Leipzig University
Faculty of Physics and Earth Science
Leipzig Institute for Meteorology

MASTER THESIS

Variation of cloud horizontal sizes and cloud fraction over Europe 1985–2018 in high-resolution satellite data

Author: Olivia Linke
1st Supervisor: Prof. Dr. Johannes Quaas
2nd Supervisor: Jun.-Prof. Dr. Heike Kalesse

March 31, 2020
Abstract

Aerosol-cloud interactions (ACIs) are a major uncertainty in estimating anthropogenic radiative forcing on the climate system. Adjustments of cloud properties to an aerosol perturbation concern among others the cloud fraction (CF), and have been proven as particularly complex.

European aerosol inventories have shown major reductions in the emission of anthropogenic particle species, such as sulphate and carbonaceous aerosols, since the late 1980s. Previous studies suggest that there was an overall decrease in cloud albedo as result of the aerosol emission decline. Changing cloud radiative properties due to adjustments have been part of observational and model-based studies, but remain elusive given the lack of observational constraints.

Cloud adjustments may generate important responses on the distribution of cloud horizontal sizes, that might shift due to forcings and feedbacks embedded in the adjustments, thereby providing an observational parameter.

This work mainly serves the development of a method for deriving cloud size distributions from satellite data to ultimately carry out long-term trends in CF over Europe during 1985–2018, where major aerosol reductions have been taken place. Landsat data with high spatial resolution of 30 m was pre-processed by the web-based platform Google Earth Engine to address the obstacle of high computational effort and time to handle the comprehensive data archive.

The method successfully derived multidecadal trends in CF, indicating a large-scale increase during 1985–2018, with local statistical significance. Overall, there was a decrease in the number of small clouds (several 10–100 m cloud length), whereas larger clouds (1 km and more) increased regarding their frequency of occurrence. The observations do not provide clear evidence of a large-scale impact from the aerosol retreat over Europe and are likely the result of other confounding factors (e.g., the local meteorology) that obscure potential ACIs.

Untangling the actual aerosol response, and the complex effects of those confounding factors can further expose the pure aerosol effect on clouds, but is left as an outlook for subsequent studies.
# Contents

1 Introduction ................................................................. 2
2 Aerosol-cloud interactions ............................................. 3
   2.1 Radiative forcing from aerosol-cloud interactions .......... 3
   2.2 Adjustments of cloud properties .............................. 3
3 State of knowledge ......................................................... 6
   3.1 Aerosol emission trends ........................................ 6
   3.2 Changes in cloud statistics .................................... 8
4 Landsat and Google Earth Engine (GEE) ............................ 15
5 Method ........................................................................ 17
   5.1 GEE pre-processing ............................................. 17
   5.2 Cloud sizes ...................................................... 19
   5.3 Cloud size distribution ....................................... 20
   5.4 Cloud fraction .................................................. 20
   5.5 Summary .......................................................... 22
6 Results ...................................................................... 23
   6.1 Binary cloud mask .............................................. 23
   6.2 Cloud sizes ...................................................... 25
   6.3 Cloud size distribution ....................................... 25
   6.4 Cloud fraction .................................................. 30
7 Discussion ................................................................. 35
   7.1 Trend analysis .................................................... 36
8 Conclusion and Outlook ................................................ 40

Technical supplement ............................................................. 42
   I Access to the GEE data catalogue ............................. 42
   II Landsat data structure on GEE ................................. 42
   III Landsat scene processing methods ............................ 42
   IV Landsat band characteristics ................................ 44
   V Three cloud masking algorithms ............................... 46
References ................................................................. 57
List of Figures ............................................................. 59
1 Introduction

Clouds are important regulators of the Earth’s energy balance due to their strong impact on fluxes of incoming shortwave radiation (SWR), and outgoing longwave radiation (LWR) that is emitted by the Earth (Cubasch et al., 2013). Since cloud interactions with both SWR and LWR fluxes are large, small changes in cloud properties may have important implications on the climate system (Boucher et al., 2013).

Several human impacts have the potential to alter cloud characteristics and abundance, among these, a variable atmospheric aerosol burden arising from changes in anthropogenic emissions (Lohmann and Feichter, 2005; Boucher et al., 2013). Aerosols alter the Earth’s energy budget through direct interactions with radiation, but also by serving as nuclei for cloud particles, thereby impacting cloud radiative properties (Lohmann and Feichter, 2005). Aerosol-cloud interactions (ACIs) have been part of the estimations of the latest report of the Intergovernmental Panel on Climate Change (AR5; IPCC, 2014) in terms of their top-of-atmosphere (TOA) radiative forcing (RF) during the period 1750–2011, but they are still poorly understood. In fact, AR5 identifies ACIs as the largest contribution to the total uncertainty range in quantifying anthropogenic perturbation.

In recent years, substantial improvements were made in the understanding of ACIs and the underlying processes. It is commonly known that changes in the amount of aerosols can impact the number concentration of cloud droplets, thereby affecting the cloud albedo (Twomey, 1974). However, large uncertainties remain, especially in estimating cloud adjustments to the instantaneous radiative forcing due to ACIs. Those adjustments concern the aerosol effect on cloud properties like the cloud lifetime, liquid water path (LWP), and the cloud fraction (CF), that can strongly impact the effective cloud forcing (Boucher et al., 2013).

During recent decades, the European aerosol burden has experienced substantial decreases, mainly due to air quality legislations and technical improvements of emission factories concerning the industrial sector, traffic, and power generation (Crippa et al., 2016).

This work aims to provide an observational linkage between decreasing aerosol concentrations and a potential CF response, by estimating cloud size distributions from satellite data. High-resolution Landsat data with a spatial resolution of 30 m was applied, and pre-processed by the web-platform Google Earth Engine (GEE) to manage the vast amount of data.

Within the next section, elaborations are given on known hypothetical pathways through which changing aerosol concentrations impact cloud characteristics. Section 3 outlines observed changes in aerosol and cloud statistics over Europe during recent decades, and introduces the idea of the cloud size distribution to provide an observational constraint. Within section 4 information are given about the Landsat satellite mission and its potential to carry out long-term cloud records by using GEE. The methodical part outlines all steps from the original Landsat data products to long-term records of CF, including the pre-processing via GEE. The technical details that concern GEE’s data collection and tools are separately given within the supplement. Results are presented successively, by analysing each intermediate product, starting from the binary cloud mask. Finally, results are discussed in terms of potential ACIs over Europe 1985–2018, before summarizing the main outcome, and giving an outlook.
2 Aerosol-cloud interactions

Changes in atmospheric aerosol concentrations have the potential to impact the Earth’s energy balance via altering radiative fluxes both in a direct and indirect manner. Aerosol particles scatter and absorb radiation at solar and terrestrial wavelengths (Lohmann and Fiechter, 2005). Additionally, light-absorbing aerosols emit thermal radiation, which heats the aerosol layer and stabilizes the air column below. The subsequent suppression of cloud formation, together with enhanced evaporation rates due to absorption heating likely results in a reduction of cloudiness (Boucher et al., 2013). These impacts on clouds are commonly summarized as semi-direct effects.

2.1 Radiative forcing from aerosol-cloud interactions

Aerosol particles are part of almost all liquid droplet nucleation processes as they provide cloud condensation nuclei (CCN). Consequently, a changing aerosol concentration has the potential to transform cloud characteristics of both warm clouds and ice/mixed-phase clouds, initiated from the liquid phase (Lohmann and Fiechter, 2005). The addition of aerosol particles likely increases the cloud droplet number concentration (CDNC), which in turn enhances the cloud albedo. The instantaneous negative radiative forcing due to aerosol-cloud interactions (RF_{ACI}) is commonly referred to as “Twomey effect” (Twomey, 1974, 1977). Moreover, an increase in the CDNC potentially alters cloud radiative properties, which can affect the albedo of a given scene containing liquid water clouds. Those adjustments concern the aerosol impact on CF, LWP, cloud lifetime, or precipitation (Albrecht, 1989). However, there are conflicting results on the sign of the radiative forcing arising from the adjustments, making it particularly challenging to understand the full implication of ACIs (e.g., Jiang et al., 2006; Xue et al., 2008; Small et al., 2009).

2.2 Adjustments of cloud properties

Precipitation suppression

It is likely that an increased state of pollution corresponds to more numerous droplets, which would decrease the mean droplet size at initially unchanged LWP (Twomey, 1974). One hypothetical pathway through which clouds adjust to smaller droplets suggests the suppression of precipitation in warm boundary layer clouds, which can increase LWP and cloud cover (Albrecht, 1989). This thought chain is widely known as “lifetime effect” as it potentially extends the residence time of cloud condensate receiving the aerosol perturbation, thereby adding to a negative RF (Lohmann and Fiechter, 2005).

However, previous studies reveal responses that are incongruent with the monotonic lifetime effect; LWP and CF may increase or decrease with higher aerosol concentrations, which complicates conclusions about the lifetime response (Small et al., 2009). Xue et al. (2008) ran large eddy simulations of trade cumulus and confirm a monotonic precipitation response due to an aerosol perturbation, but show both positive and negative responses in the CF. Figure 1 implies that higher aerosol concentrations tend to increase CF as long as precipitation is suppressed within low-pollution conditions. However, as the particle number concentration further increases, the CF tends to decrease, which violates the lifetime effect.
Results of both model-based and observational studies reinforce that a non-linear relationship exists within the cloud adjustments. Other processes can feed back on cloud radiative properties as a consequence of initially increased CDNCs (e.g., Grmpeert et al., 2018).

**Evaporation-entrainment feedbacks**

The idea of evaporation-entrainment feedbacks is schematically expressed in Figure 2 (Small et al., 2009). It suggests the modification of the basic circulation features of small cumulus clouds, including the core of updrafts, a shell of downdrafts at the cloud edges, and cloud-top turbulent entrainment.

Figure 2: Schematic illustration of the evaporation-entrainment feedback given by Small et al. (2009) for small cumulus clouds. Basic circulation features of either (A) clean, and (B) polluted clouds (top and bottom, respectively), include a positively buoyant core of updrafts, negatively buoyant shells of downdrafts at the cloud edges, and cloud-top turbulent entrainment. A measure of buoyancy, $\Theta_v$, is represented by dashed lines.
Smaller droplets in more polluted clouds evaporate at shorter time scales, provoking higher evaporative cooling around the cloud periphery. An intensification of the horizontal buoyancy gradient emerges between the positively buoyant updraft of an actively growing cumulus, and the negatively buoyant shell of downdrafts. The production of additional vorticity enlarges the cloud-top entrainment of dryer air masses and enhances local evaporation rates, thereby completing a positive feedback mechanism. The reduction in LWP and CF is a likely response that is incongruent with the tendency of pollution to increase cloudiness via precipitation suppression (Jiang et al., 2006; Small et al., 2009).

**Sedimentation-entrainment feedbacks**

It has been further suggested that sedimentation-entrainment feedbacks provoke an additional reduction of cloudiness as a consequence of increased aerosol concentrations (Ackerman et al., 2004; Bretherton et al., 2007). The cloud droplet sedimentation rate is inversely related to the CDNC due to reduced fall speeds of smaller droplets. Bretherton et al. (2007) argue that an increasing entrainment rate is the consequence of removing less water from the cloud-top entrainment zones via droplet sedimentation. That enhances both entrainment-induced evaporative cooling, and LWR cooling at the cloud top, thereby decreasing the liquid water content of a given cumulus.

The possibility of reduced cloudiness due to entrainment feedbacks has the potential to offset a significant fraction of the RF_{ACI}/Twomey effect, but the strength of the feedbacks likely depends on atmospheric soundings such as ambient relative humidity (e.g., Ackerman et al., 2004). Graspeerd et al. (2018) suggest that LWP adjustments to an increased CDNC could offset up to 60% of the Twomey effect as an upper bound of the positive RF expected from LWP reductions. The existence of multiple effects that adjust cloud radiative properties to an aerosol perturbation shows that ACIs act within a highly interactive system ranging from the microscale to cloud-process scales for cloud-top turbulent entrainment and updrafts (Mülenstäd and Feingold, 2018).

In addition, cloud adjustments might feed back on the CDNC, obscuring the causal impact of the initially changing cloud. For instance, higher CDNCs likely suppress precipitation, which reduces the wet scavenging of aerosols from the cloud layer. In turn, more particles are available to serve as CCN, thereby closing a positive feedback loop (Grysspeerdt et al., 2018).

Consequently, not only the quantification of existing hypotheses appears as a challenge, an incomplete understanding of cause from effect further contributes to an overall uncertainty in estimating ACIs (Mülenstäd and Feingold, 2018).
3 State of knowledge

3.1 Aerosol emission trends

Changes in the concentration of anthropogenic aerosol species, such as sulphate and carbonaceous particles, have the potential to substantially alter radiative fluxes (Lohmann and Feichter, 2005).

Sulphate as a secondary particle is mainly an oxidation product of sulphur gases which largely originate from anthropogenic fossil fuel and biomass burning. Sulphate particles are very hygroscopic and are readily activated as CCN which tends to make clouds more reflective. The effect on clouds together with the large scattering ability of the particles in the solar spectrum results in an overall negative RF due to increased sulphate aerosol concentrations (Smith et al., 2011; Boucher et al., 2013). Black Carbon (BC) on the other hand, is likely to have a warming impact since it efficiently absorbs solar radiation, and poorly contributes to CCN spectra, thereby mainly acting through semi-direct effects on clouds (Stjern and Kristjansson, 2015). BC particles are primary products of the combustion of fossil fuel, biofuel, and biomass (Forster et al., 2007; Boucher et al., 2013).

Streets et al. (2006) show simulation-based emission trends of sulphur dioxide (SO₂) and BC for different industrialized areas during two decades preceding the year 2000 (Figure 3). Attention should be payed to the curves representing the former Soviet Union, eastern Europe, and the area of the European Organization for Economic Co-operation and Development (OECD), which accounts for changes in central and western Europe.

![Figure 3: Results of simulations run by Streets et al. (2006) showing trends in both SO₂ (left), and BC emissions (right), during the period 1980–2000 for different industrialized regions, including European areas.](image)

Emissions of both SO₂ and BC largely decreased at the end of the 1980s as consequence of the break-up of the Soviet Union, and the subsequently dropping industrial production within the eastern sector. During the same period, in western and central Europe (OECD) emissions decreased, especially in the SO₂ inventory, which is likely the result of environmental regulations and European “clean-air” policies (Streets et al., 2006; Norris and Wild, 2007).
Crippa et al. (2016) investigate the improvements in European air quality onward from 1970 by addressing the spatial distribution of pollution trends in Figure 4.

![Figure 4: Simulations of Crippa et al. (2016) showing hotspots of avoided emissions of anthropogenic gaseous (SO₂, NOₓ, CO) and particulate (PM₁₀, BC) air pollutants due to air pollution policies, and improved technology in Europe.](image)

European hotspots of avoided emission show the reduction in selected gaseous pollutants (SO₂, NOₓ, CO), and particulate matter (PM₁₀, BC), respectively.

The SO₂ emission decline is mainly attributed to a combined effect of improved energy generation and the switchover to cleaner fuels. Larger areas of high emission decline are located mainly in western European urban areas around Paris, London, Berlin, the Benelux region, while some point sources (power plants and industrial regions) are spread all over Europe.

Looking at the reduction in PM₁₀ and BC emissions, large areas of decline are found in the western European countries, with particularly high decrease in the industrialized regions, e.g., the Benelux. Crippa et al. (2016) relate the emission decline mainly to the EURO standards on road traffic and the implementation of particulate filters on power plants and industries.

In summary, there is a distinct reduction in the European aerosol burden since the late 1980s due the break down of the industrial sector in the former Soviet union, and stricter environmental regulations in western Europe. Those regulations address air quality legislation and technical improvement of emission factories within the industrial sector, traffic system, and power generation. The strong decrease in the aerosol concentration over Europe has potentially contributed to a large-scale radiative forcing mediated by ACIs.
3.2 Changes in cloud statistics

Cloud albedo

The Twomey effect is a widely known theoretical concept that states the modification of cloud brightness from an aerosol perturbation, by impacting number and size of liquid cloud droplets (Twomey, 1974, 1977). But are there observational evidences for cloud-albedo responses as a result of the multidecadal aerosol decline over Europe? The trend in surface solar radiation over Europe shows a widespread decrease during three decades preceding 1990, followed by a recovery of the downward solar flux (e.g., Pinker et al., 2005; Wild et al., 2005). The sequence of decrease and increase is commonly referred to as “solar dimming” and “solar brightening”, respectively. As mentioned earlier, sulphate and BC particles behave differently in terms of their direct interaction with solar radiation, but they reduce either way the solar flux reaching the surface.

Two studies are now introduced in depth, namely Norris and Wild (2007), and Krüger and Graßl (2002) to address the instantaneous RF over Europe that is potentially mediated by ACIs during recent decades.

Norris and Wild (2007) attribute the observed brightening event onward from 1990 to the direct radiative effect imposed by simultaneously decreasing aerosol concentrations. But did ACIs further contribute to the downward solar flux trends at the surface? Figure 5 illustrates two downward solar flux anomalies during 1965–2004 as pan-European time-series of global radiation (top panel), and shortwave cloud cover radiative effect (CCRE; middle panel). The latter addresses all changes in the surface downward solar flux that are linearly related to changes in cloud cover. The data originates from both surface and satellite-based observations, and a radiative transfer model (color coding; see Figure description).

![Figure 5: Time series of (low-frequency) monthly pan-European flux anomalies in the global radiation (top panel), the shortwave CCRE (middle), and the residual after removing the CCRE from the global radiation flux (bottom). The color coding accounts for the respective data source, with surface-based observations in black, fluxes estimated from synoptic cloud-observations in blue, and simulated radiative fluxes based upon satellite-derived cloud properties in red.](image-url)
The plot indicates a high correspondence between global radiation, and the shortwave CCRE anomalies. Consequently, the shortwave CCRE represents the greatest contribution to surface downward flux variations.

As mentioned before, the shortwave CCRE time series concerns that part of cloud albedo anomaly that is linearly related to changes in cloud cover. On the other hand, the global radiation anomaly includes all impacts of varying cloud radiative properties. After removing the shortwave CCRE from the global radiation, a residual time series emerges (bottom panel), with statistically significant trends, matching the periods of solar dimming and brightening.

Onward from the year 1987, there is a clear increase in the residual which is in line with the large-scale aerosol decline. Moreover, the global radiation indicates an overall positive trend onward from 1987, but there is a decrease in the shortwave CCRE time series. Norris and Wild (2007) conclude, that an increasing cloud cover (i.e., decreasing shortwave CCRE) has likely offset parts of the global radiation increase, so that cloud cover variations do not contribute to the solar brightening trend. Moreover, shortwave CCRE variations are highly correlated with the North Atlantic Oscillation (NAO), so that it is likely an effect of weather and climate variability rather than being related to the aerosol emission decline.

Theoretically, the residual time series is composed of variations in the clear-sky solar flux and cloud albedo changes not linearly related to cloud cover changes. Norris and Wild (2007) separate both contributions due to their possible correspondence to direct aerosol-radiation interactions and ACLs, respectively.

Both flux contributions show an increase onward from 1987. Despite lacking statistical significance, they ascribe the increase in the clear-sky flux anomaly to the previously articulated direct aerosol radiative effect, and the shortwave cloud effect residual to the aerosol influence on the cloud albedo. Recalling the Twomey effect, aerosol number and cloud albedo are connected and in case of a decreasing aerosol burden, potentially contributed to the solar brightening.

Krüger and Graßl (2002) find an average cloud albedo decline by comparing the two periods 1985–1989 and 1996–1999 over central Europe using satellite data. In agreement with the study of Norris and Wild (2007), they attribute the trend to a large-scale weakening of the Twomey effect. Moreover, the decreasing cloud albedo appears more intense within industrial areas, indicating an anthropogenic influence on the cloud albedo that is stronger within polluted regions, where the emission decline carries more weight.
Krüger and Graßl (2002) further compare data between remote sites, and adjacent regions of known emission peaks for a given time period. Figure 6 illustrates the mean cloud reflectance for winter (left), and summer time (right) during the period 1985–1989. The summer data shows a gradient of reflection from remote to polluted regions, with maximum mean reflectance in regions of the so-called “Black Triangle” (BT; border-triangle shared by Poland, the Czech Republic, and Germany), where numerous coal plants contributed to extensive sulphur emission peaks during the mid-1980s. During winter, the gradient of reflection changes sign and main emission centers indicate local minima in reflectance (e.g., around London, Paris, Leipzig, and Prague). Krüger and Graßl (2002) argue that cloud reflectivity trends to not entirely relate to the Twomey effect, but also to the absorbing effect of carbonaceous particles in extreme-pollution conditions. The latter seems to dominate the albedo response in winter time due to lower boundary heights and the slow formation of secondary sulphate particles from precursor gases. However, Norris and Wild (2007) argue against an important long-term contribution of semi-direct effects over Europe during the past decades.

In summary, both introduced studies on the effect of aerosol on cloud brightness state an overall decrease in cloud albedo onward from the late 1980s, which likely mediated a positive radiative forcing associated with the Twomey effect. However, there has been no consideration of the broad range of possible cloud adjustments as contribution to a changing cloud statistic. Norris and Wild (2007) recognize an overall increase in the cloud cover due to the decline in the shortwave CCRE time series. They attribute this variation entirely to synoptic and climate variations, but there might be a connection to the European aerosol decline.
Adjustments of cloud properties

Norris and Wild (2007) state that since a long-term European increase in cloud cover occurred at the same time as aerosol loadings decreased, substantial impacts of cloud adjustments are unlikely. But they only consider a possible response in terms of the traditional lifetime effect, were the cloud cover increases with aerosol addition, and vice versa for the European aerosol decrease. However, as discussed in section 2.2, microphysical feedbacks might cause an effect on the cloud cover that is incongruent with a response exclusively mediated by precipitation suppression.

Recalling the modelling study of Xue et al. (2008), it has been concluded that the CF of shallow cumulus increases as precipitation is inhibited, but within higher ranges of aerosol pollution, the response might have the opposite sign. Two major regimes have been proposed to categorize the response; first, a precipitation regime, showing an increase in CF corresponding to the lifetime effect, and second, a non-precipitating regime where an increase in aerosol concentration reduces CF due to the droplet-size impact on entrainment rates.

Xue and Feingold (2006) speculate on the role of cloud horizontal sizes and propose that small and large clouds likely show variable responses to changing aerosol concentrations. The addition of aerosols may increase the size of large clouds as they are less susceptible to entrainment drying. They refer to Kaufman et al. (2005), who used data from the near-global moderate resolution imaging spectroradiometer (MODIS), and observe an increase in CF with higher aerosol loadings, consistent with the lifetime effect. But MODIS data is comparatively coarse and cannot capture the smallest clouds (Jiang et al., 2006). Modelling studies show opposing aerosol responses for small cumulus with decreasing CF and simultaneously increasing aerosol concentrations (Jiang and Feingold, 2006; Xue and Feingold, 2006). They argue that smaller clouds are more exposed to entrainment due to their higher surface-to-volume ratio, which feeds back positively on the cloud size.

Small et al. (2009) support that entrainment feedbacks are unlikely to affect heavy precipitation regimes, but when small, non-precipitating cumulus clouds experience an aerosol perturbation, they might respond in a manner incongruent with the lifetime effect.

However, generalizations of the aerosol response in these small clouds is premature due to the dependency on controlling factors, such as the ambient relative humidity, which makes the overall response highly uncertain (Ackerman et al., 2004).

Deriving a relationship between aerosol amount and cloud properties has been part of many studies in the past relying on models and observations, however, scientists have not reached a consensus on the overall response and the resulting climate effect.

For the LWP, Grysspeerdt et al. (2018) show that the relationship with the CDNC can be either positive or negative, depending on the degree of pollution; the LWP decreases with enhanced particle concentration in lower pollution regimes, falling together with regions where precipitation is expected. For higher states of pollution, the LWP-CDNC relationship is negative, which is evident for areas where precipitation is rather unlikely. Those results show similarities with the two-regime theory proposed by the modelling study of Xue et al. (2008), who showed the switch from positive to negative susceptibility by passing from precipitation regime to non-precipitation regime. Overall, Grysspeerdt et al. (2018) conclude that LWP reduction dominates the aerosol response on a global average.
For the CF, Gruseck et al. (2016) show an overall positive correlation between aerosol amount and CF. However, difficulties remain in distinguishing the aerosol effect on cloud properties from confounding factors, e.g., local meteorology, or aerosol hygroscopic growth. Moreover, satellite retrieval errors potentially generate a relationship between aerosol number and the cloud response, that can be falsely interpreted as correlation. Regarding the role of atmospheric sounding, Rosenfeld et al. (2019) further isolate the effect of aerosols from the meteorological impact on cloudiness. They found that by extracting the pure aerosol effect, cloudiness rises monotonically with increasing particle concentration, to an extent much larger than previously estimated (Figure 7). Rosenfeld et al. (2019) attribute lower or even negative relationships from previous studies to meteorological controls on cloudiness that obscure the actual aerosol effect.

![Figure 7:](image)

Figure 7: Previous model and observational studies investigated the effect of increased aerosol amount on cloudiness during the past two decades. Quantifications are presented in terms of cloud susceptibility (variation of the cloudiness due to changing aerosol amount). The most recent study by Rosenfeld et al. (2019) concludes a higher (positive) susceptibility than studies before by extracting the pure aerosol effects on clouds from the meteorological impact. The plot was adopted from Sato and Suzuki (2019).

In summary, a combination of multiple processes embedded in the cloud adjustments likely determines the overall aerosol effect on cloud parameters, which likely depends on cloud type and the state of atmospheric sounding (Christensen et al., 2016). Moreover, for individual clouds, diverging effects on the cloud size are possible for clean and polluted conditions, depending on which mechanism dominates. This reinforces the hypothesis that through addition of aerosols the cloud size distribution might shift either way due to lifetime and/or entrainment feedbacks. This potentially
provides an observational constraint for aerosol effects on cloud fraction and lifetime, which are often assumed to be correlated (Jiang et al., 2006). High-resolution satellite imaging provides a convenient method to derive cloud horizontal size distributions, since surface-based/airborne cloud products are spatially limited and only provide cloud distributions over a restricted size range. The next section introduces the distribution of cloud horizontal sizes as parameter to derive the change in cloud fraction over Europe 1985–2018.

The role of cloud size distribution

A common way to categorize clouds based upon their horizontal dimension is via the cloud chord length, which can be derived from remote sensing techniques and aircraft observations (Kleiss et al., 2018). The cloud distribution $n$ is commonly known to follow a negative sloping power-law relation in terms of the cloud length $L$ (or area)

$$n(L) = \alpha L^{-\beta},$$

where $\alpha$ and $\beta$ are the intercept and the slope on a log-log plot describing the relation (Feingold et al., 2017). The power-law distribution was previously documented for satellite observations and large eddy simulations, and also applies for global data sets, as demonstrated by Wool and Field (2011). They applied MODIS satellite data, and derived size distributions by adding up the total number of clouds $N_i$ with lengths between (approximately logarithmic) bin boundaries $L_i^-$ and $L_i^+$, and dividing by the respective bin width and total transect length $D_{tot}$;

$$n(L) = \frac{N_i}{D_{tot}(L_i^+ - L_i^-)}.$$  

Figure 8 displays the size distribution for two years of MODIS data (2008 and 2009) in a log-log spaced plot, together with aircraft-based observations and simulations from a global high-resolution numerical weather prediction (NWP) model.

![Figure 8: Cloud chord length $n(L)$ derived from MODIS satellite (filled circles), aircraft (open triangles), and UM (diamonds) data. The middle sloping line represents the linear fit regression to the data. The Figure is adopted from the study of Wood and Field (2011).](image-url)
Wood and Field (2011) indicate that $n(L)$ follows a negative power law distribution, spanning over 4 orders of magnitude with horizontal scales between 0.1 and 1000 km. Therefore, the connection between cloud sizes and their frequency of occurrence simply depends on the slope $\beta$, and the intercept $\alpha$ of the corresponding log-log linear plot. Steep negative slopes, together with large intercepts indicate a relative abundance of small clouds, whereas the combination of shallow negative slopes and small intercepts suggests a relative abundance of large clouds (Feingold et al., 2017).

For larger cloud fields, a scale break, $L_s$, arises where the distribution deviates from the power law, which can be described by the relation $L^{-\beta} \exp\left(-\frac{L}{L_s}\right)^2$.

The size distribution can be integrated to determine the contribution of clouds with a certain horizontal extent to the cloud fractional cover $f_c$ (or cloud fraction; CF). The cloud fraction is thereby defined as

$$f_c(L_{\text{min}}, L_{\text{max}}) = \int_{L_{\text{min}}}^{L_{\text{max}}} Ln(L)dL,$$

where the lower integration boundary is related to the minimum detectable cloud length.

Wood and Field (2011) show that clouds with logarithmic intervals of size contribute equally to the cloud cover for the case that $\beta$ equals 2. If $\beta$ rises beyond this threshold, small clouds increasingly contribute to the cloud cover, whereas for exponents below 2, large cloud fields become the main contribution (results from integrating $Ln(L)$ for $L$, with $n(L) = \alpha L^{-\beta}$; see Wood and Field, 2011, for more details).

The global size distribution indicates a power-law exponent with $\beta = 1.66$, matching the widespread conclusion that large cloud fields dominate the cloud cover, whereas small clouds are globally more numerous than large clouds.

Due to the coarse resolution of satellite instruments like MODIS, the majority of clouds cannot be captured within the satellite scene, which is why higher-resolution data is needed to fill the demand.

Landsat provides a powerful tool to resolve cloud fields at 30 m horizontal size, thereby allowing the presentation of small cumulus clouds within the data set. The disadvantage is the reduced temporal resolution, given the fact that Landsat captures each point of the Earth every 16 days. This causes problems within diurnal cloud studies, but the data is still suitable for deriving interannual trends of cloudiness (Kleiss et al., 2018).

The next section introduces the main characteristics of the Landsat mission and its potential to provide observational data for large-scale cloud-statistical analyses.
4 Landsat and Google Earth Engine (GEE)

The Landsat mission is a joint project of the U.S. Geological Survey (USGS) and the National Aeronautics and Space Administration (NASA), and provides imagery onward from 1972. Up to now, the data archive includes a continuous imaging of the Earth’s surface from seven successfully launched satellites. With over 40 years of consistent data acquisition, Landsat forms the longest record of satellite-based observations at visible and thermal wavelength (L. Williams et al., 2006; Loveland and Dwyer, 2012; Roy et al., 2014). Where costs and access to Landsat imagery limited the data use before, recent open-access policies of the Landsat data archive enable an improved understanding of the Earth’s environmental systems and the way it is changing (Woodcock et al., 2008).

The following information summarizes the main characteristics of Landsat’s satellite acquisitions and the primary sensors aboard, referring to the details given in the Landsat fact sheet of the U.S. Geological Survey (2016).

The first satellite of the Landsat mission was launched in 1972 as Earth Resources Technology Satellite (ERTS-1) in a sun synchronous orbit at 920 km altitude, giving global coverage every 18 days. ERTS-1 was later renamed Landsat 1 and followed by the partner satellites Landsat 2 and Landsat 3 in 1975, and 1978, respectively. The primary sensor aboard Landsat 1 to 3 was the Multispectral Scanner (MSS), a four spectral-band instrument, acquiring data in the visible green to the near-infrared (NIR) wavelengths at a 79-meter spatial resolution. Landsat 4 and 5 satellite launches followed in 1982, and 1984 respectively, with Landsat 5 covering a total record length of 28 years and 10 months, thereby setting the record for longest-operating Earth observation satellite. Both Landsat 4 and 5 carried the MSS, along with the new Thematic Mapper (TM) sensor, including three additional spectral bands, two in the shortwave infrared (SWIR) part of the spectrum, and one thermal infrared (TIR) band. The spatial resolution within the TM sensor improved to 30 m for the visible through middle infrared channels, and thermal data was collected at 120-meters resolution. The sun synchronous orbits of Landsat 4 and 5 where chosen differently from the predecessors at altitudes of 705 km with a 16-day full-Earth coverage cycle. The successor satellites where Landsat 7, and 8 in 1999, and 2013, respectively, continuing the mission since Landsat 6 failed to achieve orbit in 1993. Both Landsat 7 and 8 satellites are still contributing to the data collection on orbits similar to those of Landsat 4 and 5, and offset to each other to allow 8-day repeat coverages of each scene. Landsat 7 carries the Enhanced Thematic Mapper Plus (ETM+), with 30-meter visible, NIR, and SWIR bands, a 60-meter thermal band, and a new panchromatic band. The latter covers the complete visible spectrum and collects a larger amount of radiation compared to the other bands, allowing “pan-sharpening” of multispectral imagery to a higher resolution of 15 m. Note that during the year 2003 Landsat 7 suffered a Scan Line Corrector (SLC) failure, which left wedge-shaped spaces of missing data on either side of Landsat 7’s images and limits the data use for certain applications.

Landsat 8 is the most recent satellite that carries a push-broom Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). OLI covers the visible, near-IR, and parts of the SWIR spectrum at 30-meters resolution and again, at 15-meter resolution in the panchromatic band. OLI additionally includes a deep blue band for coastal-aerosol studies, and a band for cloud detection. The latter is placed within spectral ranges where the atmosphere absorbs nearly all radiation to allow the extraction of the highly reflecting
clouds. TIRS contains two thermal bands, which are acquired at 100 m, respectively. Note that all thermal data collected within Landsat 4–8 was originally acquired at lower resolutions compared to the shorter wavelengths, but is later resampled to 30 m. The Landsat band structure is summarized in Table 3 within section IV of the supplement, and further discussed in the context of different Landsat processing methods.

From now on, Landsat data is referred to in the general form “LXSS”, following the official USGS name convention with

\[
\begin{align*}
L, & \quad \text{the Landsat constant,} \\
X, & \quad \text{the Landsat sensor, with } C = \text{OLI} / \text{TIRS}, O = \text{OLI-only}, \\
T = \text{TIRS-only}, E = \text{ETM+}, T = \text{TM}, \text{and } M = \text{MSS}, \text{ and} \\
SS, & \quad \text{the satellite, with } 05 = \text{Landsat 5}, 07 = \text{Landsat 7}, \text{ and } 08 = \text{Landsat 8},
\end{align*}
\]
as defined in Zanter (2017).

Many advantages are induced by opening the petabyte-scale Landsat archives for public use. However, taking full advantage of these information requires a high technical expertise and effort in order to acquire and store the massive data amount (Gorelick et al., 2017). Google Earth Engine (GEE) was developed to address these problems by providing access to high-performance computing resources to process large geospatial data volumes. GEE is a cloud-based data analysis platform that includes, among other archives (e.g., MODIS and Sentinel1), the entire time series of Landsat images (Hansen et al., 2013; Gorelick et al., 2017).

Landsat’s continuous observation of the Earth’s surface was developed primarily to track land use and land cover changes (Roy et al., 2014; U.S. Geological Survey, 2016). With the help of Google’s massive computational capabilities, recent studies carried out products to address long-term changes of e.g., regional and large-scale forest cover (Hansen et al., 2013), global surface water (Pekel et al., 2016), and urban land cover within high-resolution mapping (Liu et al., 2018). For those observations, the vulnerability of optical imagery to cloud cover is often treated as problem embedded in remote sensing. To address this matter, the Landsat community developed several methods for masking out clouds from the imagery, leading to a large selection of different algorithms available within the GEE environment. Here, this circumstance is used as an advantage to carry out long-term cloud records over Europe.
5 Method

This section presents the methodical proceedings of this work. The first part concerns all steps run via the GEE computational facilities and follows three major tasks; first, selecting suitable data from the Landsat data catalogue, second, filtering among the images, and third, the cloud-masking process, which separates clear-sky and cloudy pixels to produce a binary product.

The technical supplement includes information about the GEE platform that concern those tasks. How to get access to data and tools within GEE in the first place is briefly described in supplement section I.

5.1 GEE pre-processing

Data selection and filtering

The center of this study was to develop a method for deriving cloud horizontal size distribution and cloud fraction trends over Europe during 1985–2018, from high-resolution Landsat imagery. For this purpose, the data was taken from the USGS Collection-1 Tier-1 TOA reflectance archive.

Landsat images are provided as either Tier-1 (T1), Tier-2 (T2), and Real-Time (RT) data, where T1 classification meets the highest radiometric and geometric quality requirements. More information are given within section II of the supplement, and are based upon the USGS documentation on Landsat Collection 1 Tiers.

GEE provides Landsat data from different processing methods, namely the conversion of raw images to the TOA reflectance, and Surface Reflectance (SR), respectively. The TOA-reflectance calibration considers the variation of solar elevation on a daily basis, and the seasonally varying Earth-Sun distance, which makes it the better choice for multitemporal studies (details given in supplement section III).

After defining and accessing the main data set, the GEE platform further allowed to filter the entire collection for certain requirements. Landsat data is provided as multispectral GeoTIFF images, containing metadata that are stored as set of properties. By accessing the metadata, each collection was filtered for acquisition date during 1985–2018, and location.

From Landsat 5 and 8, all available images were considered within the period of interest, with data acquisition periods of 1985–2011, and 2013–2018, respectively. Landsat 7 data was included only within 1999–2003 due to the SCF which would have affected the results onward from May 2003. The year 2012 is not represented at all due to the fail of Landsat 6 and the consequential data gap.

Landsat images are provided within so called footprints, or “tiles”, spatially classified by satellite path and row according to Worldwide Reference System-2 (WRS-2). Figure 9 gives a map of all tiles covering the research domain. In total, 143 Landsat tiles were chosen to represent a wide range of the European continent.
Binary cloud mask

In the next step, all files within the filtered data collection were passed on to the cloud-masking process to break down each multispectral image to a binary mask of ones (clouded) and zeros (not clouded). The mask was provided by a GEE internal algorithm, and treated as a function within the Python API (see section I and V within the supplement).

As mentioned before, Landsat 5–8 provide remote sensing over a wide range of radiation bands reaching from the visible to thermal part of the spectrum. Landsat additionally provides bands for quality assessment (QA) to further identify pixels based upon instrumental, atmospheric, and surficial conditions (supplement section IV).

Those QA bands typically contain bit-packed information about the status of cloudiness of a given pixel, which is expressed as either true, false, or as confidence level. Two bands, the "pixel_qa" (pixel quality assessment) and "bqa" (band for quality assessment), have been considered within this work to produce binary cloud masks. The two QA bands are contained within the SR, and TOA-reflectance collection on GEE, and are based upon the C Function of Mask (CFMask) algorithm, and the Automated Cloud-Cover Assessment (ACCA) algorithm, respectively.

The operational approaches where validated against a third cloud mask, the internal "SCS" (Simple Cloud Score) algorithm, that is provided by the GEE community. Details about the three pre-existing masking approaches are given in supplement section V and all references therein. Section V is concluded by an argumentation of which algorithm provided the best choice for this work.

To summarize the latter, the SCS code was selected over the operational algorithms to produce long-term trends of cloudiness over Europe. The main argument was the tendency of operational masking procedures to overmask cloudiness within the scene, even for the cloud-confidence bits. Automated masking algorithms are convenient for observing changes at the Earth surface where cloud cover is considered an obstacle. Therefore, it is preferable to overmask rather than having cloud contamination alter the trend as false
negatives (Foga et al., 2017; Mateo-García et al., 2018). The SCS provides a rather simple method for scoring Landsat pixels by their relative cloudiness, but allows controlling the trade-off between commission (overinclusion) and omission (underinclusion), by adjusting the lower cloud-scoring threshold. The algorithm considers certain spectral characteristics of clouds, namely brightness in the visible blue and all visible bands, brightness in all IR bands, relatively low cloud temperatures, and the Normalized-Difference Snow Index (NDSI) for the discrimination from snow.

The GEE pre-processing is started by sending a batch-export command to the engine. Only now, each image within the filtered collection is sequentially masked for clouds and then exported at original resolution of 30 m.

The main data set included around 52 500 Landsat GeoTIFF files, which gives an average of 367 images per satellite footprint. A typical Level-1 GeoTIFF product accounts for a data size of several 100 MB up to few GB, thereby requiring a large amount of time, storage, and computational effort to download and process the files at original size. With the help of GEE, the entire collection of multispectral images was digitized into binary cloud masks. The resulting images account for only few MB, given a reduction in size to 1–0.1 % of the original data volume.

The only disadvantage of the pre-processing via GEE is the design of the batch export; all tasks, one per image, are sequentially delivered to the engine. 3000 tasks are allowed in line for the processing at the most, allowing only step-wise ordering procedures. However, this problem could be addressed in the future by parallelization from different Google accounts.

Until now, each step was exclusively run via the Google engine, where no downloading or private computational resource was necessary. The remaining parts build on the pre-processing, but are realized without Google’s computational power.

5.2 Cloud sizes

For deriving cloud horizontal sizes from the binary mask, the cloud length in two different sampling directions (in North-South and West-East alignment) was used as a measure. Within in each image, cloud sizes are derived for both sampling directions, along each stripe of a given scene. Taking as an example Landsat images acquired for WRS-2 path 191, row 25; all images contain 7308 pixels in the North-South, and 7989 pixels in the West-East direction, respectively. That accounts for 7989 (7308) stripes in North-South (West-East) alignment, each of them with 7308 (7989) px, or approximately 220 (240) km length. The length of each stripe is hereinafter referred to as sample length $D$.

The algorithm for registering cloud lengths along each stripe simply counts horizontally contiguous cloudy pixels between the clear boundaries. If the first or last pixel of a given stripe contains a cloud, the corresponding cloud length is initially excluded due to the chance of extending beyond the satellite scene. Cloud sizes are stored in the unit pixel and are later weighted by the total number of pixels within the image.
5.3 Cloud size distribution

Corresponding to equation 2, the cloud size (length) distribution is computed by adding up the total number of clouds $N$ that showed horizontal lengths $L$ within logarithmically scaled size bins, and dividing by the respective bin width, and the total sample lengths, $D_{tot}$. For a given sampling direction, $D_{tot}$ is essentially the length of pixel stripes within that direction, multiplied by the number of those stripes, i.e., the total number of pixels $P_{tot}$ of a given image;

$$n(L) = \frac{N_i}{2 \cdot P_{tot} \cdot (L_{i+} - L_{i-})}. \quad (4)$$

Note that the pre-factor 2 within equation 4 comes from the consideration of two sampling directions that are considered within the computation of $n(L)$, which means that each image is completely screened for clouds two times.

For deriving an average size distribution including more than one satellite scene (e.g., when deriving annual values) $P_{tot}$ is simply derived by adding up the pixel number of all images that go into the calculation.

By ignoring all clouds touching the scene boundary, a size-dependent sampling bias arises since large clouds are more likely being excluded. Wood and Field (2011) address this error by introducing a correction term to equation 2. However, applying the correction to the Landsat data was set aside here, since the main interest was to quantify how cloud size distributions changed over the years. Since the image size among all Landsat files does not show distinct variations, the error can be considered an equal impact on the results and thereby cancels by looking at temporal changes.

As mentioned before, the cloud distribution typically follows a single power law, so that the relative abundance of either small or large clouds can be classified by the intercept and slope of the corresponding plot in log-log space (Feingold et al., 2017).

To estimate changes in the frequency of occurrence of cloud sizes, annual size distributions were derived by grouping together the data within each year during 1985–2018, before applying equation 4. To quantify interannual changes, equation 1 was fitted to use the horizontal exponent (or slope) $\beta$ and the way it changes for estimating the pivoting of the size distribution, i.e., the variation of frequency of occurrence among the cloud-size classes.

5.4 Cloud fraction

For a given cloud field that conforms to a power law size distribution, the cloud fraction can be derived by integrating $Ln(L)$ over $L$.

If $L_{min}$ is the smallest, and $L_{max}$ the largest detectable scale, the CF contribution from clouds within these size ranges is computed according to equation 3, where $n(L)dL$ is the number of clouds with horizontal lengths between $L$ and $L + dL$ per unit length (Wood and Field, 2011).

Finally, the CF is calculated as

$$CF = \int_{L_{min}}^{L_{max}} Ln(L)dL = \frac{\alpha}{2 - \beta} \left( L_{max}^{2-\beta} - L_{min}^{2-\beta} \right), \quad (5)$$
with power-law parameters $\alpha$ and $\beta$ derived from fitting the observations by $n(L) = \alpha L^{-\beta}$, and $\beta \neq 2$.

Theoretically, clouds can exist on all scales ranging from zero to infinity, or in practice, up to the planetary scale. However, each measuring device has its limitations and can only register clouds within a certain size range (Wood and Field, 2011). The lower integration boundary is typically related to the minimum detectable cloud length, i.e., 1 px, corresponding to 30 m for Landsat data. The upper integration boundary is set to the maximum cloud scale that follows the power-law distribution, i.e., a straight line on the corresponding log-log plot.
5.5 Summary

1) Landsat Data Catalogue

Definition of collection (USGS Landsat 5 to 8 Collection-1 Tier-1 TOA-reflectance), and filtering images for acquisition date and location, according to the research domain.

2) Landsat multispectral GeoTiff images

Cloud masking via GEE internal SCS algorithm, which considers
- brightness in the visible blue,
- brightness in all visible bands,
- brightness in all IR bands,
- low temperatures, and
- NDSI for discrimination from snow,

and adds a cloud score band to the collection, with adjustable cloud score threshold as lower limit.

3) Binary cloud masks

Stripe-wise extraction of cloud lengths for two sampling directions in North-South, and West-East alignment, by counting horizontally contiguous cloudy pixels.

4) Cloud horizontal lengths $L$

Grouping together all cloud sizes $N_i$ for each year within 1985–2018 (per grid box or for whole domain), and applying

$$n(L) = \frac{N_i}{2 \cdot P_{tot} \cdot (L_i+ - L_i-)}.$$ 

Fitting $n(L) = \alpha L^{-\beta}$ to the observational data set to estimate (annual) size distribution parameters, that are passed on to

$$CF = \frac{\alpha}{2 - \beta} \left( L_{max}^{2-\beta} - L_{min}^{2-\beta} \right),$$


5) Single cloud length distribution $n(L)$

Figure 10: Overview of the methodical steps to derive cloud trends over Europe 1985–2018 from the Landsat GEE data catalogue. In-between products are marked blue, where 1) – 4) represent data sets with original spatio-temporal assignments, i.e., each file within the data set is attributed to a certain Landsat scene of given time and location. Product 5) represents single annual cloud distributions produced by grouping together the data for each year within 1985–2018. From the last intermediate product, long-term records of CF are derived as pan-European trends, and grid-box trends per individual footprint center coordinates.
6 Results

6.1 Binary cloud mask

For masking multitemporal satellite images, the SCS code implemented in GEE was chosen over the operational algorithms provided within the quality assurance of Landsat. The cloud likelihood within SCS is presented on a scale of 0–100, where larger values indicate a greater chance of a pixel being clouded. The choice of the cloud likelihood threshold as lower limit was based upon the following idea; for all cloud masking approaches a continuity between haze and clouds exists, and defining a limit between them is subjective (Hagolle et al., 2010). By using a value of 20 as threshold, haze and thin clouds, in some cases, still pass which was confirmed by manually looking at randomly chosen binary masks.

Figure 11 (a) exemplarily shows a Landsat-5 scene, taken on 3 August 1986 over WRS-2 path 191, row 25, as true color (RGB) composite. Panels (b) to (d) present binary cloud masks resulting from the SCS function, with cloud score threshold of 40, 30, and 20, respectively.

Three regions (A, B, C) are marked within the RGB image to comment on the mask performance. In accordance with the definition of the SCS function, enhancing the cloud likelihood threshold from 20 to 40 drastically decreases the number of masked pixels due to the presence of hazy pixels within the image. This becomes particularly evident by looking at the large cloud field in region A. SCS 40 masks the sharp cloud edges without including hazy surroundings, particularly evident in the left (western) part the region. The use of SCS 30 results in an overmasking of these blurry regions around the cloud field, which becomes even more distinct within the SCS 20 binary product. Region B presents a collection of smaller clouds of several 10–100 m in length. SCS 40 and 30 generate similar results for this region, while further lowering the cloud score threshold to 20, again, includes more regions between the small cumulus clouds that presumably contain haze. Regarding the cloud collection in region C, similar performances are registered for all cloud thresholds within SCS. The difference between the products of SCS 20, 30, and 40 was more or less distinct, not only for different regions within a given image, but also among other Landsat images that were randomly tested. In most cases, the SCS performance was strongly coupled to the atmospheric clearness and the sharpness of cloud horizontal edges.

The SCS-40 algorithm was chosen for deriving cloud records to avoid false positives to create a permanent disturbance within the cloud record. Additionally, by manually looking at different image cases, SCS 40 showed the best overall correspondence between true color images and respective binary masks.

Panel (e) and (f) present cloud mask products of Landsat’s quality assurance bands, generated by the CFMask and ACCA algorithm, respectively. The use of operational algorithms for performing long-term cloud studies was already being excluded due to the risk of high commission errors (Mateo-García et al., 2018). In the case of Figure 11, CFMask tends to overmask in regions B and C, whereas the performance of the ACCA product corresponds well with the RGB image, similar to the SCS 40 binary mask. However, the ACCA product indicated an erroneous masking procedure for a large number of cases on GEE. A further investigation of this problem was set aside, but an example image is shown within the supplement (Figure 24) for illustration matters.
Figure 11: USGS LT05 Collection 1 Tier 1 scene, acquired over WRS-2 path 191, row 25 on 3 August 1986, given as (a) TOA reflectance displayed as a true color (Red, Green, Blue; RGB) composite, and (b) to (d) binary cloud mask derived with the SCS algorithm for cloud score thresholds of 40, 30, and 20, respectively, and (e), (f) binary cloud masks derived with CFMask (pixel_qa band), and ACCA (bqa band) algorithms, respectively.
6.2 Cloud sizes

Cloud sizes are registered as cloud horizontal lengths in two different sampling directions within horizontal stripes (each 1 px in width) from North to South, and West to East, respectively.

Table 1 gives an overview of the data set, accounting for the entire research domain and time during 1985–2018. The term “cloud number” is no representation of the real number of horizontally enclosed clouds, but represents the number of registered cloudy stripes. In total, $3.69 \cdot 10^{10}$ cloudy stripes were registered which accounts for $7.03 \cdot 10^5$ stripes per satellite image. The entire data set includes $3.07 \cdot 10^{12}$ pixels, from which $5.33 \cdot 10^{11}$ were registered as cloudy.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>total image number</td>
<td>$5.25 \cdot 10^4$</td>
</tr>
<tr>
<td>total pixel number</td>
<td>$3.07 \cdot 10^{12}$</td>
</tr>
<tr>
<td>number of columns</td>
<td>$4.10 \cdot 10^8$</td>
</tr>
<tr>
<td>number of rows</td>
<td>$3.88 \cdot 10^8$</td>
</tr>
<tr>
<td>total number of cloudy pixels</td>
<td>$5.33 \cdot 10^{11}$</td>
</tr>
<tr>
<td>cloudy pixels in column</td>
<td>$5.35 \cdot 10^{11}$</td>
</tr>
<tr>
<td>cloudy pixels in row</td>
<td>$5.31 \cdot 10^{11}$</td>
</tr>
<tr>
<td>total cloud number</td>
<td>$3.69 \cdot 10^{10}$</td>
</tr>
<tr>
<td>cloud number in columns</td>
<td>$1.89 \cdot 10^{10}$</td>
</tr>
<tr>
<td>cloud number in rows</td>
<td>$1.80 \cdot 10^{10}$</td>
</tr>
</tbody>
</table>

Table 1: Data set of cloud horizontal sizes accounting for the entire research domain covering wide ranges of the European continent during 1985–2018.

The data set is expressed separately for the two different directions, were “row” accounts for sampling in West-East, and “column” for sampling in North-South alignment, respectively. The “total” cloud number accounts for an average value derived from both directions.

6.3 Cloud size distribution

The calculation of cloud horizontal size distributions considers the number of clouds registered in latitudinal and longitudinal alignment, respectively, and follows equation 4. First, all cloud sizes were grouped together for the entire time during 1985–2018 to account for both a pan-European cloud size distribution, and locally resolved distributions per individual Landsat footprint. Second, the data set was further untangled by grouping cloud sizes for each year individually, to ultimately carry out interannual records of cloudiness over Europe. To ensure the comparability of the different results, the number of logarithmic cloud size bins was set to $i = 50$ for quantifying the changes.

Temporal average $n(L)$

33 years of Landsat data was grouped together to produce a single pan-European cloud length distribution, $\overline{n(L)}$, represented in Figure 12. The results account for cloud numbers
within logarithmically spaced length bins, plotted against a linear scale (left), and logarithmic scale (right), respectively. The observed cloud size distribution from 1–1000 px (black triangles) was passed on to a fitting approach, assuming the single power-law case \( n(L) \propto L^{-\beta} \). The power-law exponent \( \beta \) is essentially the linear regression slope of \( \ln n \) against \( \ln L \), and was further used as parameter to quantify and compare the results.

The pan-European cloud size distribution closely matches a single power-law relation that spans over three orders of magnitude of horizontal scales from 1–1000 px, corresponding to 30 m–30 km cloud length. The right panel of Figure 12 presents the observational data in the log-log space. The middle sloping straight line (black) shows the best linear fit regression, with absolute correlation of \( r = 0.996 \) between observation and fit. The overall power-law exponent (slope) is estimated with \( \beta = 1.73 \), while the blue, and green line represent exponents of \( \beta = 1.5 \), and \( \beta = 1.9 \), respectively, to visualize the impact of variable power-law parameters on the size distribution.

For scales larger than approximately 30 km (represented as grey triangles), the distribution deviates from the power law which is attributed to the presence of a horizontal scale break.

Looking at each Landsat footprint individually, again, all size distributions closely match a power-law relationship (not shown here), with very high individual correlations between observation and fit. The power-law exponent over Europe spatially varies with \( 1.63 \leq \beta \leq 1.83 \), so that the overall scaling exponent can be quantified as \( \beta = 1.73 \pm 0.10 \) to account for regional variations throughout the research domain.

Figure 13 shows all characteristic scaling exponents \( \beta \) on a map, as temporal average during 1985–2018. Grid boxes are placed on the center coordinates of the respective satellite footprint. Regions marked with a cross indicate values of \( \beta \) below the European average.
Figure 13: Map of power-law exponent $\beta$, averaged for all 33 yr of Landsat data. Center coordinates of all Landsat footprints were used for defining the geographic grid. Over Europe, the power-law exponent to the cloud size distribution varies with $1.63 \leq \beta \leq 1.83$. The power-law fit closely matches the observations with correlations giving $0.994 \leq r \leq 0.998$ throughout the grids. Crosses within the regions indicate values below the domain average value of $\beta = 1.73$.

Larger connected areas of high $\beta$ values (above the domain average) are evident within regions of the Alps and their foothills, for large parts of Poland, just as northern and central France. In theory, larger horizontal exponents indicate a steeper slope on the corresponding log-log straight line fitting the cloud distribution, accounting for the relative abundance of small-scale clouds, compared to larger cloud fields. However, even though a geographic pattern arises within the spatial distribution of $\beta$, the variation remains moderate.

Until now, the observational results account for temporal averages during 1985–2018. Within the next section, a trend estimation account for further untangling the data set to derive cloud size distributions for each year individually.

**Long-term trends of $n(L)$**

Trends in the size distribution are carried out by histogramming cloud lengths for each year separately, so that each time step accounts for an annual average of the distribution $n(L)$. Figure 14 shows annual values of $n(L)$ over WRS-2 path 196, row 24, and path 197, row 24, representing regions of the Benelux, and western Germany around the Ruhr area, respectively. Both plots show the cloud size distribution for some years during 1985–2018 in 6-yr steps, thereby only presenting a small portion of the total data to serve a first impression of a temporally changing $n(L)$. 
For a better illustration, the cloud size distribution is derived for \( i = 15 \) logarithmically spaced bins and plotted against lengths from 1–1000 px, where no scale break is expected.

![Annual cloud size distribution per Landsat footprint; \( n(L) \) for years between 1985 and 2018, in 6-year steps over satellite WRS-2 path 196, row 24 (left), and WRS-2 path 197, row 24 (right), respectively.](image)

The illustration of annual size distributions, again, supports the power-law scaling even for smaller amounts of data. Both regions reveal a similarly shifting cloud size distribution, with an overall dropping number in small-scale clouds (1–10 px, i.e., 30–300 m horizontal length), versus enhanced numbers of larger cloud fields (cloud lengths larger \(~100\) px, i.e., 3 km). This trend is consistent with a pivoting power-law distribution displayed in the log-log space, to lower values of the negative slope \( \beta \).

For quantifying the variation of cloud size distribution within the entire period and region of interest, annual values of \( \beta \) are used to derive long-term trends. Again, for the quantification, 50 log-spaced size bins are used for deriving size distributions from the observation, but estimating power-law parameters only accounts for cloud sizes prior to the scale break.
The pan-European trend of the horizontal exponent $\beta$ during 1985–2018 is represented in Figure 15.

(a) Without filtering, includes 143 Landsat tiles  
(b) After filtering, includes 21 Landsat tiles

Figure 15: Pan-European trends of the power-law scaling exponent $\overline{\beta}$ 1985–2018, (a) without filtering, and (b) after filtering, as raw data (empty squares), and low-frequency variability (5-yr moving average time series; filled squares). Error bars are indicated as standard error of mean, and a linear regression fit is used to derive the overall trend 1985–2018. $\beta$ was derived by fitting $n(L) \propto L^{-\beta}$ to the observations, for $1 \leq L \leq 1000$ px, respectively. The filtering excludes all Landsat footprints with individual correlation coefficients $r < 0.4$, and a RMSE > RMSE, respectively. From a total of 143 tiles, 21 passed the filtering.

Panel (a) of Figure 15 accounts for a trend estimate derived from all 143 Landsat tiles covering the research domain. The data indicates an overall dropping horizontal exponent over Europe, which is more clear by looking at low-frequency changes (after applying a 5-yr low-pass filter, i.e., moving average). The overall trend is not monotonic and includes a period of increasing $\beta$ during 2000–2010.

Not all time series within individual Landsat footprints showed trends that were statistically significant. Panel (b) only includes the data of satellite footprints with robust trends. Filtering among the tiles excludes all local trends with linear correlation coefficients below 0.4, and a root-mean-square error (RMSE) above the domain-average value to ensure the practicability of the linear trend model.

Both trend estimates within panel (a) and (b) reveal an overall decrease in $\beta$ during the period of interest. However, including only robust individual time-series increases the trend from $-1.6 \cdot 10^{-3}$ to $-3.7 \cdot 10^{-3}$ per year with larger correlation, while increasing the error range due to the reduced amount of data; only 21 grid boxes passed the filtering procedure from a total of 143 Landsat footprints. The decline in $\beta$ appears to be continuous for the filtered time series, with an exception during 2000–2005, showing an increase in $\beta$ within the low-frequency time series.

However, the overall filtered trend appears to be statistically significant with an absolute correlation coefficient of $r = 0.81$. The corresponding regions that passed the significance filter are marked in Figure 16, showing the spatial distribution of trends widespread over Europe; from the 21 tiles that passed the filtering, 17 are found on land and appear in the map as crosses within the respective region.
Figure 16: Grid-box trends of power-law exponent $\beta$ during 1985–2018. Regions marked with 
cross indicate statistically significant trends with individual linear correlation coefficient $r > 0.4$, 
and RMSE < RMSE. From a total of 143 Landsat tiles, 21 met the significance criteria, from 
which 17 are found on land as indicated within the map.

All statistically significant grid-box trends support a long-term decrease in the power-law 
exponent. The corresponding regions are mostly found over Germany, and in some parts of France, Great Britain and the Benelux. Considering all tiles, regions with negative 
trends dominate the geographic pattern.

It can be concluded that there was an overall dropping horizontal scaling exponent 
describing the cloud size distribution over Europe during 1985–2018, with local significant 
trends supporting the overall result. The negative trend corresponds to a large-scale piv-
oting of the negative power-law slope fitting the observation as result of both a decreased 
number of smaller clouds, and an increase in the number of larger cloud fields. The latter 
was not specifically shown here, but confirmed by looking at the pan-European trend of 
cloud number, separately within the size bins 1–10 px, 10–100 px, and 100–1000 px, re-
spectively. From those size bins, the first one, representing the smallest clouds of several 
10–100 m, indicated a negative trend, whereas for larger clouds the number increased over 
the years.

6.4 Cloud fraction

In a final step, records of cloud fraction are derived by the integral according to equation 5. 
The contribution of clouds with horizontal scales from 1–1000 px to the CF is henceforth 
referred to as “total” CF. This should not be associated with the real cloud cover, but 
the partial CF from clouds within the size range satisfying a power-law distribution. The 
actual CF was reduced not only by considering a limited range of cloud horizontal sizes, 
but also since all cloud fields outranging the Landsat footprint dimension (≈200 km) are 
automatically neglected. The same applies for clouds touching the boundaries of the satel-
ite scene. Moreover, Landsat T1 data precautionarily excludes heavily clouded scenes 
from the data set due to low quality conditions (see supplement section II about the GEE 
data structure for more details). Therefore, it can be assumed that in reality the CF was 
much larger.

Figure 17 shows the pan-European trend in the total CF over Europe (blue curve), and 
moreover, the individual contribution from different cloud size bins. The partial CF is de-

erived as contribution from clouds with 1–10, 10–100, and 100–1000 px length, respectively, 
by correspondingly adjusting the limits of integration.
Again, empty squares show the actual data, whereas filled squares represent the low-frequency variability in CF. Error bars are indicated as standard error of mean for calculating the domain-average for each data point. Again, the statistical significance filter is applied to the time series (right panel), excluding all Landsat tiles that give individual trends with absolute linear correlation \( r < 0.4 \), and an individual RMSE above the domain average.

![Figure 17: Pan-European trends of the cloud fraction CF 1985–2018, (a) without filtering, and (b) after filtering for statistical significance, as raw data (empty squares), and low-frequency variability (5-yr moving average; filled squares). Error bars are indicated as standard error of mean, and a linear regression fit is used to derive the overall trend 1985–2018. The CF was derived by fitting \( n(L) = \alpha L^{-\beta} \) to the observations, and integrating \( Ln(L) \) over \( dL \), for \( 1 \leq L \leq 1000 \text{px} \), respectively. CF trends were derived separately as the contribution from clouds within different size ranges, i.e., 1–10\,\text{px} (light grey), 10–100\,\text{px} (dark grey), and 100–1000\,\text{px} (black), by correspondingly adjusting the limits of integration. All clouds with 1–1000\,\text{px} length are contributing to the “total” CF (blue). The filtering excludes all Landsat footprints with trends showing \( r < 0.4 \), and RMSE > RMSE. From the total of 143 tiles, 9 passed the filtering.](image)

The total CF indicates an overall increase during 1985–2018 over Europe with 0.04\% per year accounting for the entire domain, and 0.15\% per year within the filtered data set. An increasing CF trend is further notable for the contribution from medium-sized and large clouds, but with decreasing significance towards smaller cloud sizes. For the smallest size bin, a slight reduction in the contribution to the CF is evident from the corresponding curve.

Since trends were determined separately for each time series, it cannot be expect that the sum of individual trends exactly adds up to the change in total CF.

The contribution of partial CF to the total CF is dominated by large-scaled clouds, which is in agreement with the observed power-law exponent \( \beta \) below the scaling threshold 2. Consequently, the overall trend in CF is also dominated by changes within larger cloud fields.

By applying a statistical filter to the time series, each individual trend (partial and total CF) is intensified and appears increasingly monotonic. However, from a total of 143 tiles, only 9 (8, 6, 12) passed the thresholds to contribute to the filtered trend of total CF (partial CF from 1–10, 10–100, 100–1000\,\text{px}, respectively). The corresponding regions are
marked in the map of Figure 18, showing the spatial distribution of trends in total CF during the period 1985–2018.

Figure 18: Grid-box trends of total CF during 1985–2018. Regions marked with cross indicate statistically significant trends with individual linear correlation coefficient $r > 0.4$, and RMSE $< \text{RMSE}$. From a total of 143 Landsat tiles, 9 met the significance criteria, from which 6 are found on land and indicated within the map.

Most grid-box trends from 1985 to 2018 show a widespread increase in CF over Europe. From 9 regions that passed the filtering, 6 are found on land and are marked in the map as crosses within the respective region. Grids with significant trends are exclusively positive, supporting the filtered time series that gives the average of these areas in panel (b) of Figure 17. Some larger connected parts of Europe indicate negative trends, especially in eastern Germany extending towards the Czech Republic, and the Benelux, but with little statistical significance so that it can be concluded that there was a widespread increase in CF over Europe during 1985–2018. However, even though the overall pan-European trend in CF is positive, there are time intervals of alternating decrease in increase evident in the unfiltered time series in panel (a) of Figure 17.

In a second step, the statistical filter is applied on individual robust trend intervals, instead of filtering the cloud record throughout the entire time period. According to the maximum linear correlation coefficient within a certain time range, the trend intervals were determined as the periods (A) 1985–1993, (B) 1994–2002, and (C) 2003–2011. Due to enlarged trend significance within the individual periods, the lower threshold of linear correlation coefficient was set to 0.5. The RMSE threshold to validate the linear trend model was used as before.
Figure 19 gives the pan-European trend in total CF, as shown before in Figure 17 (blue curve), and the interval-filtered time series (red curve).

![Pan-European trend in total CF](image)

Figure 19: Pan-European trends of total CF, without filtering (blue), and after interval filtering (red). Filtering was done separately for the trend intervals (A) 1985–1993, (B) 1994–2002, and (C) 2003–2011, and excludes all trends with individual absolute correlations coefficients \( r < 0.5 \), and \( \text{RMSE} > \text{RMSE} \). From a total of 143 tiles, 35 passed the filtering within interval (A), 33 within (B), and 19 within (C). A linear regression fit was used to derive individual trends.

The application of the statistical filter to the periods (A), (B), and (C) further intensifies the overall trend within each individual time interval. All trend results are summarized in Table 2, for both the entire period 1985–2018, and the individual trend intervals.

<table>
<thead>
<tr>
<th>period</th>
<th>total CF trend [% yr(^{-1})]</th>
<th>absolute linear ccor</th>
<th>robust trends</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>unfiltered</td>
<td>filtered</td>
<td>unfiltered</td>
</tr>
<tr>
<td>1985–2018 (entire period)</td>
<td>+0.04</td>
<td>+0.15</td>
<td>0.40</td>
</tr>
<tr>
<td>1985–1993 (A)</td>
<td>−0.31</td>
<td>−0.64</td>
<td>0.79</td>
</tr>
<tr>
<td>1994–2002 (B)</td>
<td>+0.30</td>
<td>+0.59</td>
<td>0.84</td>
</tr>
<tr>
<td>2003–2011 (C)</td>
<td>−0.20</td>
<td>−0.49</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 2: Pan-European trends in total CF for all time periods investigated. Estimations of the CF change per year, and the correlation coefficient (ccor) for the linear trend are given for both the unfiltered and filtered time-series. The number of regions that passed the filtering (due to robust trends) are given as portion among all grid boxes, in % of domain.

Each period shows linear correlations higher than the overall trend from 1985 to 2018, even before applying the filter. The correlation coefficients increased in absolute values from 0.79, 0.84, and 0.60 to 0.93, 0.99, and 0.96 for the intervals (A), (B), and (C), respectively. The trend intervals give a statistically significant change in CF that intensified by filtering the data from \(-0.31\%\) yr\(^{-1}\) to \(-0.64\%\) yr\(^{-1}\), \(+0.30\%\) yr\(^{-1}\) to \(+0.59\%\) yr\(^{-1}\), and \(-0.20\%\) yr\(^{-1}\) to \(-0.49\%\) yr\(^{-1}\), for (A), (B), and (C), respectively.

The number of tiles passing the statistical thresholds increased by filtering within individual trend intervals, with 35, 33, and 19 regions with robust trends during periods (A),...
(B), and (C), respectively. Again, corresponding grids (on land only) are marked in the map in Figure 20.

![Maps showing annual change of CF for different time periods](image)

(a) 1985 to 1993  
(b) 1994 to 2002  
(c) 2003 to 2011

Figure 20: Grid-box trends of total CF during trend intervals (A) 1985–1993, (B) 1994–2002, and (C) 2003–2011, respectively. Regions marked with *crosses* indicate statistically significant trends with individual (absolute) linear correlation coefficient *r* > 0.5, and RMSE < \(\overline{\text{RMSE}}\).

Panel (a) representing the period 1985–1993 shows that most of the grid boxes support the widespread decrease in CF. Relatively connected areas of significantly decreasing CF are found in regions around Paris, the Benelux, Berlin, Leipzig, and Prague, just as some regions of Poland. A subsequent increase in CF is notable for all robust trend regions in panel (b), representing the time during 1994–2002. The positive trend is significant in larger connected areas of eastern France extending towards the Benelux and western Germany, as well as eastern Germany, Czech Republic and Poland.

In summary, period (A) shows a widespread decrease in CF with locally robust trends, whereas period (B) indicates an increasing CF and widespread regional statistical significance within similar regions as highlighted in (A), but even more pronounced. In contrast to period (A) and (B), the last period (C) from 2003–2011 shows an overall noisy signal among the grid cells with no clear emergence of connected areas that show significant individual trends. Therefore, even though the average trending curve suggests an overall robust decrease in CF during 2003–2011, the result should be treated with some caution.
7 Discussion

For deriving large-scale and multitemporal trends of cloudiness over Europe, cloud lengths are simply registered as the number of horizontally contiguous cloudy pixels. The method is in strong agreement with the work of Wood and Field (2011), who additionally show that the results are not sensitive to whether the size distribution is determined using cloud segment length or cloud area as measure for the horizontal size. The study further indicates high agreement between size distributions derived from a single MODIS footprint, and from several granules conjoined together. Therefore, it can be assumed that the approach of using each Landsat footprint separately to extract cloud sizes is a good presentation of the real cloud distribution.

Cloud sizes registered in latitudinal or longitudinal alignment are essentially indistinguishable regarding their frequency of occurrence. To prove that, the pan-European size distribution corresponding to Figure 12 was separately derived for both sampling directions, and correlates with a value of $r = 0.9998$, which indicates that the cloud length distribution is horizontally isotropic (Guillaume et al., 2018). This further excludes the possibility of a large error induced by the fact that all Landsat images show an unequal number of pixels in width and height, which would theoretically have an influence on the size-induces sampling bias; for smaller sampling lengths, clouds are more easily excluded from the registration due to the higher chance of extending beyond the satellite scene. However, the image width-to-height ratio (expressed in Figure 21) of each satellite image gives values close to one, so the deviation from a quadratic shape is small compared to the image size. In general, the footprint size is not uniform among the geographic grid, with increasing pixel number towards higher latitudes due to altitude variations of the polar-orbiting satellite.

![Average Width-to-Height Ratio](image)

Figure 21: Scatter plot showing the horizontal dimension of all Landsat images covering the research domain. The number of pixels in x and y direction (i.e., image width and height) per image, are correspondingly assigned to x and y-axis, respectively.
The primary result of deriving an overall cloud size distribution over Europe is the emergence of a power-law scaling that spans over three orders of magnitude of cloud horizontal scales, ranging from 30 m to approximately 30 km. The power-law is further valid when breaking down the data set to account for regional and annual distributions. The quantification of the overall exponent with $\beta = 1.73$ is in good agreement with previous studies using data from satellite, dropsonde, and in situ aircraft observations, but also from numerical simulations (Guillaume et al., 2018). There are some attempts to interpret the repeated occurrence of similar horizontal exponents in observational data sets. One theoretical approach is the Kolmogorov–Obukhov theory, that states an energy spectrum exponent of $-5/3$ ($\sim 1.67$) for describing how energy in a turbulent flow follows a cascade from larger to smaller scales for homogeneous and isotropic turbulent flows (see Guillaume et al., 2018, and references therein for more information). However, gaps remain in physically interpreting how this relates to the distribution of cloud horizontal scales.

In another step, the data set was locally resolved, showing that the characteristic power-law exponent $\beta$ varies spatially among the grids within a geographic pattern. However, the local variations appear to be modest with values ranging from 1.63–1.83, accounting for $\sim 94\%$, and 106% of the average horizontal exponent. Overall, the exponent remained below the scaling threshold 2, which supports the common knowledge that the cloud cover is dominated by larger clouds, whereas the contribution to the number density increases towards smaller cloud sizes.

By breaking down the data set into annual distributions, a 33-yr trend analysis is carried out to derive long-term records in cloud fraction and the potential attribution to retreated aerosol emissions within the temporal frame.

### 7.1 Trend analysis

The leading scientific questions of this work are defined as the following:

1) What cloud changes have occurred during recent decades and what factors contributed to the trend within the observational data set?
2) Are the trends real or spurious, and if real, can they be attributed to the multidecadal aerosol retreat that has been experienced over Europe since the late 1980s, or are they rather a manifestation of natural weather and climate variability?
3) Have limitations and uncertainties within the satellite remote sensing generated relationships that could be falsely interpreted as trends?

The pan-European trend of the power-law cloud distribution indicates a large-scale decrease in the number of small clouds, together with an increase in the relative abundance of larger cloud fields of several 100 m length and more. Statistically significant grid boxes that support this trend are mostly found within regions of central and western Europe. The reduction in the number of small clouds reflects in the corresponding trend of partial CF, whereas for larger clouds the number increase contributes to a positive trend in the corresponding CF curve. Overall, there was an increase in the total CF during 1985–2018 widespread over Europe, which becomes monotonic by filtering the entire data for robust trends among the grid boxes.
The cloud record is in alignment with the findings of Norris and Wild (2007), that showed an overall dropping downward solar CCRE due to an increasing cloud cover during 1987–2002. They estimate the enhancement in cloudiness with $+0.9 \pm 1.7\%$ per decade. The Landsat data set accounts for a CF trend of $+0.4\%$ per decade within the unfiltered trend, and $+1.5\%$ per decade for the filtered time-series, which matches the estimated range of Norris and Wild (2007). Given the high correspondence between the quantifications of cloud cover increase, the first attempt is to interpret the Landsat-derived trends as real. Norris and Wild (2007) attribute the cloud record to natural weather and climate variability since a long-term increase in cloud cover occurred at the same time as aerosol concentrations decreased, which violates the lifetime effect. This argumentation alone might not be enough to exclude a major impact of European aerosol retreats, since clouds can adjust through other processes than precipitation suppression. However, according to recent conclusions of Rosenfeld et al. (2019), the isolation of the pure aerosol effect most likely results in a positive correlation, which is indeed mainly mediated by the aerosol effect on coalescence and precipitation.

Another point of interest is the role of cloud horizontal sizes in the cloud adjustments. It has been suggested that individual cloud-size classes might respond in a diverging manner to an aerosol perturbation (Xue and Feingold, 2006; Small et al., 2009). Again, smaller clouds are more exposed to entrainment drying due to their surface-to-volume ratio. The observational data, however, suggests a retreat in the number of smaller clouds and the corresponding partial CF, even though aerosol loads have been dropping. The opposite is true for larger clouds, which would not be expected from a pure aerosol impact. Rosenfeld et al. (2019) attribute negative aerosol-CF relationships from previous studies to an effect of the meteorological setup. The CF response appears to be strongly tied to atmospheric conditions (e.g., ambient relative humidity) that obscure the actual aerosol effect within observational data sets. But do local meteorological conditions reflect in the observed CF trend?

For individual cloud records among the grid boxes, the trends were mostly weak and of low statistical significance, so that the signal-to-noise ratio would justify the attribution to natural variability. Moreover, local minima within the cloud record fall together with noteworthy drought years, e.g., 2003, 2015, and 2018 (see again Figure 17 (a) and (b), without low-frequency filter). Those drought events are mostly driven by precipitation deficits and rising temperatures (Hanel et al., 2018), and provide another indication of a likely impact of local meteorology on the CF trend.

The remaining question is whether cloud changes are entirely explained by natural internal variability or whether they can be attributed to ACIs by further resolving the data set spatially and temporally. By mapping annual grid-box trends, a regional pattern arises with larger connected areas supporting the overall trend during 1985–2018, but only few, sporadically distributed robust trends among the grids.

When looking at individual significant trend intervals, the periods (A) 1985–1993, and (B) 1994–2002, show a uniform distribution of trends with larger connected areas of statistical significance, indicating an overall decrease in CF during (A), followed by an increase within (B). Even though robust grid-box trends are found in regions of known aerosol emission changes (just as for the Benelux region, the BT or the Ruhr area), interpreting the results within the context of ACIs is premature. It remains unclear what else impacts the cloud
record and to what extent local trends are individually affected. Errors within the satellite retrieval might have occurred, e.g., due to shifting satellite orbits during the investigation period. Figure 22 shows the annual average overflight time (in hours UTC) of each Landsat footprint in the left panel, with color coding representing the satellite paths from 1985–2015.

![Figure 22](image)

Figure 22: Left: time-series of the center-footprint acquisition time (satellite time stamp) for each Landsat tile within the research domain. The color coding refers to each individual Landsat path. Right: mean anomaly of data acquisition time (in minutes; black curve), together with the pan-European total CF trend 1985–2018. The correlation of both time series gives $r = 0.461$.

The mean overflight time of all Landsat paths deviates within a 2-hour range due to the measure in UTC. The local time, however, is equal for each given scene due to the polar orbit. The right panel of Figure 22 gives the annual time stamp anomaly with respect to 1985, together with the pan-European trend in total CF (without low-pass filtering; Figure 17).

Theoretically, clouds experience a diurnal cycle regarding their abundance and size, and thereby the contribution to the cloud cover. Landsat passes the equator at approximately 10:00 a.m. local time, so that data acquisition over Europe covers time ranges within 10:00–11:00 a.m. local time. During that time, an increase in cloud amount is typically expected, especially during sunlit hours within summer season (Noël et al., 2018), thereby suggesting a positive CF-time stamp relationship within that temporal frame. The question is, whether the varying acquisition time has affected the overall observed cloud trends over Europe.

While Landsat 8 is kept relatively stable since its launch in 2013, there is larger variation in the orbit of Landsat 5 within the 20th century. Both curves within panel (b) correlate with a coefficient of $r = 0.461$, indicating a moderate correspondence of the time series. Even though the correlation includes annual values (and therefore only 34 data points), there is a high visual correspondence of both trends throughout the 20th century; during 1985–1995 the time stamp was decreasing, with a maximum anomaly of −20 min in 1995, followed by an increase until 2001. This sequence of decrease and increase matches the CF trend periods (A) 1985–1993, and (B) 1994–2002, indicating a positive correlation. The third trend period (C) 2003–2011 showed a noisy pattern of sporadically distributed trends over Europe. The corresponding years within Figure 22 (b), indicate a less visual correspondence between CF and time-stamp anomaly. The results suggest that since
periods of stronger CF-time stamp relationships and large-scale significant trend periods fall together, a systematic error is likely to generate the robust trends within periods (A) and (B). The trend intervals of (A), CF increase, and (B), CF decrease are not evident within the statistically filtered time series in Figure 17, which instead shows a monotonically increasing CF during 1985–2018.

Consequently, it is possible that the impact of shifting satellite orbits is rather weak and only a decisive factor if no robust individual trend is evident in the first place. For significant trends, however, the signal might be strong enough to outweigh spurious trends, which can be further investigated in the future.

Note that the shifting overflight time suggests an inconsistent effect on different cloud-size classes; the CF-timestamp relationship for different cloud sizes gives correlations of \( r = 0.167 \) (1–10 px), \( r = 0.588 \) (10–100 px), and \( r = 0.458 \) (100–1000 px), respectively. A possible explanation is the diurnal cycle of clouds that is different for each cloud type, suggesting data acquisition happens during a time range where smaller clouds are generally less variable regarding their frequency of occurrence.

In summary, the observed CF increase over Europe 1985–2018 from independent satellite data is considered a real trend within this work. However, based upon current knowledge about the CF response, it is likely that other confounding factors contributed to the trend rather than the dropping aerosol burden over Europe during recent decades. The impact of local meteorology and climate variability are likely to obscure a potential aerosol impact. Moreover, inconsistencies within the satellite retrieval should be considered as source for systematic errors that might be generating spurious trends.
8 Conclusion and Outlook

Several anthropogenic drivers have the potential to impact cloudiness by modifying cloud radiative properties. Amongst these drivers are changes in atmospheric aerosol concentrations by affecting CCN availability which is crucial for almost all liquid droplet nucleation processes (Lohmann and Feichter, 2005). During recent decades, the European aerosol burden experienced a large-scale decrease, mainly due to the implementation of air quality policies on power generation and road transport, together with technological advancement to reduce anthropogenic emissions (Crippa et al., 2016). It has been suggested before, that European aerosol reductions have led to a large-scale decrease in cloud albedo onward from the late 1980s, which likely induced a positive RF associated with the Twomey effect (Krieger and Graßl, 2002; Norris and Wild, 2007). However, ACIs might have been additionally mediated by cloud adjustments concerning properties like the CF, lifetime, the LWP and precipitation, adding complexity to a highly coupled system. Improving the understanding of the CF response to a changing aerosol amount has the potential to generate relationships to other cloud properties that are tied to the CF (Gryspeerdt et al., 2016). So far, scientists have not reached a consensus on the sign of the aerosol-CF relationship due to the coexistence of precipitation suppression (Albrecht, 1989), and evaporation feedbacks (Ackerman et al., 2004; Small et al., 2009), which might either increase, or decrease cloudiness by adding aerosols. The cloud size distribution has been suggested to serve as observational link, but many satellite instruments provide coarse-resolution data which hampers the representation of the majority of clouds within a given scene (Kleiss et al., 2018). The Landsat mission provides over 40 years of continuous data acquisition at visible and thermal wavelengths, and high resolution of 30 m. Due to its general spectral characteristics and long-term data record it has the potential to carry out multidecadal cloud studies.

Even though the scientific motivation of this work is clear, the main emphasis was placed on developing a method for horizontal scaling of Landsat cloud imagery by using the platform Google Earth Engine. The observation of cloud changes over Europe, with data covering over 30 years at high-resolution requires high technical expertise and computational power. Pre-processing via GEE has led to a significant reduction in time and data volume by filtering suitable images through their meta data, and masking all multispectral files before exporting binary cloud masks. Cloud sizes were extracted by a simple algorithm, counting horizontally contiguous cloudy pixels along each pixel-stripe of a given scene. Cloud size distributions followed the expected power-law relation, as in earlier studies. The covariation of slope and intercept of the power-law distribution defines the relative abundance of either small or large clouds, and was further used to derive the contribution to the cloud fraction. The observations indicate an increase within the pan-European CF during 1985–2018 including some regions of significant trends, thereby showing a negative relationship between aerosol amount and CF. Considering the results of recent studies, the isolated aerosol-CF effect is likely to add up in a positive correlation (Gryspeerdt et al., 2016; Rosenfeld et al., 2019), thereby suggesting that the observed trends are rather explained by factors other
than that mediated by cloud adjustments.
A negative aerosol-CF relationship has been found in several studies before, and can be interpreted as effect of local meteorology and climate variations (Sato and Suzuki, 2019). The conclusion is supported by both the overlap of local minima within the cloud record with noteworthy drought years, and the low signal-to-noise ratio due to natural variability. Moreover, shifting satellite orbits have likely contributed to periods of alternating CF decrease and increase during the 20th century, given the diurnal cycle of cloud abundance and size.

As articulated before, additional data (e.g., trends in temperature and precipitation) needs to be considered within the chain of causality between aerosol amount and CF to isolate potential adjustments embedded in ACIs. More attention should be payed to robust trends indicated within individual cases, as they might outweigh both the effect of natural variability, and spurious trends from satellite retrieval anomalies. This topic is left as an outlook and could be pursued in subsequent studies.
Technical supplement

I Access to the GEE data catalogue

GEE allows computations by two programming languages, JavaScript and Python, the latter only via an Application Programming Interfaces (API). In order to make use of the vast amount of data in a scalable manner, GEE enables the use of algorithms developed by the community, or the generation of self-developed codes to run via Google’s computational facilities (Mateo-García et al., 2017). Heavy satellite data can be visualized and processed through massive parallelizations, with no need of downloading the files before. More information about the algorithm parallelization on GEE can be found in Mateo-García et al. (2017).

Access to the engine is granted by a web-based integrated development environment (IDE), but only by using the JavaScript programming language. The IDE provides a large variety of tools for loading a whole Landsat collection, filtering images by meta data, and for performing band math operations, e.g., within the cloud masking procedure. In addition, the IDE allows the immediate visualization of images and other products like tables and charts. Even though the online Earth Engine Code Editor uses JavaScript exclusively, a GEE Python module was developed to connect to the engine via an API. The Python API provides the same advantages by sending requests to GEE, but the visualization tools are somewhat less advanced, so that immediate looking-at images is granted only via a thumbnail URL request. However, the Python API was chosen to serve the methodical part of this work due to its more convenient way to handle long running tasks, which was essential to derive data records of high temporal and spatial expense.

II Landsat data structure on GEE

The USGS provides Landsat data in 3 categories, namely Tier 1 (T1), Tier 2 (T2) and Real-Time (RT) collections. The collection-based data classification ensures an easier identification of suitable scenes for pixel-based analyses within the Landsat data archive. Landsat imagery classified as T1 meet the highest geometric and radiometric quality requirements and are considered as appropriate usage for multitemporal analyses. More information about the classification procedure and requirements for T1 data can be found in the technical guide of Zanter (2017), and on the USGS site for Landsat Collection 1. Satellite scenes that do not meet the T1 requirements are categorized as T2 data, e.g., due to less accurate orbital information, extensive cloud cover, or ground-control insufficiency. The RT collection exclusively provides newly acquired data that has not been classified to either T1 or T2 due to a transition delay of 14 to 26 days. New images are added to the RT collection on a daily basis, but once being categorized, they are removed from the RT data set.

III Landsat scene processing methods

Landsat imagery on GEE is available in its raw form, as Surface Reflectance (SR), and Top-of-Atmosphere (TOA) reflectance. Comprehensive information about the processing methods are given within the respective chapter of the GEE guide.
Raw Landsat images are represented by digital numbers (DNs) which account for scaled radiance values. A linear transformation converts the DNs to at-sensor radiance by applying coefficients stored in the respective image metadata (Chander et al., 2009).

From the at-sensor radiance, the exoatmospheric TOA reflectance (or at-sensor reflectance) is derived by additionally taking into account the diurnal cycle of solar elevation, and seasonally varying Earth-Sun distance. The exact transformation from at-sensor radiance to TOA reflectance follows

\[
\rho_\lambda = \frac{\pi \cdot L_\lambda \cdot d^2}{ESun_\lambda \cdot \cos \Theta_s},
\]

as defined by Chander et al. (2009), with

- \(\rho_\lambda\), the planetary TOA reflectance (no unit),
- \(L_\lambda\), the spectral radiance at the sensor’s aperture [W m\(^{-2}\) sr\(^{-1}\) \(\mu\)m\(^{-1}\)],
- \(d\), the distance between Earth and Sun [astronomical units],
- \(ESun_\lambda\), the mean exoatmospheric solar irradiance [W m\(^{-2}\) \(\mu\)m\(^{-1}\)], and
- \(\Theta_s\), the solar zenith angle [degrees].

The use of the TOA reflection imposes crucial advantages when comparing data from different Landsat missions; first, the cosine effect of different solar zenith angles is compensated, and second, variations in the Earth-Sun distance throughout the acquisition period are removed.

The TOA method further converts the thermal bands from the at-sensor radiance to at-sensor brightness temperature by considering the Earth’s surface a black body. The conversion is given by

\[
T = \frac{K2}{\ln(K1/L_s) + 1}
\]

with

- \(T\), the effective at-sensor brightness temperature [K], and
- \(K1\) and \(K2\), calibration constants 1 and 2, respectively.

More details about the computation of TOA reflectance or brightness temperature are given in Chander et al. (2009), and references therein.

A different approach of providing Landsat data on GEE is via the Surface Reflectance collection. The SR is the amount of incoming solar radiation that is reflected from the Earth’s surface towards the sensor, and accounts for atmospheric effects such as aerosol scattering, and the presence of water vapor or ozone. Due to the atmospheric corrections embedded in the computation, it is an appropriate method for investigating Earth-surface changes. In case of Landsat 4/5 and Landsat 7, the images have been atmospherically corrected using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) software (Masek et al., 2006; Schmidt et al., 2013), while for Landsat-8 data, the Land Surface Reflectance Code (LaSRC) was applied (Vermote et al., 2018).

To visualize the difference between both processing methods, an example Landsat-5 image over WRS-2 path 191, row 25, acquired on 3 August 1986 is shown in Figure 23 for SR, and TOA reflectance, respectively.
Figure 23: USGS LT05 Collection-1 Tier-1 scene, acquired over WRS-2 path 191, row 25 on 3 August 1986, processed to SR (left), and TOA reflectance (right), and displayed as true color (Red, Green, Blue; RGB) composite, respectively.

The Earth Engine snippet for getting access to a collection is given as

```python
ee.ImageCollection("LANDSAT/LXSS/C01/T1_TOA")
```

for TOA-reflectance or alternatively, by replacing “TOA” with “SR” within the collection identification, for SR. The Landsat mission would be specified by adjusting LXSS to LT05, LE07, or LC08.

**IV Landsat band characteristics**

The different bands and their characteristics are summarized in Table 3, and further discussed in the context of the two processing methods of Landsat images to either SR, or TOA reflectance. More information about the band characteristics and processing methods of Landsat images can be found in the Landsat GEE data catalogue by navigating to the respective Landsat mission and collection.
### Table 3: Landsat 5 (LT05), 7 (LE07), and 8 (LC08) band characteristics. Landsat images contain four visible bands B1-B4, a near-infrared (NIR) band B5, two short-wave infrared (SWIR) bands B6 and B7, a panchromatic band B8 (LE07 and LC08 only), a cirrus band B9 (LC08 only), and several QA bands to identify pixels with certain instrumental, atmospheric, or surficial conditions. Bands with the remark “S” are processed to orthorectified SR, and bands with the remark “T” are calibrated to TOA reflectance.

<table>
<thead>
<tr>
<th>LT05</th>
<th>LE07</th>
<th>LC08</th>
<th>R</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>B1</td>
<td>0.44–0.45 μm</td>
<td>30 m ultra blue, coastal aerosol</td>
</tr>
<tr>
<td>B1 0.45–0.52 μm</td>
<td>B2 0.45–0.51 μm</td>
<td>30 m blue</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>B2 0.52–0.60 μm</td>
<td>B3 0.53–0.59 μm</td>
<td>30 m green</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>B3 0.63–0.69 μm</td>
<td>B4 0.64–0.67 μm</td>
<td>30 m red</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>B4 0.77–0.90 μm</td>
<td>B5 0.85–0.88 μm</td>
<td>30 m Near infrared (NIR)</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>B5 1.55–1.75 μm</td>
<td>B6 1.57–1.65 μm</td>
<td>30 m Shortwave infrared 1 (SWIR1)</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>B6 10.40–12.50 μm</td>
<td>B10 10.60–11.19 μm</td>
<td>Thermal Infrared 1 (TIR1)</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>B10 11.50–12.51 μm</td>
<td>30 m Thermal Infrared 2 (TIR2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B7 2.08–2.35 μm</td>
<td>B7 2.11–2.29 μm</td>
<td>30 m Shortwave infrared 2 (SWIR2)</td>
<td>S/T</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>B8 0.52–0.90 μm</td>
<td>15 m Panchromatic</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>-</td>
<td>B9 1.36–1.38 μm</td>
<td>15 m Cirrus</td>
<td>T</td>
<td></td>
</tr>
</tbody>
</table>

**Quality Assessment (QA) bands:**

- sr_aerosol
- sr_atmos_opacity
- pixel_qa
- radsat_qa
- bqa

Aerosol attributes
Atmospheric opacity
Pixel quality attributes from CFMask
Radiometric saturation QA
QA Bitmask from ACCA

Besides small differences, Landsat bands within the visible, NIR, SWIR, and TIR spectrum are contained within both the SR and TOA-reflectance collection. Only the panchromatic band of Landsat-7 and 8 data catalogues, as well as the cirrus band of the Landsat-8 mission applies to TOA reflectance only, since it doesn’t require any atmospheric correction.

Besides the radiation bands, both TOA-reflectance and SR collections include different Quality Assessment (QA) bands to further identify pixel information for advanced uses. The QA bands can be used to filter pixels by their instrumental, atmospheric, or surficial conditions in order to exclude unsuitable and low-quality data, or to meet application-specific requirements (Zanter, 2017).

The SR product within Landsat 4/5 TM, as well as Landsat 7 ETM+ contains four internal QA bands. The SR cloud QA band, “sr_cloud_qa”, is generated by the LEADAPS atmospheric correction code. It provides bit-packed cloud and surface state flags, expressed as either true or false. Specific attributes for this band are for pixels containing dark dense vegetation (which is used to estimate aerosol optical thickness), clouds (or pixels adjacent to clouds), cloud shadow, snow, and land/water (U.S. Geological Survey, 2018).

In contrast to the LEADAPS-assessed sr_cloud_qa band, the “pixel_qa” band originates from an alternative approach for quality assessment, which is the C Function of Mask.
(CFMask) algorithm (see Foga et al., 2017 for more details). The band includes bits for water, snow, cloud, cloud shadow, and cloud-confidence, and is expressed as either true, false, or as confidence level within the respective bit. For most applications, the cloud flag within the CFMask code is recommended over the LEDAPS product, since only these cloud (and cloud shadow) bits have been properly validated (Foga et al., 2017; Dwyer et al., 2018).

The radiometric saturation band, “radsat qa”, provides bit-packed information about the saturation status of the sensor during data acquisition. Pixels that are unsaturated during acquisition are classified “valid”, whereas saturated pixels are disqualified as “invalid”.

The internal SR atmospheric opacity band, “sr_atmos_opacity”, is generated by the LEDAPS atmospheric correction code and used for further assessing data suitability. The band contains a single scaled float value that is proportional to the aerosol optical depth, so that pixels with large atmospheric opacity are considered less reliable (Dwyer et al., 2018).

Similar QA bands are included in the LC08 collections, however, Landsat 8 provides more bands than the predecessor missions, so that some QA bands contain more bits. The Landsat-8 pixel_qa band is equally derived by the CFMask correction, but contains additional bits for cirrus confidence from OLI-band 9. Moreover, Landsat-8 OLI provides a higher radiometric resolution and dynamic range, which is why saturation is less common within the radsat_qa band, compared to prior missions. The standard cases where radiometric saturation leads to image disqualification within Landsat 8 is for data acquisition over volcanoes or wildfires (Dwyer et al., 2018).

Instead of the sr_cloud_qa band, the Landsat-8 SR collection provides an aerosol QA band, “sr_aerosol”, generated by the LaSRC algorithm. This band is designed to determine the quality of the atmospheric correction within the SR computation. Indications within the band determine e.g., whether the aerosol retrieval within the atmospheric correction is valid or not (Dwyer et al., 2018).

In contrast to the SR collections within Landsat 5 to 8, the TOA-reflectance product includes only one common band for quality assessment, the “bqa” band. This band is commonly provided for standard Landsat Level-1 products and gives a decimal value per pixel, which represents bit-packed information of surface, atmosphere, and sensor condition (U.S. Geological Survey, 2017). Again, a cloud flag is included within the bqa nominal product, which is generated by the Automated Cloud-Cover Assessment (ACCA) algorithm.

More details about specific cloud-masking algorithms are given in the following section, with focus on the cloud flags within pixel_qa and bqa bands, and another cloud-masking approach provided by the GEE community.

V Three cloud masking algorithms

Satellite-based cloud detection is normally based upon certain cloud characteristics that allow the discrimination from the underlying surface. Prominent spectral indices rely on cloud physical properties such as the high reflectivity in the visible ranges, or the relative cold temperature of clouds compared to the underlying ground. Common challenges for all cloud masking algorithms are the discrimination from bright targets, such as snow,
roof tops, or deserts, and the recognition of optically thin clouds (Mateo-García et al., 2018).

For Landsat, multiple approaches exist to identify Landsat pixels contaminated by clouds. This section concentrates on three different cloud masks that are provided by the GEE platform. Each of these masks is capable of working across multiple Landsat sensors, using many, or all of the Landsat bands within certain spectral tests or decision trees. Two of the approaches follow operational algorithms, namely CFMask and ACCA within the Landsat quality assurance. The third algorithm is provided as an internal code implemented in GEE, the so-called Simple Cloud Score (SCS).

ACCA

The cloud flag included in the bqa nominal product is generated by the Automatic Cloud Cover Assessment (ACCA) algorithm to derive a scene-wide cloud score (Hallahan and Prepperneau, 2013). The code requires prior conversion from raw DNs to TOA reflectance (see section III for the conversion to TOA reflectance) which is the basis for a normalized comparison between imagery acquired within different times (Irish, 2000). The cloud screening was originally designed for LE07 images and later adapted to other missions (Scaramuzza et al., 2012; Hallahan and Prepperneau, 2013). In the following explanation, the bands used for the cloud filtering refer to LE07 products.

The code within ACCA classifies clouds through a series of filters by applying multiple bands. Each filter classifies a pixel as either clouded or not cloud based upon certain thresholds, considering brightness (among the visible bands) and temperature (TIR spectrum). Detailed information can be found in Irish (2000). Most importantly, the first-pass proceedings are built upon per-pixel decision trees created by spectral indices to create per-pixel cloud masks. The first four of these indices are described more detailed, since they appear in all three masking algorithms described in this section, however in a different constellation.

As a first step within the ACCA code, the brightness test applies the visible blue band (Landsat-7 band 3) to compare to a threshold, which is set at 0.08. This index relies on the fact that clouds are highly reflecting in the visible spectrum. Therefore, only pixels exceeding the threshold will be passed on to the second filter, which is built upon the concept of the Normalized-Difference Snow Index (NDSI). The NDSI is a common method to exploit the unique spectral signature of snow to discriminate from clouds. The index applies LE07 bands 2 and 5 within the visible green and SWIR spectrum, respectively, and is calculated as

\[ \text{NDSI} = \frac{B2 - B5}{B2 + B5}. \]

The reflectance in the SWIR spectrum is very high for clouds but shows a drop in the spectral signature of snow due to the high capability to absorb at these wavelengths. However, in the visible green spectrum, both snow and clouds show similar high reflectance values (threshold tests done by e.g., Dozier, 1989, and Hall et al., 1995). Within the ACCA code, NDSI values above 0.7 are classified as snow. The third filtering applies a temperature threshold by defining a cloud temperature maximum of 300 K in the thermal bands. Pixels that fall below this value are passed on to filter four. The next test exploits
the high reflectivity within the SWIR range, and the coldness of clouds in a combined test, given as

\[
\frac{B_5}{B_6} \text{ composite} = B_6 (1 - B_5).
\]

This composite allows the discrimination from cold land surfaces by using a threshold of 255. Pixels that maintain below this value are given to the next filter, however, pixels above the threshold are marked as ambiguous and reconsidered within the pass-2 procedure. The remaining filters form ratios between different bands to further exclude non-cloudy pixels from the mask and can be studied in detail in Irish (2000). Pixels that passed all filters included in the pass-1 proceedings are finally classified as clouds and further categorized as either warm or cold. Pixels marked as ambiguous during the filtering are passed on to further analysis. The following pass-2 operation performs a statistical aggregation of the output of pass 1 to reduce remaining ambiguity. First, a spectral signature for both cloud types (warm and cold) is defined, based upon the statistical outcome of pass-1 spectral analyses. Afterwards, a thermal threshold test exclusively uses the thermal bands to untangle the ambiguous pixels. The last step, pass 3, serves as a combination of pass-1 and 2 results and finally performs a hole-filling procedure to generate a scene-wide cloud-cover score.

Note that by masking Landsat images on GEE, some inconsistencies occurred within the bqa binary product; a large number of images accounted for an erroneous masking procedure for some cloudy regions within the given scene. This matter is not further investigated, but illustrated within the example of Figure 24 for the Landsat-5 scene of WRS-2 path 191, row 25 from 28 April 1996.

![Figure 24: Failed masking procedure of the bqa binary product illustrated for the Landsat scene acquired over WRS-2 path 191, row 25 on 28 April 1996; an erroneous masking is evident for most regions covered by clouds.](image)

**CFMask**

The USGS implementation of the C Function of Mask (CFMask) algorithm is carried out as a quality assurance band, `pixel_qa`, within the SR collection on GEE (Foga et al., 2017). CFMask is the C-code based version of the Function of Mask (FMask) algorithm,
which was originally written in MATLAB and described in the work of Zhu and Woodcock (2012). In principle, the cloud screening within FMask has similarities with the ACCA algorithm by relying on a combination of spectral thresholds and tests. Four filters appear in the FMask algorithm that are combined via “and” statements to mark potentially clouded pixels. A “basic test” uses a combination of a SWIR threshold together with a brightness temperature threshold, following the strategy of filter 4 within the ACCA. Two other indices are included in the basic test, namely the already introduced NDSI, as well as the Normalized Difference Vegetation Index,

$$\text{NDVI} = \frac{B4 - B3}{B4 + B3},$$

to distinguish potential cloud pixels from vegetation.

The final Basic test reads

$$\text{Basic test} = B7 > 0.03 \text{ and } BT < 300K \text{ and } \text{NDSI} < 0.8 \text{ and } \text{NDVI} < 0.8.$$ 

The next filter uses cloud-reflectance in the visible range where clouds appear white, and is expressed by the “whiteness test” as

$$\text{Whiteness Test} = \sum_{i=1}^{3} |\text{band } i - \text{MeanVis}| < 0.7,$$

with

$$\text{MeanVis} = \frac{B1 + B2 + B3}{3},$$

the mean of all visible bands. An upper threshold of 0.7 separates potentially clouded pixels from clear-sky pixels that show high reflectance variability in the visible range, leading to an increase of whiteness-test value.

The next filter does not have a complement in the ACCA and is referred to as “HOT” (Haze Optimized Transformation) test. The statement reads

$$B1 - 0.5 \cdot B3 - 0.08 > 0,$$

and is generated empirically from regressions of DN values under clear sky (see Zhang et al., 2002 for more information). The HOT test is built up to separate among pixels containing haze, thin clouds, or clear sky. The assumption is that under clear-sky conditions, land surface areas should have highly correlating spectral signatures in the visible range, which is not the case for haze and thin clouds. Therefore, the spectral differences between visible blue and red bands serve as a separation method.

The last test is the “B4/B5” test and gives the ratio of NIR band 4 and SWIR band 5, which should remain below 0.75 for pixels being clouded. This filter is applied from Irish (2000) and appears in the same form within the ACCA proceedings (as filter 7, not mentioned here) to further remove bright land surfaces from the mask. The idea is that highly reflecting surface features, such as bright rock or desert areas, tend to be more reflective in the SWIR than in the NIR, whereas the reversed accounts for clouds.
The final pass-1 statement to identify a potential cloud pixel connects the four above-
mentioned tests via and statements, to produce one single filter. In subsequent procedures, more tests are runs for each pixel to combine to a layer of potential clouds, from which the final cloud mask is extracted. All details are given in Zhu and Woodcock (2012). In summary, both FMask and ACCA use single-image approaches and mainly rely on thresholds and rules to identify clouds based upon a set of spectral indices. The final result, the binary cloud mask, is produced for each Landsat image in an automated way and stored in the respective ready-to-use band within the GEE data catalogue.

SCS

The Simple Cloud Score (SCS) is a GEE-internal cloud-masking algorithm and treated as a function (“ee.Algorithms.Landsat.simpleCloudScore”) in both the Python API and the web-based code editor. For filtering pixels by their relative cloudiness, the SCS code assigns a cloud score band based on different cloud criteria. The criteria are in high correspondence with spectral indices used in CFMask and ACCA, using a combination of Landsat bands within the visible, NIR, SWIR and thermal range. Landsat images are filtered for brightness in the blue band, and all visible bands combined. Moreover, the reasonable high reflectivity in all IR bands appears within the filtering procedure, just as the cold nature of clouds considering the thermal Landsat bands with a threshold close to the one used in the ACCA and FMask, respectively. Again, the NDSI provides the test for discriminating clouds from snow. The entire code implementation is given in detail within the code editor sample script on GEE.

The resulting cloud mask is assigned to a new band called “cloud” which accounts for a cloud score/likelihood value. The cloud score is represented on a scale ranging from 0 to 100, where lower values indicate a smaller likelihood of a pixel being clouded. The threshold is subjective, but GEE suggests the use of a threshold within the range 20–40 (see the Introduction to Google Earth Engine and Cloud Free Composites). Pixels showing a higher cloud score than the threshold, are marked as clouded and thereby excluded from the respective Landsat image.

Multi-temporal cloud masking

Three pre-existing algorithms have been introduced for cloud masking via GEE; two of them rely on the operational CFMask and ACCA approaches which are accessible through the QA bands included in both SR and TOA-reflection collections on GEE, respectively. The third one, SCS, is implemented as a function on GEE, and represents a more simple method for scoring Landsat pixels by their relative cloudiness. All algorithms show some similarities as they rely on spectral tests and decision trees, incorporating most of Landsat’s bands within a threshold-based approach.

Mateo-García et al. (2018) found for some cases, a more simple approach of masking clouds can be advantageous for multi-temporal cloud studies. They further suggest that for a more advanced use it is profitable to control the trade-off between commission (overinclusion) and omission (or underinclusion), depending on the research goal. The argument is that for cloud screening within operational algorithms, usually overmasking errors are more accepted than the underinclusion of cloudy pixels, in order to avoid false
negatives. High commission errors are found within results from the CFMask approach, and moderate commission errors for ACCA, which still outrun the omission cases (Mateo-García et al., 2018). One explanation for the overmasking tendency of CFMask is the so-called dilation strategy, which fills any clear pixels surrounded by clouds, thereby adding to false positives. However, even for the CFMask cloud confidence band, which does not include the cloud dilation process, commission outruns omission errors (Foga et al., 2017). Automated masking algorithms are convenient for studies of land-use and land-cover changes due to the primary goal of the Landsat mission to observe the Earth’s surface. Therefore, overmasking is favored over cloud contamination that potentially affects the trend analysis as false negatives (Foga et al., 2017). One might be attempted to conclude, that for the opposite case of observing cloud trends, the underinclusion of cloudy pixels should be preferred in order to avoid false positives within the cloud record.
Bibliography


# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Surface precipitation rate and cloud fraction as a function of aerosol concentration from simulations run by Xue et al. (2008)</td>
</tr>
<tr>
<td>2</td>
<td>Schematic illustration of the evaporation-entrainment feedback by Small et al. (2009)</td>
</tr>
<tr>
<td>3</td>
<td>SO₂ and BC emission trends for different industrialized regions 1980–2000 from simulations run by Streets et al. (2006)</td>
</tr>
<tr>
<td>4</td>
<td>Map showing hotspots of reduced pollution over Europe 1970–2010 from simulations run by Crippa et al. (2016)</td>
</tr>
<tr>
<td>7</td>
<td>Results of previous model and observational studies that have estimated the effect of increased aerosol amount on cloudiness, collected by Sato and Suzuki (2019)</td>
</tr>
<tr>
<td>8</td>
<td>Global cloud size distribution from satellite observations, aircraft data and simulations given by Wood and Field (2011)</td>
</tr>
<tr>
<td>9</td>
<td>European map showing all Landsat footprints covering the research domain</td>
</tr>
<tr>
<td>10</td>
<td>Flow chart as summary of the methodical part of this work</td>
</tr>
<tr>
<td>11</td>
<td>Landsat-5 scene of WRS-2 path 191, row 25 from 3 Aug 1986 as true color composite, and as binary cloud mask from three different masking approaches, respectively</td>
</tr>
<tr>
<td>12</td>
<td>Observed pan-European cloud size distribution over Europe as temporal average 1985–2018</td>
</tr>
<tr>
<td>13</td>
<td>European map of grid-box derived power-law exponent β as temporal average 1985–2018</td>
</tr>
<tr>
<td>14</td>
<td>Annual cloud size distribution for some years within the period 1985–2018 for WRS-2 path 196, row 24, and path 197, row 24, respectively</td>
</tr>
<tr>
<td>15</td>
<td>Pan-European trend of power-law exponent β 1985–2018, both with and without statistical significance filter</td>
</tr>
<tr>
<td>16</td>
<td>European map showing grid-box trends of power-law exponent β during 1985–2018, with indicated significant trends among the grids</td>
</tr>
<tr>
<td>17</td>
<td>Pan-European trends of total, and partial CF 1985–2018, both with and without statistical significance filter</td>
</tr>
</tbody>
</table>