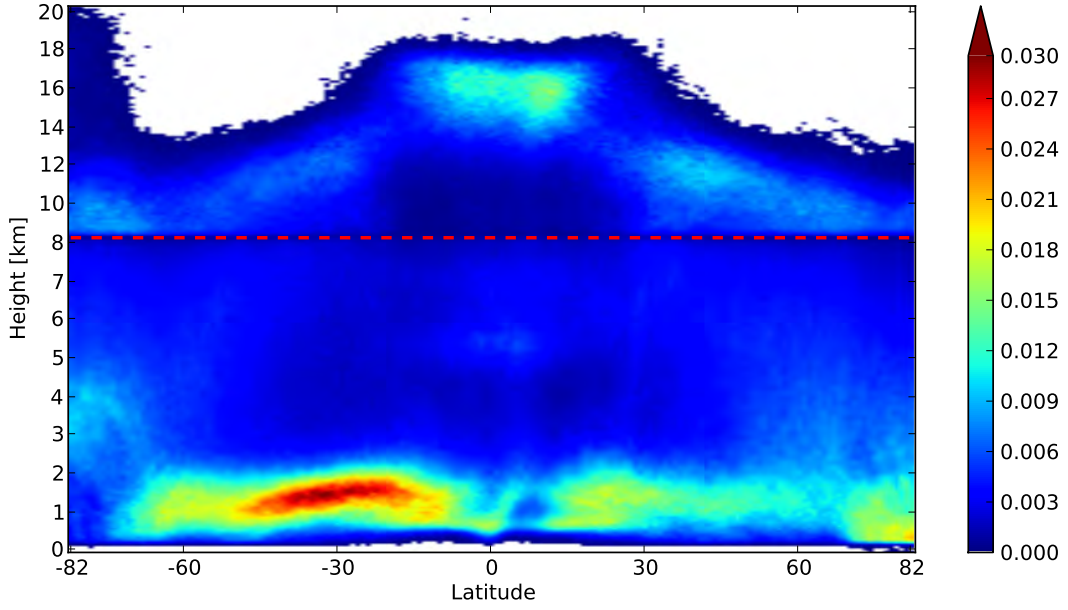


Figure 30: Cloud top height versus latitude distribution of the merged CALIPSO 2007 data. The lower troposphere below 8.2 km with a vertical resolution of 60 m is supplemented by the upper troposphere with a vertical resolution of 120 m. In total 374,285,708 cloud tops were detected. The global distribution can be seen in (a). A peak at the Arctic region arises due to the polar orbit of the satellite. To better understand the local vertical distribution at each latitude (b) the histogram was normed over each rounded latitude degree. The red dashed line represents the boundary at which the 333 m product migrate to the 5 km product.

For the southern hemisphere, the occurrence frequency of high level clouds decreases from around 20 km to a local minimum at around 60 °S, before it starts rising again to a plateau region of 18 km ranging from 15 °S to 15 °N. The high altitude clouds above Antarctica are caused by a phenomenon called Polar Stratospheric Cloud (PSC) [Pitts et al., 2009] and they occur annually in the Antarctic summer (see Figure 31c). They can reach an altitude up to 26 km, which was confirmed by a cloud event with a top height of 26.4 km found in the merged data set. The majority of such ice clouds have been observed near the Antarctic Peninsula or near the Transantarctic Mountains. The primary formation mechanism are orographic waves. Due to the less stable Arctic polar vortex and warmer temperatures the PSC occurrence in the Arctic is significantly smaller, which can also be seen in the histogram [Pitts et al., 2009]. Consequently, the high level cloud amounts increase from higher to lower latitudes in the northern hemisphere. The normed cloud height histogram clearly shows a global peak for clouds ranging from 0.5 km up to 2.5 km with a maximum from 45 °S to 15 °S at around 1.4 km. Interestingly this peak correlates with the cloud top flatness peak discussed in Section 6.2. Mid level clouds up to 8 km occur mainly in the horse latitudes (30 °N and 30 °S) up to the polar fronts with an accumulation from 50 °N to 70 °N in the northern hemisphere. Additionally local minima near 20 °N and 30 °S at around 6 km correspond to the descending circulation of the Hadley cell. The elevated clouds above the Antarctic Peninsula are caused by the topographic elevation of the Antarctic itself. Generally the three maxima (see Figure29a) at the upper troposphere (15-16 km), the mid-troposphere (5-6 km) and the lower troposphere (1-2 km) were previously identified as precipitating clouds ranging from deep cumulus convection to shallow convection and boundary-layer cumuli stratus respectively, as mentioned by [Wu et al., 2011]. The occurrence of the high level cloud band ranging from 9.5 km to 12 km with an uplift of 4 km above the equator (here cirrus clouds are overlaying deep cumulus convection clouds) is based on optically thin cirrus clouds [Schwartz and Mace, 2010, Veglio and Maestri, 2011]. While cirrus clouds can be geometrically thick, they rarely exceed an optical thickness of $\tau > 2.0$. With the merging threshold of $\tau \geq 0.5$, only relatively dense cirrus clouds compose the cirrus band in the histogram.

In a next step the seasonal variation and the prevailing circulations were assessed by computing seasonal averages similar to Chapter 5. Still the global dominance of low-level clouds is preserved. However, a clear seasonally induced horizontal shift of the ITCZ can be observed. Following the solar heating and the warmest surface temperatures it moves towards the northern hemisphere from April through September and reverses from October through March.

Along with the move of the ITCZ into the northern hemisphere a decline in low level clouds with a peak in July to September can be observed. Contrary, the accumulation

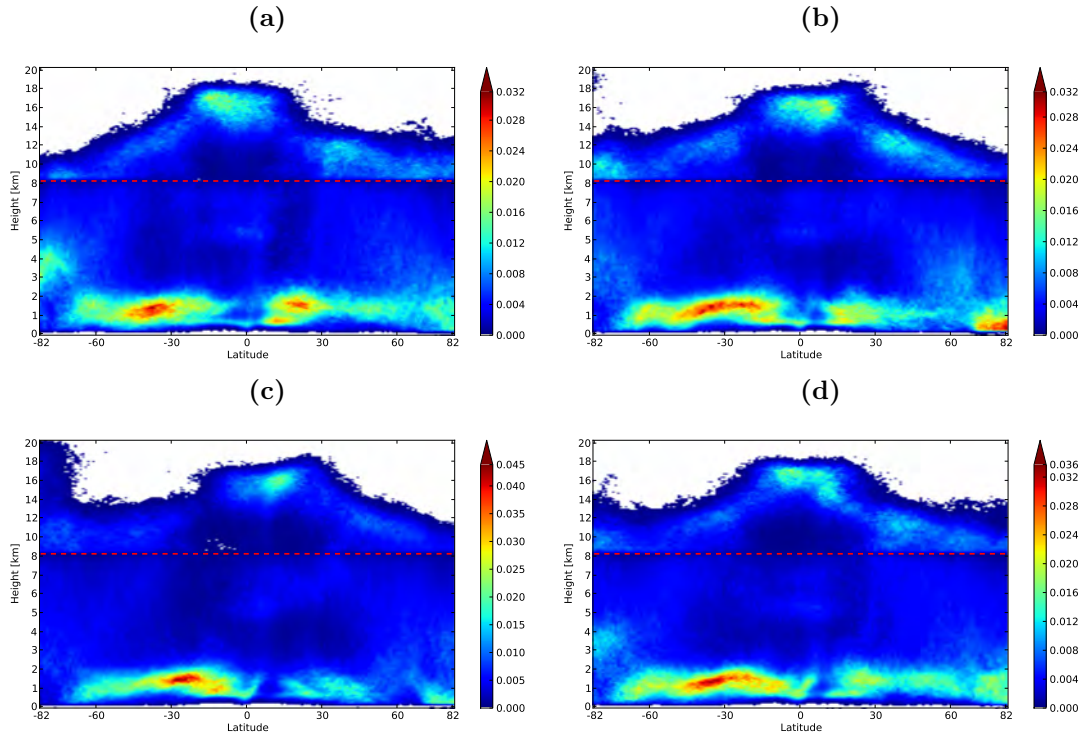


Figure 31: The same as for Figure 30b but for different seasons JFM (a), AMJ (b), JAS (c), OND (d).

of low-level clouds above Antarctica in winter vanishes and the accumulation of PSC's in the summer months becomes the dominant cloud type. While the cloud plateau above the equator at 18 km remain stable and moves with the ITCZ, the high level cloud occurrence in the northern hemisphere varies with a maximum in winter and a minimum in summer. Mid-level cloud amounts (5-7 km) over the midlatitudes also oscillate with the seasons and the same holds for low level (1-2 km) clouds.

7.2 Conclusion

In summary, the global cloud top height distribution is in accordance with recent studies and especially for [Wu et al., 2011] which also used 5 km CALIPSO L2 Cloud Layer data for the entire year 2007. Consequently the merged data set describes well the atmospheric situation (temporal and spatial) and provides the most realistic simulation for a future lidar mission. With global cirrus coverage, PSC's and deep convective cumulus the merging process added crucial cloud information of the upper troposphere that affects the measurement performance of MERLIN's methane retrieval. Therefore the data set presented within this thesis is the best possible approximation of space-lidar measurements with a similar configuration as MERLIN.

Chapter 8

Conclusion

8.1 Space lidar for methane measurements

While current observational networks operated by various initiatives and organizations (NOAA, WDCGG, etc.) are able to retrieve large scale patterns of CH_4 mixing ratios in the atmosphere, their structure and geographical distribution prohibit a continuous global high resolution observation. To circumvent this lack and to enhance the scientific understanding and the monitoring capabilities of natural and anthropogenic methane sources, a German-French climate monitoring initiative, MERLIN, was proposed by DLR and CNES in 2010. With such a space lidar at hand, methane measurements could be obtained with an accuracy of 0.1 % and 1 % precision with a nominal resolution of 50 km which would complement today's observation possibilities. To ensure such a high resolution CH_4 retrieval from a platform that is at a distance of 503 km above the Earth's surface while traveling about 7 km/s, not only the IPDA lidar instrument and the retrieval algorithms have to work efficiently, but also the atmospheric constraints have to be assessed. With actually no space lidar dedicated to global carbon emission present and due to the innovative approach of MERLIN to measure methane with an active instrument, a simulation and performance assessment of such a mission is challenging. Therefore the best possible approximation of the measurement environment MERLIN would encounter, was achieved by using CALIPSO cloud layer data as they best correspond to MERLIN's measurement geometry. This study uses single shot CALIPSO data from the lower troposphere with a horizontal resolution of 333 m and 30 m vertical resolution together with cloud data of the upper troposphere with an optical thickness greater than 0.5 to simulate atmospheric conditions, that would most likely cause a complete attenuation of the laser signal and would therefore lead to a lack of surface measurements necessary to compute a methane mixing ratio. Disregarding

the existing differences of both satellite missions, with MERLIN flying 200 km lower on a dawn/dusk orbit which is 4:30 h ahead of CALIPSO's equatorial crossing time, and the lower pulse energy for the IPDA instrument, only slightly regional differences in the atmospheric cloud cover are expected. While MERLIN might encounter morning ground fog because of its dawn/dusk orbit, CALIPSO has a higher probability of convective cloud detection due to its afternoon equatorial crossing. Generally such regional differences are suppressed by large scale seasonal changes which are attended by radical changing atmospheric compositions. As the only possibility of a performance assessment is based on statistical analysis of long period data sets, regional and short term differences cannot be accurately assessed. Consequently all results presented in this thesis only assess the mean probability of a successful surface detection with a resolution of 2 ° x 2 ° in latitude/longitude. While the horizontal resolution of 333 m generally allows the assessment of small broken cloud occurrences, the period of one year together with the 16 day repeat cycle of CALIPSO only gives approximately 23 discrete measurements of each individual measurement point. Thus, a spatial averaging is necessary to get sufficient statistical relevance and to derive valid regional conclusions. The same holds for the assessment of cloud top flatness events, which were computed for 1 ° x 1 ° grid points in latitude/longitude. With only approximately 1.3 % of all cloudy measurements contributing to such events, the statistical relevance is questionable and only the location and height trend is of interest because such signals could allow partial vertical methane column retrievals down to the flat cloud top. Additionally, broken cloud scenes with occasionally successful surface measurements could be combined with partial columns of sharp edged lower troposphere clouds to construct a vertical profile of methane concentration. As methane sources lie within altitudes of 0 to 3 km and are originated in densely populated areas, boreal and tropical wetlands, as well as in permafrost regions, this would enable a separation of the local sources from the methane background concentration, at least on such occasions.

8.2 Estimated measurement density

Finally, the expected performance derived from one year of cloud layer data of CALIPSO and various statistical assessments is presented in Figure 32. To account for the different sources of possible methane retrievals the resolution was degraded again with a contour representation indicating possible measurement regions. This performance assessment is more of quantitative nature indicating regions of highest probability for a methane retrieval. For the underlying calculations of cloud free fraction, cloud gap occurrence frequency, global cloud gap distribution, cloud top flatness and global cloud top height distribution, the reader is referred to Chapter 5,6 and 7.

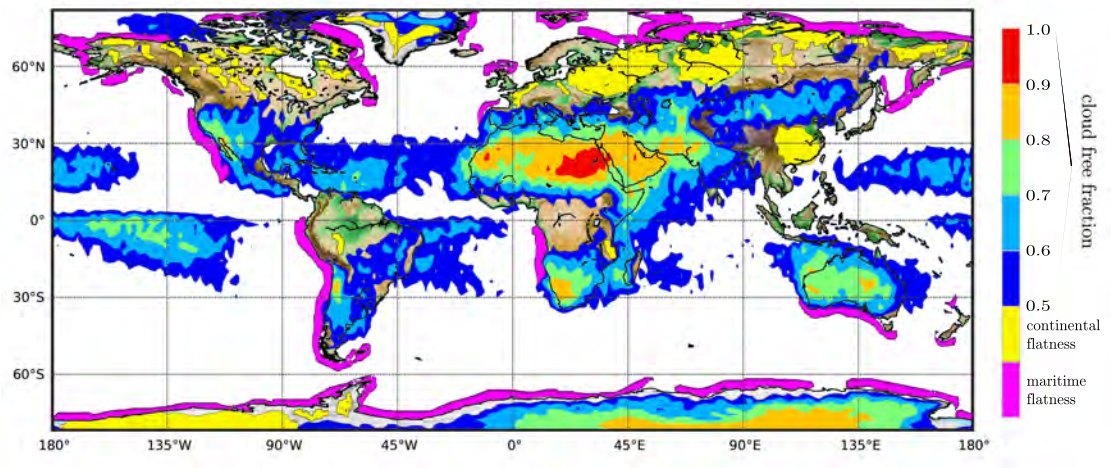


Figure 32: The overall estimated performance based on the cloud free fraction (see Figure 16) and the cloud flatness events (see Figure 26). Those events are separated in continental seasonal cloud top flatness events indicated in yellow and near continental maritime events indicated in purple. The underlying global relief map for the continents is based on the Etopo world map of NOAA (`http://www.ngdc.noaa.gov/mgg/global/`). In this map the white areas designate regions with cloud free fractions < 0.5 where space lidar measurements are less performant.

In summary 40.5 % of all measurements were cloud free with above average cloud free fraction in the Antarctica, the subtropics and the tropics with an exception of Indonesia, north South America and Mid Africa. Thereby all cloud gaps found follow a global power law with exponent $\alpha = 1.51 \pm 0.01$ with anomalies in the Sahara, the Arabian Peninsula, Australia, the west coast of North and South America as well as in Greenland, Antarctica and all high altitude mountain ranges. In 2007 a mean cloud gap length of 7.41 km and a median cloud gap length of around 1 km were found. Cloud gaps most often occur above tropical and subtropical oceans whereas cloud top flatness events with at least 50 km horizontal extent over all oceans ranging from $\pm 82\,^\circ$ to $\pm 30\,^\circ$ and on the west coast of each continent. These events are complemented by seasonal dependent events above the interior of continents. All of them do not exceed 3 km height above the underlying surface and therefore could be used for a partial column retrieval.

Taking all these results into account and applying a threshold of 50 % cloud free fraction while focusing on interesting maritime methane sources along the continental seabed under the exclusion of open water retrievals, gives an overview of preferable measurement regions (see Figure 32). Nearly all land masses possess a high density of methane measurements except Indonesia, India, the Tibetian Plateau, great parts of Siberia, North-West Europe, Scandinavia, Mid Africa, great parts of North America and New Zealand. However, seasonal variations and the shift of the ITCZ can improve the performance for higher latitudes on the northern hemisphere in summer. Also worth mentioning is

the fact that nearly all continental boundaries accumulate events of cloud top flatness. Consequently the overall performance for a future space lidar looks quite promising.

8.3 Outlook

While MERLIN is currently in phase B and all mission components are planned in detail, this analysis basically supports past studies [Kiemle et al., 2011] and for the first time ever allows conclusions about cloud gaps. Consequently not only MERLIN could benefit from this study but also any other future space lidar as well as passive remote sensing missions whose measurements often are biased by clouds. With ADM Aeolus and EarthCare which are part of the Earth Explorer Program of ESA, two other near future lidar satellites are expected to go into space in the next few years. Thanks to the software developed as part of this thesis, a future modified or extented analysis could be performed in a very short period of time. For instance an analysis of airborne lidar campaigns that fly in close geographical proximity to the CALIPSO ground track could be performed without the need of additional programming. With a launch date of MERLIN aimed at 2017, many validation and testing campaigns are likely to happen. Furthermore, the statistical results of this thesis are a starting point for future similar performance assessments. It is also worth mentioning that for the first time ever CALIPSO 333 m cloud layer products were used for such cloud statistics. This approach will certainly be used in future studies on clouds and cloud gaps.

Appendix A

Detailed results and additional information

The following additional information and results serve the reader as a reference for the composition of the used and converted CALIPSO cloud layer products (see Chapter 3). Additionally to the merged data structure (see Table 11 and 12), detailed cloud free fraction results for different latitude zones are presented for each monthly data set as well as on an annual mean (see Table 13). As the presented power law fit of the cloud gap occurrence frequency needs prior logarithmic binning, the process is briefly explained. In addition the σ and x_{min} values for the best power law fit of 2 ° x 2 ° grid points (see Figure 22 are presented in Figure 33. Finally seasonal cloud flatness events (see Figure 35 as well as the total amount (unnormed) of the cloud top height distribution is presented in Figure 36.

CALIPSO converted data set

The converted and combined data sets consist of the data arrays presented in Table 11 and 12.

Parameter	Data Type	Units	Elem/Rec
Profile ID	int32	NoUnits	1
Latitude	float32	deg	1
Longitude	float32	deg	1
Profile UTC Time	float64	NoUnits	1
Day Night Flag	int8	NoUnits	1
Scattering Angle	float32	deg	1
Parallel Column Reflectance 532	float32	NoUnits	1
Perpendicular Column Reflectance 532	float32	NoUnits	1
Column Integrated Attenuated Backscatter 532	float32	sr^{-1}	1
IGBP Surface Type	int8	NoUnits	1
NSIDC Surface Type	uint8	NoUnits	1
Lidar Surface Elevation	float32	km	1*
DEM Surface Elevation	float32	km	1
Number Layers Found	int8	NoUnits	1
Layer Top Altitude	float32	km	5
Layer Base Altitude	float32	km	5
Integrated Attenuated Backscatter 532	float32	sr^{-1}	5
Integrated Attenuated Backscatter Uncertainty 532	float32	sr^{-1}	5
Overlying Integrated Attenuated Backscatter 532	float32	sr^{-1}	5
Feature Classification Flags	uint16	NoUnits	5

Table 11: Lidar 1/3 km Cloud Layer and Column Descriptor Record

Parameter	Data Type	Units	Elem/Rec
Profile ID	int32	NoUnits	2
Latitude	float32	deg	3
Longitude	float32	deg	3
Profile UTC Time	float64	NoUnits	3
Day Night Flag	int8	NoUnits	1
Column Optical Depth Cloud 532	float32	NoUnits	1
Column Optical Depth Cloud Uncertainty 532	float32	NoUnits	1
Column Optical Depth Aerosols 532	float32	NoUnits	1
Column Optical Depth Aerosols Uncertainty 532	float32	NoUnits	1
Column Optical Depth Stratospheric 532	float32	NoUnits	1
Column Optical Depth Stratospheric Uncertainty 532	float32	NoUnits	1
IGBP Surface Type	int8	NoUnits	1
NSIDC Surface Type	uint8	NoUnits	1
Lidar Surface Elevation	float32	km	7*
DEM Surface Elevation	float32	km	4
FeatureFinderQC	uint16	NoUnits	1
Number Layers Found	int8	NoUnits	1
Layer Top Altitude	float32	km	10
Layer Base Altitude	float32	km	10
Integrated Volume Depolarization Ratio	float32	NoUnits	10
Feature Classification Flags	uint16	NoUnits	10
Feature Optical Depth 532	float32	NoUnits	10

Table 12: Lidar 5 km Cloud Layer and Column Descriptor Record

Detailed cloud free fraction for the entire year 2007

	product	82 °S-60 °S	60 °S-30 °S	30 °S-0 °	0 °-30 °N	30 °N-60 °N	60 °N-82 °N
	333m	38.54	38.24	63.33	66.64	50.49	50.79
Jan	5km	27.49	25.03	30.34	42.30	31.17	26.25
	merged	35.13	27.02	43.71	55.45	37.07	40.88
	333m	36.29	37.38	67.40	69.57	49.84	55.29
Feb	5km	25.83	24.03	31.00	48.46	27.98	29.56
	merged	33.54	25.81	45.21	60.09	35.12	48.15
	333m	38.94	37.49	66.68	70.36	51.56	54.55
Mar	5km	25.46	24.84	33.51	46.22	29.76	30.27
	merged	34.74	25.84	46.90	58.92	37.84	48.74
	333m	42.06	36.74	67.15	70.90	54.04	47.47
Apr	5km	25.62	24.31	36.91	39.38	28.03	26.98
	merged	35.67	24.92	50.42	56.47	38.90	41.44
	333m	48.92	37.38	65.92	70.14	53.72	35.75
May	5km	27.45	24.63	45.10	33.77	25.03	20.46
	merged	38.55	24.16	54.40	51.38	38.05	31.90
	333m	51.00	38.01	63.02	69.06	53.95	40.73
Jun	5km	25.72	25.99	45.02	30.97	27.78	25.30
	merged	38.54	24.78	52.89	48.34	39.29	35.85
	333m	54.13	37.69	60.23	66.82	56.08	43.45
Jul	5km	22.70	26.36	46.70	29.37	31.86	26.33
	merged	40.33	25.30	52.08	46.36	41.48	36.86
	333m	54.14	36.57	58.43	67.15	57.74	35.56
Aug	5km	21.96	25.60	46.64	29.65	35.91	19.95
	merged	40.84	24.52	51.64	44.67	45.49	28.07
	333m	54.82	34.04	59.21	69.73	58.37	28.20
Sep	5km	25.17	23.51	44.17	32.49	37.69	14.43
	merged	41.82	23.22	51.32	47.73	45.72	20.99
	333m	51.43	34.45	58.67	71.43	55.73	28.62
Oct	5km	27.57	23.79	37.86	37.21	34.96	13.78
	merged	40.94	24.24	47.55	52.33	41.91	20.81
	333m	44.18	34.82	59.81	71.33	51.89	39.47
Nov	5km	27.55	24.51	31.55	40.33	31.49	20.13
	merged	38.33	24.95	43.54	55.63	37.94	31.10
	333m	43.73	36.45	62.64	68.54	51.59	51.11
Dec	5km	30.35	23.28	30.99	44.00	29.69	25.18
	merged	39.45	25.46	43.60	55.79	36.97	42.23
Mean	333m	46.52	36.61	62.71	69.30	53.75	42.58
2007	5km	26.07	24.66	38.32	37.85	31.03	23.21
	merged	38.16	25.01	48.61	52.76	39.65	35.59

Table 13: The mean monthly cloud free fraction in % averaged by latitudal zones and different products. The single shot (333m) product has the highest probability of being cloudless, while the 5km product provides additional clouds due to extensive averaging and post-processing. The merged product combines both data according to Chapter 3.

Logarithmic binning

By definition a bin of constant logarithmic width means that the logarithm of the upper edge of a bin (x_{i+1}) is equal to the logarithm of the lower edge of that bin x_i plus the bin width b_w. That is,

$$log(x_{i+1} = log(x_i) + b_w$$

Each binned value x_b can easily be calculated by the average of all x_i for i=1,..,n such that i falls in the bin.

$$x_b = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Note that n is the number of integers that fall into each given bin regardless of whether x_b is zero or not. Furthermore, for the last bin all integers, even if they are beyond the last data point, have to be included in n.

Power law fit uncertainty

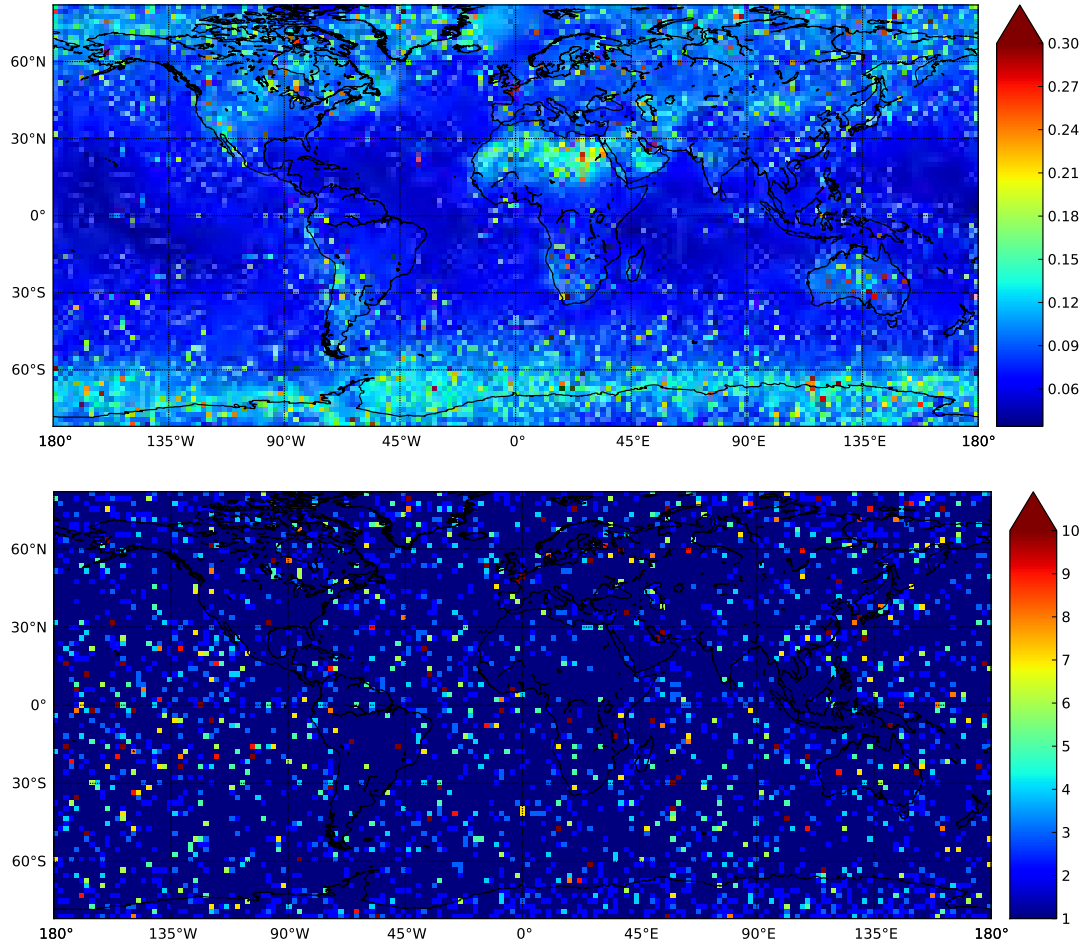
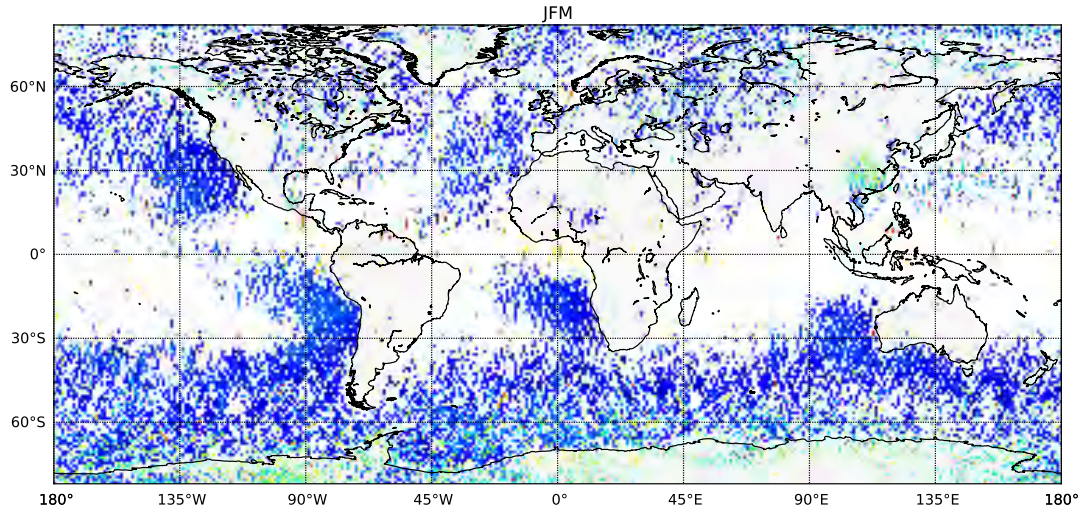


Figure 33: The resulting σ (top) and x_min (bottom) values for the discrete best fit power law exponent α computed for each 2 ° x 2 ° grid point (complete 2007 data set). x_{min} is thereby directly correlated to the cloud gap length by $x_i = i * 0.333km$. The threshold of $\sigma = 0.1$ indicate a high uncertainty in the power law distribution.

Cloud flatness events

(a)



(b)

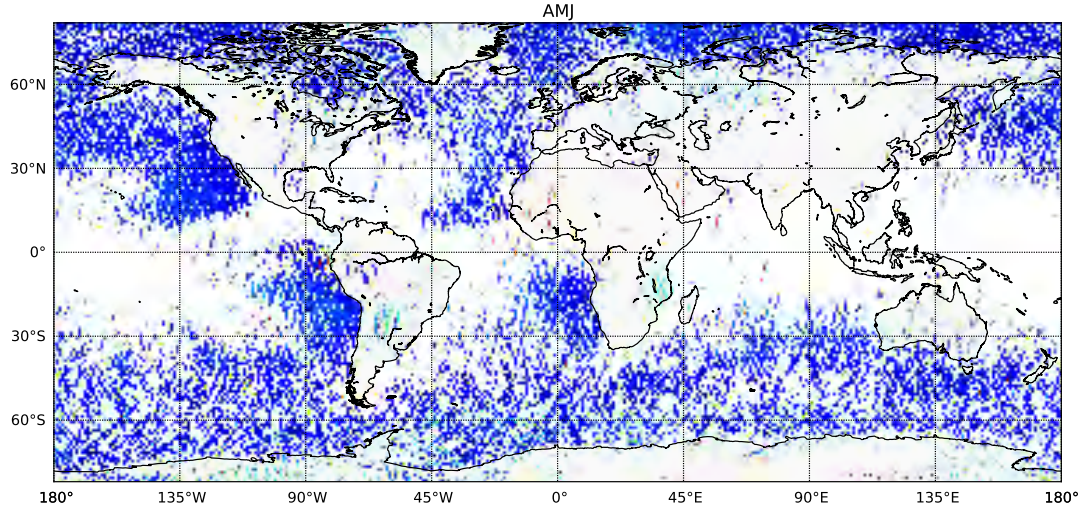
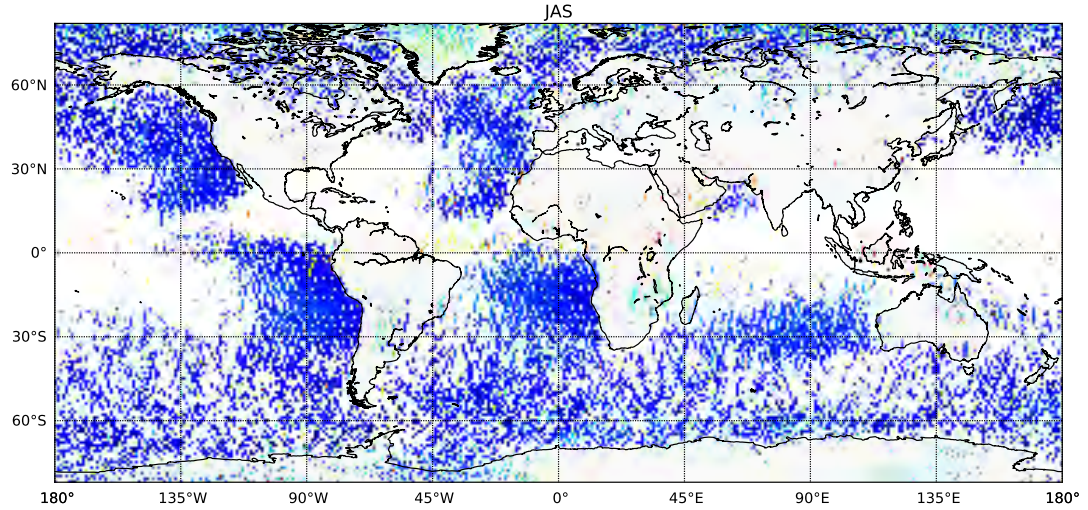


Figure 34: The same as for Figure 26a but for different seasons JFM (a), AMJ (b), JAS (c), OND (d).

(a)



(b)

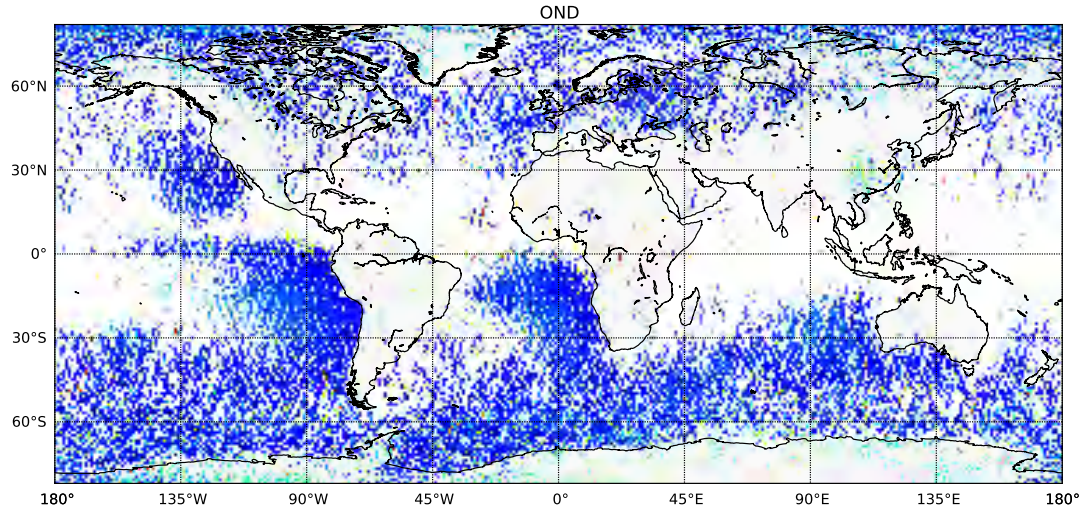


Figure 35: The same as for Figure 26a but for different seasons JAS (a), OND (b).

Cloud cloud top distribution The total amount of cloud tops vertically binned by 60 m in the lower troposphere up to 8.2 km and 120 m in the upper troposphere. The polar peaks are due to to the polar orbit of CALIPSO causing more measurements above high latitudes.

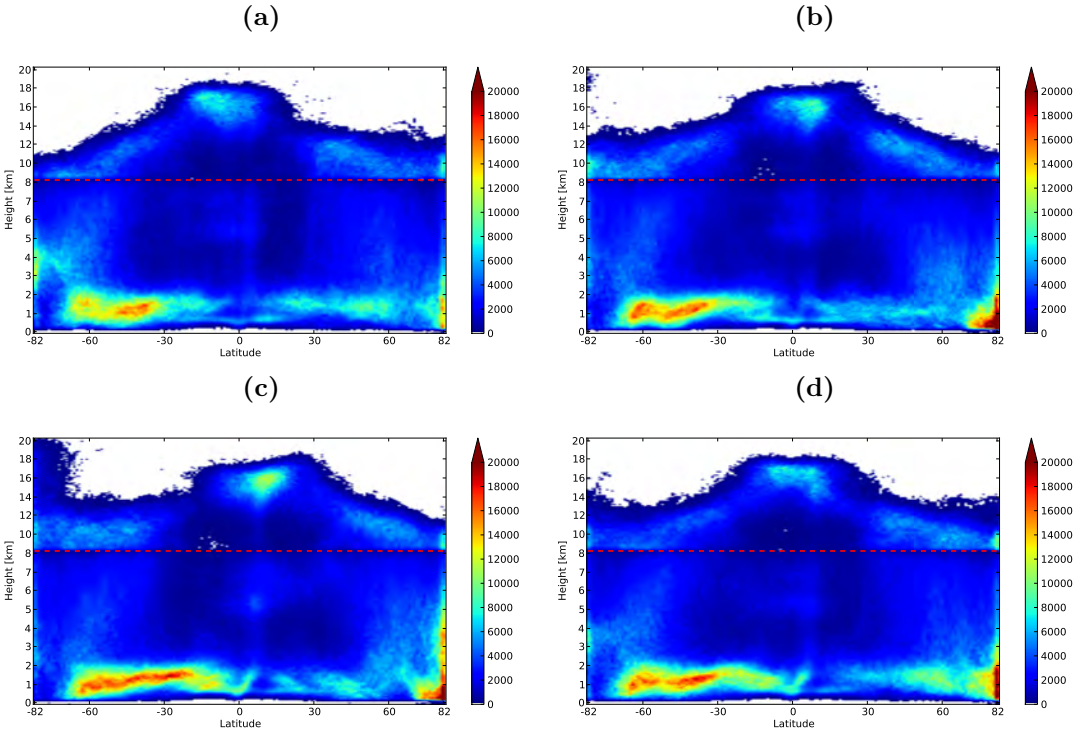


Figure 36: The same as for Figure 30a but for different seasons JFM (a), AMJ (b), JAS (c), OND (d).

Bibliography

J. Alstott, E. Bullmore, and D. Plenz. powerlaw: a python package for analysis of heavy-tailed distributions. *ArXiv e-prints*, 2013.

I. Astin and C. Kiemle. Space-borne clear air lidar measurements in the presence of broken cloud. *Annales Geophysicae*, 21(3):639–647, 2003. doi: 10.5194/angeo-21-639-2003. URL `http://www.ann-geophys.net/21/639/2003/`.

P. Bergamaschi, S. Houweling, A. Segers, M. Krol, C. Frankenberg, R. A. Scheepmaker, E. Dlugokencky, S. C. Wofsy, E. A. Kort, C. Sweeney, T. Schuck, C. Brenninkmeijer, H. Chen, V. Beck, and C. Gerbig. Atmospheric ch4 in the first decade of the 21st century: Inverse modeling analysis using sciamachy satellite retrievals and noaa surface measurements. *Journal of Geophysical Research: Atmospheres*, 118(13):7350–7369, 2013. ISSN 2156-2202. doi: 10.1002/jgrd.50480. URL `http://dx.doi.org/10.1002/jgrd.50480`.

S. Berthier, P. Chazette, J. Pelon, and B. Baum. Comparison of cloud statistics from spaceborne lidar systems. *Atmospheric Chemistry and Physics*, 8(23):6965–6977, 2008. doi: 10.5194/acp-8-6965-2008. URL `http://www.atmos-chem-phys.net/8/6965/2008/`.

M. A. Chan and J. C. Comiso. Cloud features detected by modis but not by cloudsat and caliop. *Geophysical Research Letters*, 38(24):n/a–n/a, 2011. ISSN 1944-8007. doi: 10.1029/2011GL050063. URL `http://dx.doi.org/10.1029/2011GL050063`.

P. Chazette, J. Pelon, and G. Mégie. Determination by spaceborne backscatter lidar of the structural parameters of atmospheric scattering layers. *Appl. Opt.*, 40(21):3428–3440, 2001. doi: 10.1364/AO.40.003428.

A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009. ISSN 0036-1445. doi: 10.1137/070710111.

E. J. Dlugokencky, L. Bruhwiler, White, J. W. C., L. K. Emmons, P. C. Novelli, S. A. Montzka, K. A. Masarie, P. M. Lang, A. M. Crotwell, J. B. Miller, and L. V. Gatti. Observational constraints on recent increases in the atmospheric ch4 burden. *Geophysical*

Research Letters, 36(18):n/a–n/a, 2009. ISSN 1944-8007. doi: 10.1029/2009GL039780. URL http://dx.doi.org/10.1029/2009GL039780.

E. J. Dlugokencky, E. G. Nisbet, R. Fisher, and D. Lowry. Global atmospheric methane: budget, changes and dangers. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1943):2058–2072, 2011. ISSN 1364-503X. doi: 10.1098/rsta.2010.0341.

G. D. Emmitt, Simpson Weather Associates, Charlottesville, VA, S. Greco, D. M. Winker, and Y. Hu. Cflos and cloud statistics from satellite and their impact on future space-based doppler wind lidar development. In *Symposium on Recent Developments in Atmospheric Applications of Radar and Lidar*, volume Poster Session 2, Recent Developments in Atmospheric Applications of Radar and lidar. 2012. URL https://ams.confex.com/ams/88Annual/techprogram/paper_134230.htm.

P. Forster, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, Fahey, D. W., J. Haywood, J. Lean, Lowe, D. C., G. Myhre, J. Nganga, R. Prinn, G. Raga, M. Schultz, and R. van Dorland. Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change: Changes in atmospheric constituents and in radiative forcing, 2007. URL http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-wg1-chapter2.pdf.

W. H. Hunt, D. M. Winker, M. A. Vaughan, K. A. Powell, P. L. Lucker, and C. Weimer. Calipso lidar description and performance assessment. *J. Atmos. Oceanic Technol.*, 26(7):1214–1228, 2009. ISSN 0739-0572. doi: 10.1175/2009JTECHA1223.1. URL http://dx.doi.org/10.1175/2009JTECHA1223.1.

K.-G. Karlsson and E. Johansson. On the optimal method for evaluating cloud products from passive satellite imagery using calipso-caliop data: example investigating the cm saf clara-a1 dataset. *Atmospheric Measurement Techniques*, 6(5):1271–1286, 2013. doi: 10.5194/amt-6-1271-2013. URL http://www.atmos-meas-tech.net/6/1271/2013/.

C. Karney and R. Deakin. F.w. bessel (1825): The calculation of longitude and latitude from geodesic measurements. *Astronomische Nachrichten*, 331(8):852–861, 2010. ISSN 1521-3994. doi: 10.1002/asna.201011352. URL http://dx.doi.org/10.1002/asna.201011352.

C. Kiemle, M. Quatrevalet, G. Ehret, A. Amediek, A. Fix, and M. Wirth. Sensitivity studies for a space-based methane lidar mission. *Atmospheric Measurement Techniques*, 4(10):2195–2211, 2011. doi: 10.5194/amt-4-2195-2011. URL http://www.atmos-meas-tech.net/4/2195/2011/.

A. Klaus, S. Yu, and D. Plenz. Statistical analyses support power law distributions found in neuronal avalanches. *PLoS ONE*, 6(5):e19779, 2011. doi: 10.1371/journal.pone.0019779. URL http://dx.doi.org/10.1371%2Fjournal.pone.0019779.

L. V. Leahy, R. Wood, R. J. Charlson, C. A. Hostetler, R. R. Rogers, M. A. Vaughan, and D. M. Winker. On the nature and extent of optically thin marine low clouds. *Journal of Geophysical Research: Atmospheres*, 117(D22):n/a–n/a, 2012. ISSN 2156-2202. doi: 10.1029/2012JD017929. URL http://dx.doi.org/10.1029/2012JD017929.

Z. Liu, A. Omar, Y. Hu, M. A. Vaughan, and D. M. Winker. Part 3: scene classification algorithm. In D. M. Winker, C. A. Hostetler, M. A. Vaughan, and A. Omar, editors, *CALIOP Algorithm Theoretical Basis Document*. 2006. URL http://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202_Part3_v1.0.pdf.

Z. Liu, M. A. Vaughan, D. M. Winker, C. Kittaka, B. Getzewich, R. Kuehn, A. Omar, K. Powell, C. Trepte, and C. Hostetler. The calipso lidar cloud and aerosol discrimination: Version 2 algorithm and initial assessment of performance. *J. Atmos. Oceanic Technol.*, 26(7):1198–1213, 2009. ISSN 0739-0572. doi: 10.1175/2009JTECHA1229.1. URL http://dx.doi.org/10.1175/2009JTECHA1229.1.

Z. Liu, R. Kuehn, M. Vaughan, D. Winker, A. Omar, K. Powell, C. Trepte, Y. Hu, and C. A. Hostetler. The calipso cloud and aerosol discrimination: version 3 algorithm and test results. In *Proceedings of the 25th International Laser Radar Conference, St.-Petersburg, 5-9 July 2010*. 2010.

M. C. Pitts, L. R. Poole, and L. W. Thomason. Calipso polar stratospheric cloud observations: second-generation detection algorithm and composition discrimination. *Atmospheric Chemistry and Physics*, 9(19):7577–7589, 2009. doi: 10.5194/acp-9-7577-2009. URL http://www.atmos-chem-phys.net/9/7577/2009/.

A. Razavi, C. Clerbaux, C. Wespes, L. Clarisse, D. Hurtmans, S. Payan, C. Camy-Peyret, and P. F. Coheur. Characterization of methane retrievals from the iasi spaceborne sounder. *Atmospheric Chemistry and Physics*, 9(20):7889–7899, 2009. doi: 10.5194/acp-9-7889-2009. URL http://www.atmos-chem-phys.net/9/7889/2009/.

D. L. Reinke and Vonder Haar, T. H. Probability of cloud-free-line-of-sight (pcflos) derived from cloudsat and calipso cloud observations. In *2011 EUMETSAT Meteorological Satellite Conference*. 2011. URL ftp://ftp.cira.colostate.edu/ftp/CloudSat/EUMETSAT_SAT_CONF_2011_Reinke_PCFLOS.pdf.

J. A. Rice. *Mathematical statistics and data analysis*. Duxbury Press, Belmont and Cal, 1995. ISBN 0-534-20934-3.

M. Sasakawa, K. Shimoyama, T. Machida, N. Tsuda, H. Suto, M. Arshinov, D. Davydov, A. Fofonov, O. Krasnov, T. Saeki, Y. Koyama, and S. Maksyutov. Continuous measurements of methane from a tower network over siberia. *Tellus B*, 62(5), 2011. ISSN 1600-0889, url = http://www.tellusb.net/index.php/tellusb/article/view/16583.

M. C. Schwartz and G. G. Mace. Co-occurrence statistics of tropical tropopause layer cirrus with lower cloud layers as derived from cloudsat and calipso data. *Journal of Geophysical Research: Atmospheres*, 115(D20):n/a–n/a, 2010. ISSN 2156-2202. doi: 10.1029/2009JD012778. URL `http://dx.doi.org/10.1029/2009JD012778`.

C. Stephan, M. Alpers, B. Millet, G. Ehret, P. Flamant, and C. Deniel. Merlin: a space-based methane monitor. *Proc. SPIE*, 8159, 2011. doi: 10.1117/12.896589. URL `http://dx.doi.org/10.1117/12.896589`.

M. A. Vaughan, D. M. Winker, and K. A. Powell. Part 2: Feature detection and layer properties algorithms. In D. M. Winker, C. A. Hostetler, M. A. Vaughan, and A. Omar, editors, *CALIOP Algorithm Theoretical Basis Document*. 2006. URL `http://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202_Part2_rev1x01.pdf`.

M. A. Vaughan, K. A. Powell, D. M. Winker, C. A. Hostetler, R. E. Kuehn, W. H. Hunt, B. J. Getzewich, S. A. Young, Z. Liu, and M. J. McGill. Fully automated detection of cloud and aerosol layers in the calipso lidar measurements. *J. Atmos. Oceanic Technol.*, 26(10):2034–2050, 2009. ISSN 0739-0572. doi: 10.1175/2009JTECHA1228.1. URL `http://dx.doi.org/10.1175/2009JTECHA1228.1`.

P. Veglio and T. Maestri. Statistics of vertical backscatter profiles of cirrus clouds. *Atmospheric Chemistry and Physics*, 11(24):12925–12943, 2011. doi: 10.5194/acp-11-12925-2011. URL `http://www.atmos-chem-phys.net/11/12925/2011/`.

J. M. Wallace and P. V. Hobbs. *Atmospheric science: An introductory survey*, volume vol. 92 of *International geophysics series*. Elsevier Academic Press, Amsterdam, 2nd ed edition, 2006. ISBN 978-0-12-732951-2.

C. Weaver, C. Kiemle, S. R. Kawa, T. Aalto, J. Necki, M. Steinbacher, J. Arduini, F. Apadula, H. Berkhout, J. Hatakka, and S. O'Doherty. Retrieval of methane source strengths in europe using a simple modeling approach to assess the potential of space-borne lidar observations. *Atmospheric Chemistry and Physics Discussions*, 13(7):19559–19582, 2013. doi: 10.5194/acpd-13-19559-2013. URL `http://www.atmos-chem-phys-discuss.net/13/19559/2013/`.

E. P. White, B. J. Enquist, and J. L. Green. On estimating the exponent of power-law frequency distributions. *Ecology*, 89(4):905–912, 2008. ISSN 0012-9658. doi: 10.1890/07-1288.1. URL http://dx.doi.org/10.1890/07-1288.1.

D. M. Winker, Y. Hu, M. C. Pitts, J. Tackett, C. Kittaka, Z. Liu, and M. A. Vaughan. Calipso at four: results and progress. URL http://ntrs.nasa.gov/search.jsp?R=20110012417.

D. M. Winker, C. A. Hostetler, M. A. Vaughan, and A. Omar. Part 1: Caliop instrument, and algortihms overview release 2.0. In D. M. Winker, C. A. Hostetler, M. A. Vaughan, and A. Omar, editors, *CALIOP Algorithm Theoretical Basis Document*. 2006. URL http://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202.Part1_v2-Overview.pdf.

D. M. Winker, M. A. Vaughan, A. Omar, Y. Hu, K. A. Powell, Z. Liu, W. H. Hunt, and S. A. Young. Overview of the calipso mission and caliop data processing algorithms. *J. Atmos. Oceanic Technol.*, 26(11):2310–2323, 2009. ISSN 0739-0572. doi: 10.1175/2009JTECHA1281.1. URL http://dx.doi.org/10.1175/2009JTECHA1281.1.

D. M. Winker, Y. Hu, M. C. Pitts, M. Avery, B. J. Getzewich, J. Tackett, C. Kittaka, Z. Liu, and M. A. Vaughan. The calipso mission: results and progress. *Proc. SPIE*, 7832, 2010a. doi: 10.1117/12.864901. URL http://dx.doi.org/10.1117/12.864901.

D. M. Winker, J. Pelon, Coakley, J. A., Ackerman, S. A., Charlson, R. J., Colarco, P. R., P. Flamant, Q. Fu, Hoff, R. M., C. Kittaka, Kubar, T. L., H. Le Treut, McCormick, M. P., G. Mégie, L. Poole, K. Powell, C. Trepte, Vaughan, M. A., and Wielicki, B. A. The calipso mission: A global 3d view of aerosols and clouds. *Bull. Amer. Meteor. Soc.*, 91(9):1211–1229, 2010b. ISSN 0003-0007. doi: 10.1175/2010BAMS3009.1. URL http://dx.doi.org/10.1175/2010BAMS3009.1.

R. Wood and P. R. Field. The distribution of cloud horizontal sizes. *J. Climate*, 24 (18):4800–4816, 2011. ISSN 0894-8755. doi: 10.1175/2011JCLI4056.1. URL http://dx.doi.org/10.1175/2011JCLI4056.1.

D. Wu, Yongxiang Hu, M. McCormick, K.-M. Xu, Z. Liu, B. Smith, A. Omar, and F.-L. Chang. Deriving marine-boundary-layer lapse rate from collocated calipso, modis, and amsr-e data to study global low-cloud height statistics. *Geoscience and Remote Sensing Letters, IEEE*, 5(4):649–652, 2008. ISSN 1545-598X. doi: 10.1109/LGRS.2008.2002024.

D. Wu, Y. Hu, McCormick, M. P., and F. Yan. Global cloud-layer distribution statistics from 1 year calipso lidar observations. *Int. J. Remote Sens.*, 32(5):1269–1288, 2011.

ISSN 0143-1161. doi: 10.1080/01431160903530821. URL http://dx.doi.org/10.1080/01431160903530821.

S. Young. Just how far does calipso's lidar profile into clouds? page 4 p, 2012.

C. Zhao, A. E. Andrews, L. Bianco, J. Eluszkiewicz, A. Hirsch, C. MacDonald, T. Nehrkorn, and M. L. Fischer. Atmospheric inverse estimates of methane emissions from central california. *Journal of Geophysical Research: Atmospheres*, 114 (D16):n/a–n/a, 2009. ISSN 2156-2202. doi: 10.1029/2008JD011671. URL http://dx.doi.org/10.1029/2008JD011671.