High-resolution satellite-based cloud detection for the analysis of land surface effects on boundary layer clouds

Julia Fuchs[1,2], Hendrik Andersen[1,2], Jan Cermak[1,2], Eva Pauli[1,2], and Rob Roebeling[3]

[1]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research, Karlsruhe, Germany
[2]Karlsruhe Institute of Technology (KIT), Institute of Photogrammetry and Remote Sensing, Karlsruhe, Germany
[3]European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), Darmstadt, Germany

Correspondence: Julia Fuchs (julia.fuchs@kit.edu)

Abstract. ~~Continental boundary layer clouds play an essential role in the climate system and are driven by processes linked to the land surface.~~ The observation of boundary layer clouds with high-resolution satellite data can provide comprehensive insights into spatiotemporal patterns of land surface-driven modification of cloud occurrence, such as the diurnal variation of the occurrence of fog holes and cloud enhancements attributed to the impact of the urban heat island. High-resolution satellite-based cloud masking approaches are often based on locally-optimized thresholds that ~~are compared against satellite-observed visible and/or infrared radiances to separate cloudy from clear-sky observations that~~ can be affected by the local surface reflectance ~~. Therefore, spatial differences in surface albedo, as found in and~~around urban areas or forests, ~~can~~ and, therefore, introduce spatial biases in the detected cloud cover~~that may impede the analysis of spatial pattern changes due to land surface influences~~. In this study, ~~two approaches for cloud masking using the High Resolution Visible channel of the Spinning Enhanced Visible and Infrared Imager aboard Meteosat Second Generation are developed and validated~~ geostationary satellite observations are used to develop and validate two high-resolution cloud masking approaches for the region of Paris to show and improve applicability for analyses of urban effects on clouds. Firstly, the Local Empirical Cloud Detection Approach (LECDA) uses an optimized threshold to separate the distribution of visible reflectances into cloudy and clear sky for each individual pixel accounting for its locally specific brightness. Secondly, the Regional Empirical Cloud Detection Approach (RECDA) uses visible reflectance thresholds that are independent of surface reflection at the observed location. Validation against in-situ cloud fractions reveals that both approaches perform similarly ~~Heidke Skill Score of 0.69~~ with a probability of detection (POD) of 0.77 and ~~0.71~~0.69 for LECDA and RECDA, respectively. ~~While the~~ Results show that with the application of RECDA a decrease of cloud cover during typical fog/low stratus conditions over the urban area of Paris for the month of November likely a result from urban effects on cloud dissipation. While LECDA is representative for the widespread usage of locally-optimized approaches, comparison against RECDA reveals that the cloud masks obtained from LECDA ~~can~~ result in regional biases of +-5 % that are most likely caused by the differences in surface reflectance in and around the urban areas of Paris. This makes the regional approach RECDA a more appropriate choice for the high-resolution satellite-based analysis of cloud cover ~~changes~~ modifications over different surface types and the interpretation of locally induced cloud processes. Further, this approach is potentially transferable to other regions and temporal scales for analysing natural and anthropogenic long-term impacts of land cover changes on clouds.

1 Introduction

The detection of continental boundary-layer clouds based on satellite data has a long tradition and is essential for the analysis of the various ways of how clouds interact with the land surface from the climate to the micro scale. Different land cover types, such as vegetation and urban surfaces exhibit distinct physical surface properties that influence the latent and sensible heat fluxes between the surface and the boundary layer and thus cloud development and spatial patterns (Shepherd, 2005; Collier, 2006; Varentsov et al., 2018).

Frequently, satellite-based analyses of the impact of land surface characteristics on boundary-layer clouds are based on comparisons of cloud fraction and its spatial anomalies with respect to different land cover types (Teuling et al., 2017; Theeuwes et al., 2019; Pauli et al.). Characteristic modifications of spatial patterns of cloudiness can emerge among others over urban areas. For example, an afternoon-cloud cover enhancement has been found over Paris during summer (Theeuwes et al., 2019), whereas fog frequencies have been observed to be lower over the urban areas of the Gangetic Plain during winter (Gautam and Singh, 2018). In this context, the urban heat island effect has been found to be a main driver of the boundary-layer cloud enhancement over Paris (Theeuwes et al., 2019). In many cases, the mechanisms by which the land-surface characteristics interact with the spatial patterns in cloudiness are not well understood (Zhong et al., 2017; Fan et al., 2016; Liang et al., 2018; Shepherd, 2005; Collier, 2006; Han et al., 2014).

Satellite data are ideally suited to map patterns of cloudiness and facilitate the analysis of land surface-atmosphere interactions at various scales. The High Resolution Visible (HRV) channel of the Spinning Enhanced Visible and Infrared Imager (SEVIRI) sensor onboard Meteosat Second Generation (MSG) is particularly useful in this context due to its high temporal (∼15min) and spatial resolutions (~~1km~~ 1 km at nadir) (Klüser et al., 2008; Deneke and Roebeling, 2010; Derrien et al., 2010; Henken et al., 2011). High resolution cloud masks enable the detection and analyses of modifications in spatial patterns in cloudiness induced by small-scale (~~smaller than~~ effective spatial resolution of HRV in this region ∼~~3~~ 2 km) features. While the higher spatial resolution is a clear advantage of the HRV channel, cloud detection from the single-spectral HRV channel is hampered by missing multi-spectral information that is available at coarser resolutions (e.g. 3 km at nadir for SEVIRI). This is shown by a large number of cloud detection algorithms that are based on multi-spectral thresholding tests to separate clear sky from different cloud types (Ackerman et al., 1998; Rossow and Garder, 1993; Saunders and Kriebel, 1988; Stowe et al., 1999; Di Vittorio and Emery, 2002; Chen et al., 2003; Hutchison et al., 2005; Cermak, 2006; Yang et al., 2007; Bley and Deneke, 2013; Andersen and Cermak, 2018).

HRV-based cloud detection methods to derive cloud masks are rare and challenging due to the limitation of observations from a single visible channel only (Schulz et al., 2012; Nilo et al., 2018). Cloud masks obtained from a single channel are often histogram-based, and seek to separate the clear sky and cloudy components of the reflectivity distribution. The separation of the distribution is typically based on a threshold that is determined by a histogram-based minima approach in a representative data set of a specific surface type and geographic region (Minnis and Harrison, 1984; Ipe et al., 2003; Bley and Deneke, 2013;

Cermak, 2006; Cermak and Bendix, 2008). For the empirical determination of the clear sky reflectivity the frequent occurrence of atmospheric and non-atmospheric features can be challenging, as aerosols, cloud shadows, thin cirrus clouds as well as fresh snow affect the retrieved reflectances (Yang et al., 2007; Ipe et al., 2003; Matthews and Rossow, 1987). In general, thresholds determined by such techniques are temporally dynamic, to account for the diurnal and seasonal variations of the solar zenith angle (Yang et al., 2007) and the vegetation period (Teuling et al., 2017; Theeuwes et al., 2019). A main problem for cloud detection is the influence of the land surface signal on the clear sky determination that can lead to severe misinterpretation of cloud occurrence over different surface types. For example, over bright surfaces a reduced contrast between clear sky reflection and cloudy-sky reflection can lead to an underdetection of clouds (EUMETSAT, 2019). To account for location-specific differences of surface albedo and to reduce the local bias, thresholds determined over land are frequently locally optimized. Teuling et al. (2017) suggested an empirical approach that flags an observation as cloudy when reflectivity exceeds a locally (per pixel) and temporally (per hour, 10-day period) adjusted clear-sky climatological value. Such locally adjusted and empirically based thresholds have a low bias in total but, still, are challenged by a degradation of detection accuracy depending on the local surface brightness (Bley and Deneke, 2013). This is caused by locally-optimized approaches that seek to find the optimal threshold to distinguish clouds and clear sky and thus have to deal with a surface-dependent bias (see explanation in section 3.2 and 1b)). For analyses of land-surface impacts on clouds it is, however, crucial that the satellite-based detection of cloud occurrence over different land cover types is ~~unbiased~~ independent of variations in surface properties, like spectral albedo. With such an application-focused algorithm there is a great potential to further improve the understanding of cloud formation and dissipation processes with respect to ~~surface properties~~regional characteristics.

In this study, two approaches to detect boundary-layer clouds on the basis of SEVIRI HRV data are presented and compared for the analysis of cloud cover changes over different surface types: the locally optimized cloud detection scheme of the Local Empirical Cloud Detection Approach (LECDA) and the Regional Empirical Cloud Detection Approach (RECDA) that is more robust to regional variations in surface reflectivity. Both approaches are processed for November of a multi-year period over a large region of Paris and its surrounding area. During this month fog is occurs frequently and can be expected to be affected by the urban region of Paris (Gautam and Singh, 2018). Results are validated with measurements of a Cloudnet station in the Southwest of Paris.

~~The paper is structured as follows.~~In Section 2 the data used is described. The methods to derive LECDA and RECDA as well as the validation with in-situ Cloudnet data are explained in Section 3. Section 4 presents the results and discussion of the comparison as well as the validation of the two cloud masking approaches, and a meteorological case using RECDA. Section 5 contains the conclusions.

2 Data

2.1 SEVIRI satellite data

The main part of this study including the generation of cloud masks is based on the analysis of SEVIRI HRV data that has a spatial resolution of 1-km at nadir and covers a broadband range from 0.4 to 1.1-μm. Together with the remaining 3 solar and

8 thermal SEVIRI channels with a spatial sampling distance of 3 km at nadir, the Earth is observed in a repeat cycle of 15 minutes covering the full disc for the low-resolution and half of the full disc for the HRV channels (Schmetz et al. (2002). It must be noted that spatial resolutions of the low and high resolution channels are effectively lower than the given sampling distance due to the satellite viewing geometry and the fact that the reflectance field is oversampled by a factor of 1.6 (Deneke and Roebeling (2010); Schn

5 , MSG Level 1.5 Image Data Format Description). For the detection and removal of snow and ice clouds channel reflectances of the visible (0.6 and 1.6 μm) and brightness temperatures of the infrared channels (8.7, 10.8 and 12.0 μm) are used, respectively (see section 3.1). In total the time period of the data set spans the month of November from 2004 to 2019 (08:00 - 16:00 UTC) resulting in $\sim$14400 time steps. The study region is the urban area of Paris and the adjacent rural areas extending from 48.0°N to 49.6°N and from 1.6°E to 3.0°E (Fig. 4b) where urban cloud modifications have already been observed and considered as

10 relevant for the urban climate of Paris during summer (Theeuwes et al., 2019). The urban area of Paris, further, serves as an ideal testbed as it is surrounded by relatively homogeneous and flat terrain with altitude ranges of $\sim$200 metres (see Fig. 6b)). The month of November was chosen to investigate a possible reduction of fog clouds over the urban area of Paris as found in other regions during winter (Gautam and Singh, 2018). Further, with the selection of one month, seasonal variations of land cover due to e.g. the vegetation period can be neglected and long-term variability of the surface reflectance can assumed to be

15 small.

2.2 Cloudnet data

The cloud masks are validated against the calibrated Cloudnet classification product (CF-1.0, Level 2) for November (2015 to 2019) at Palaiseau (48.718°N – 2.202°E) located $\sim$25km 25 km to the Southwest of Paris 156 m above sea level are used (Illingworth et al., 2007). The location of the Cloudnet station is shown in Fig. 2. Ground-based remote sensing instrumen-

20 tation including a ceilometer, a doppler cloud radar and a microwave radiometer of the atmospheric observatory SIRTA (Site Instrumental de Recherche par Télédétection Atmosphérique, Haeffelin et al. (2005)) as well as model outputs of the ECMWF Integrated Forecast System (IFS) serve as input for the Cloudnet Target classification algorithms described in ACTRIS Deliverable WP5 / D5.5 (M24). As a result a classification into 11 classes distinguishes ice and water clouds, precipitation, aerosols, insects, clear sky and combinations thereof (Appendix: Table A1) is obtained in a 30s time interval and for 719 heights above

25 sea level with a vertical resolution of 25 m (168.5-18118.5 m).

2.3 Reanalysis data

To highlight the applicability of the novel cloud mask data set generated by the regional approach to study surface-driven modifications of fog cloud cover, ERA5 data are used to filter for atmospheric conditions under which these clouds typically occur. ERA5 hourly data on single levels from 1979 to present of the European Centre for Medium-Range Weather Forecasts

30 (ECMWF; (Hersbach et al., 2018) are used. The filters consider situations averaged over the study region and SEVIRI scenes are selected only if all of the following criteria are met: 10 m wind speed $<$3ms$< 3\,ms^{-1}$ (U/V wind component), boundary layer

height $\color{red}{\sout{< 300m}}\,\underset{\sim}{\text{(blh)}} < 300\,\text{m}$, and mean sea level pressure $\color{red}{\sout{> 1020hPa}}\,\underset{\sim}{\text{(msl)}} > 1020\,\text{hPa}$. These are boundary layer conditions that characterize and facilitate the formation of fog and low stratus.

2.4 Corine land cover

For comparison of cloud fraction anomalies over different land cover types, Corine Land Cover (CLC) Level 3 data for 2012 (Version 2020_20u1) at $100\,\underset{\sim}{\text{m}}$ spatial resolution provided by the Copernicus Land Monitoring Service is used (European Environment Agency, EEA (2021)). The land use in 2012 only shows minor differences compared to the available land cover classifications in 2006 and 2018 and is chosen here to represent the period 2004–2019. CLC 2012 data are resampled to the HRV spatial resolution by using the most frequent land cover class within a HRV pixel (Fig. 4a)$\color{red}{\sout{)}}$. Only relevant land cover classes including forests, continuous/discontinuous urban fabric, arable land and pastures are considered for a more detailed analysis.

2.5 Digital elevation model

The European Digital Elevation Model (EU-DEM; version 1.1) is used to compare changes in altitude to small-scale features of cloud fraction anomalies (Fig. 6b)$\color{red}{\sout{)}}$. The data set has a spatial resolution of $25\,\underset{\sim}{\text{m}}$ and is provided by the Copernicus Land Monitoring Service (EEA, 2017; Tøttrup, 2014).

3 Methods

3.1 Snow and Ice filter

The preprocessing of the HRV data includes the elimination of ice clouds and snow in order to pertain only liquid and mixed clouds for the data analysis. This is done by using threshold tests on the SEVIRI channels at 0.6, 1.6, 8.7, 10.8 and $12\,\underset{\sim}{\mu m}$ and collocating the flagged pixels with the associated HRV pixels. In this way, data that are not relevant for the analysis of urban cloud modifications are excluded from the data set.

Snow is detected by calculating the Normalised Difference Snow Index (NDSI) based on the 0.6 and $1.6\,\underset{\sim}{\mu m}$ channels:

$$NDSI = \frac{r_{0.6} - r_{1.6}}{r_{0.6} + r_{1.6}} \tag{1}$$

where r_λ is the reflectance at wavelength $\lambda[\mu m]$ (Dozier & Painter 2004, Cermak 2006). Tests done for specific scenes revealed a threshold of 0.3 to be suitable to detect snow. Pixels that feature an NDSI above this value are excluded from the data set. This approach has been widely and operationally applied to different satellite data including MODIS and Sentinel-2 satellite sensors (Hall et al. 2001, Richter et al. 2012).

A second filter is applied to exclude ice clouds from the data set with a brightness temperature below $\color{red}{\sout{263K}}\,\underset{\sim}{263\,\text{K}}$ in the $10.8\text{-}\mu m$ channel. To exclude potentially remaining ice clouds, the difference of the 8.7 and $12\,\mu m$ channels is used in a phase test to retain only liquid clouds. Ice absorbs much more strongly than liquid water between 10 and $13\,\mu m$, while both show

a similar absorption pattern from 8 to $10\,\mu m$ (Strabala et al. 1994, Cermak 2006). The brightness difference should thus be smaller for ice clouds and is implemented as in Westerhuis et al. 2020 with the fog and low stratus confidence level CL_{FLS}:

$$CL_{FLS} = \frac{T_{LC} - (T_{12\mu m} - T_{8.7\mu m}) - T_{CCR}}{-2xT_{CCR}} \frac{T_{LC} - (T_{12\mu m} - T_{8.7\mu m}) - T_{CCR}}{-2T_{CCR}} \tag{2}$$

where T_{LC}=1.8K is the threshold for liquid clouds and T_{CCR}=1K is the cloud confidence range. Details of this approach can be found in Westerhuis et al. 2020. Pixel values with a CL_{FLS} below 0 are removed from the dataset to retain only low-level liquid clouds.

3.2 Two approaches to delineate clouds and clear sky

3.2.1 Gaussian mixture model

Based on this preprocessed data set the Local Empirical Cloud Detection Approach and the Regional Empirical Cloud Detection Approach is proposed to delineate clouds from clear sky resulting in two different cloud masks (CMloc and CMreg). The ultimate goal of RECDA is to derive spatially unbiased cloud fraction anomalies over different land cover types, while the goal of LECDA is a pixel-by-pixel cloud detection with minimum bias.

Both approaches are relatively simple and nearly independent of other channel information. The thresholds are determined empirically, based on occurrence frequencies of HRV counts in solar zenith angle (SZA) bins as shown conceptually in Fig. 1a) for a single SZA bin and pixel. The first peak of the distribution presents clear sky conditions (CS) while the second refers to cloudy (CL) conditions. In both approaches the CS composite is obtained by applying a gaussian mixture model (GMM) to the HRV reflectance histograms of each SZA bin and each pixel within the study region. A GMM is a probabilistic model that assumes that the data are generated from a mixture of a finite number (here 2: CS and CL in 1a)) of Gaussian probability distributions with unknown parameters. This method is categorized as an unsupervised learning algorithm that works in a similar way as the distance-based k-means cluster method. Instead of a distance-based model the algorithm finds the clusters based on Gaussian probability distributions. Based on the expectation-minimization algorithm, conditional expectations of the complete log-likelihood are calculated and estimates (mean and covariance) are updated iteratively until convergence to a local optimum is reached (Jain et al., 2000). While this method is limited to near-Gaussian distributions of a representative data set (excluding high SZA bins, where a bimodal distribution of clear sky and cloud is not always existent) and a sufficient amount of data it has great potential beyond the simple usage of a binary cloud mask as it provides probabilities (Sci-kit learn implementation, Pedregosa et al. (2011)).

In this setup, especially the GMM characteristics of the CS component are exploited to dynamically (pixel-based per SZA bin) derive thresholds. In the following, the computation of a local T_{loc} (1) and a regional threshold T_{reg} (2) is described based on the maximum bin count and standard deviation of the GMM CS component.
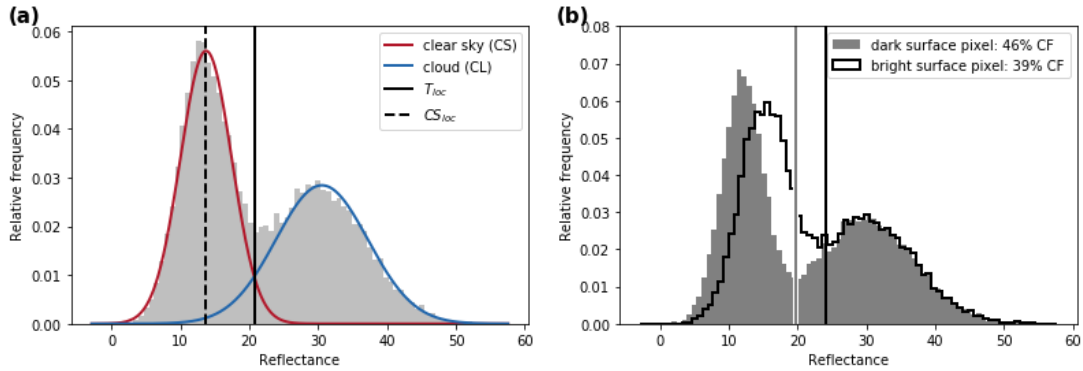
Figure 1. a) Conceptional distribution of a bimodal distribution of HRV pixel counts for one SZA bin and one pixel within the study region during November 2004–2019; 1st and 2nd component of the gaussian mixture model present clear (red) and cloudy conditions (blue) with the local clear sky reflectance CS_{loc} (vertical black dashed line) and a potential clear-cloud threshold T_{loc} (vertical black line). Clear sky reflectance is defined as the maximum count per histogram bin (maxbin) of the CS GMM component. The histogram minimum would be a potential threshold to separate the near-gaussian distributions of clear and cloudy pixels for this time and location. b) Same as (a) over a bright and a dark surface with the corresponding cloud fraction (CF). Both figures are created with 3 randomly generated gaussian distributions: a distribution for the cloudy conditions and two distributions for surfaces of different brightness (section 3.2), showing the surface-brightness-driven shift in the threshold that leads to different estimates of cloud fraction in this artificial example.

3.2.2 Local Empirical Cloud Detection Approach (LECDA)

In this approach a local HRV threshold T_{loc} is obtained based on the CS probability distribution (1st GMM component) per SZA slot (5 SZA bins 67-77; >1000 data points) and per pixel as follows. First, the local clear sky reflectance CS_{loc} (Fig. 1a))~~is~~ is determined by the reflectance value with the maximum count per histogram bin (maxbin) with a binwidth of 0.5 within

5   the $mean \pm 2\sigma$ range of the CS GMM component. Taking the maxbin approach is preferred over taking the mean of the 1st GMM component as it is representative for CS. Second, T_{loc} is computed as the sum of CS_{loc} (per pixel) and the regional median of all local 3σ of the CS GMM component. The regional median is chosen to increase the robustness of the derived T_{loc} against a sub-optimal local fit of the GMMs due to atmospheric noise as e.g. aerosols. In general, this method is assumed to be less contaminated by the effect of cloud shadows, aerosols and thin cirrus clouds compared to similar clear-sky minimum

10   approaches as it is based on the maximum bin count of the GMM clear-sky distribution (Bley and Deneke, 2013; EUMETSAT, 2019). However, the impact of these atmospheric features on the results cannot be entirely excluded and may distort the clear sky reflectance.

$$CS_{loc} = maxbin \in [mean + 2\sigma_{CSGMM}, mean - 2\sigma_{CSGMM}] \tag{3}$$

$$T_{loc} = CS_{loc} + MED_{reg}(3\sigma_{CSGMM}) \tag{4}$$

15   Thresholds obtained by LECDA as in Equ. 4 are used in e.g. Teuling et al. (2017) and present a straightforward method to approximate the "real" cloud fraction. In their study a similar local cloud mask approach is proposed where a constant thresh-

old (ten counts) is added to an empirically determined clear sky value for a given pixel based on a cumulative distribution function of reflectivity measurements. However, such locally-optimized approaches including LECDA are dependent on variations in surface albedo due to different land cover types as a constant value is added to the clear sky reflectance shown in Fig. 2 and Equ. 4. A brighter clear-sky surface pixel within the urban area of Paris will shift the clear sky distribution to higher

5 reflectance and thus clear-cloud thresholds assuming a constant cloud distribution as schematically illustrated using synthetic data in Fig. 1b). Due to the dynamic threshold method (pixel-based per SZA bin) that seeks to distinguish bright clear-sky surface pixels from clouds in a narrower bimodal distribution, a higher threshold is set in the case of brighter pixels compared to darker pixels. As a direct consequence, the application of LECDA for the retrieval of cloud fraction can potentially result in an underestimation of clouds over bright surfaces when compared to clouds over dark surfaces (example in legend of Fig. 1b)

10) given a constant cloud distribution over both surfaces.

As comparability of cloud detection over different surface albedos is not assured by LECDA and to address the potentially resulting bias, RECDA is developed that uses one regional threshold for all pixels of the study area (Equ. 5).

3.2.3 Regional Empirical Cloud Detection Approach (RECDA)

15 In this approach T_{reg} is obtained by setting the maximum of all T_{loc} as a "static" threshold that is applied to all pixels in the study area per SZA.

$$T_{reg} = max(T_{loc}) \tag{5}$$

RECDA is thought to result in a slight underestimation of clouds for all pixels as it sets the reflectance threshold to a higher reflectance value. The advantage compared to the LECDA is , however, that RECDA is independent of different surface albedo

20 values (Fig. 2) and is therefore better suitable for the comparison of regional cloud fraction anomalies over different land surfaces. The influence of the surface signal during the prevalence of thin liquid clouds may affect the cloud distribution but is assumed to be minor compared to the dependence of LECDA on the surface signal (see Sect. 4.2).

3.3 Local evaluation using Cloudnet data

Cloudnet data are used to evaluate and compare the local performance of both approaches. In order to aggregate the Cloudnet

25 Data to the temporal resolution of SEVIRI (15 minute interval), Cloudnet cloud fraction is calculated for each 15 minute time window time step based on a 1 hour time window centred around the corresponding SEVIRI time slot. This time window was chosen according to findings by Deneke et al. (2009) that compared SEVIRI satellite pixel (30 time steps). The time window of 15 minutes is approximated by the time a cloud needs to traverse the satellite reference pixel region of 3×3 pixels (∼3×5 km). This is roughly estimated based on the length of the ground track (∼ 4 km) times a length scale factor of 2 divided by the

30 average wind speed of ∼ 10 m/s at cloud top height (Greuell and Roebeling, 2009). The time window is centred around the corresponding SEVIRI time slot plus a time scan 6 km) with ground-based radiometer measurements. A correction factor of 11 minutes . This correction factor was applied was added to the SEVIRI time slot to match the SEVIRI nominal time with the
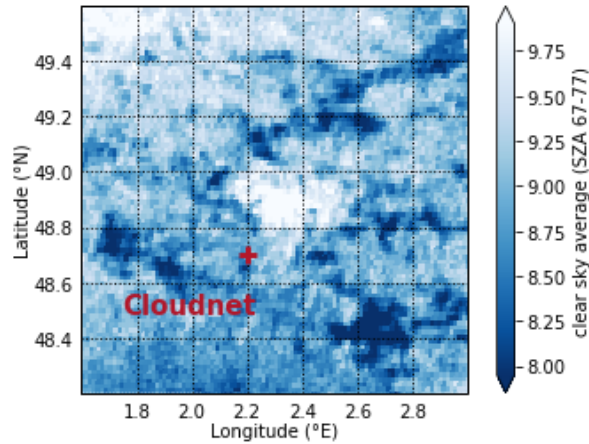
Figure 2. Average (SZA bins 67-77) of clear sky reflectance CS_{loc} obtained by Equ. 3. The location of the Cloudnet station at Palaiseau used for validation is marked in red.

actual scanning time over the validation site. Its calculation is based on the time SEVIRI needs per revolution (0.6 s) to scan 1 line (3 image lines at a time) from South to North accounting for the acquisition start at 81°S, the spreading distance of the ground resolution in S-N direction as well as the number of scan lines (MSG Level 1.5 Image Data Format Description). The W-E scan time of 30 ms is neglected. Details of the computation for the low resolution channels can be found in (Kim et al.,
5   2020) with the same nominal time as the HRV channel.

All Cloudnet Target classifications per time window ~~below an altitude of 3 km~~ are considered as relevant for the validation of the cloud mask. To create a Cloudnet cloud fraction as ground truth, classes 1 to 7 are assigned as cloudy, while 0, and 8 to 10 as clear sky (cf. Tab. A1; Veefkind et al. (2016)). If one of the 719 vertical target layers at one specific time step is classified as cloud the observation at this time step is flagged as cloud. If the resulting Cloudnet cloud fraction of ~~a~~ the defined time window
10   is above ~~0.9, the~~ 0.5, the Cloudnet flag of the corresponding SEVIRI time is assigned as cloud (1), otherwise clear sky (0). This ~~threshold~~ selected time window will ensure only persistent cloud observations to be matched with the SEVIRI cloud mask.

The temporally aggregated Cloudnet cloud mask is compared to a spatial aggregate of 3×3 HRV cloud mask pixels (∼3×6 km) centred around the cloudnet validation station. Due to the oblique viewing angle of SEVIRI a parallax correction is required resulting in a horizontally displacement of the pixel matrix surrounding the validation station by ∼2.6 km (2 pixel) to the
15   ~~South (Greuell and Roebeling, 2009)~~North (Greuell and Roebeling, 2009; Schutgens and Roebeling, 2009). If more than ~~one~~ 4 out of 9 matrix pixels are cloudy the matrix is aggregated to one value assigned as cloud (1), otherwise clear sky (0). This matrix approach was chosen in order to fully capture clouds traversing the validation sight and to reduce the effect of cloud inhomogeneity and partial cloud cover (Deneke et al., 2005).

4 Results and discussion

4.1 Comparison of the two cloud masking approaches

Fig. 3 shows cloud fractions and their regional anomalies from the Local and Regional Empirical Cloud Detection Approaches, CFloc and CFreg. Where CFloc and CFreg are defined as the number of cloudy pixels divided by the total number of pixels, and their regional anomaly ($CF_{loc}anomaly$, $CF_{reg}anomaly$) as the difference between the cloud fraction of the pixel and the average cloud fraction of the region. The spatial patterns of CF_{loc} and CF_{reg} in Fig. 3a) and c) and c are similar and show a high cloud fraction and hence a positive anomaly in Fig. 3b) and d) and d in a triangular shape east and northeast of Paris where the river valleys of the Oise, Marne and Seine are located. While CF_{loc} shows generally higher cloud fractions and a slightly larger anomaly range with more small-scale features, CF_{reg} spatial patterns are less distinct and smoother. As expected due to the algorithm differences (see section 3.2), CF_{reg} is lower (on average by 7 %) than CF_{loc}.

Differences of the two cloud-masking approaches are shown in Fig. 4a) by mapping the difference of their respective cloud fraction anomaly patterns ($CF_{loc}anomaly$-$CF_{reg}anomaly$). The differences are most prominent over the urban region of Paris and over forests (cf. Fig. 4b) and clearly connected with the clear sky reflectance of these different land cover types (cf. Fig. 2). The comparison of the respective cloud fraction anomalies reveals a relative underestimation of CF_{loc} by more than 4 % over the relatively bright urban area of Paris, while it leads to a relative overestimation of clouds over some of the relatively dark forest areas. The general dependence of LECDA on the surface reflectance is shown in Figure 5a where the difference of LECDA and RECDA is a function of clear sky surface reflectance. 75 % of the variability of the difference between RECDA and LECDA can be explained by the surface reflectance. It is assumed that the dependence on the surface reflectance is mostly attributed to LECDA, while a small portion may be attributed to RECDA in conditions of thin liquid clouds. Grouping these cloud fraction anomaly differences by land cover types confirms these observations and shows that the relative biases over the continuous urban area of Paris exceeds the biases over the forest region (Fig. 5b). The land cover type of pastures in the northwestern parts of the region also shows a relative underestimation of the local approach, however, local differences to the surrounding land cover types are not clearly visible. The contribution of land cover variability within a HRV pixel (see resampling of land cover classes, Sect. 2.4) is expected to have a marginal influence on the cloud statistics.

4.2 Analysis of cloud patterns in typical fog conditions

Patterns of cloud clearings over urban areas are mostly expected in conditions of fog or low stratus (Gautam and Singh, 2018). In order to test the application of CM_{reg} to study spatial patterns of urban cloud modification, the derived cloud mask is filtered for specific boundary layer conditions using the ERA5 reanalysis data (cf. section 2.3). Cloud fraction anomaly patterns constrained by these meteorological conditions show small-scale features that can be clearly associated with surface characteristics (Fig. 6). Notable is the distinct relative decrease of presumed fog and low stratus cloud fraction directly over the center of Paris and extending to its western edges in the order of ∼10 % relative to the regional average CF. This spatial pattern is usually associated with the dissipation of fog or low stratus by depletion of liquid water at the surface or the lifting of the cloud
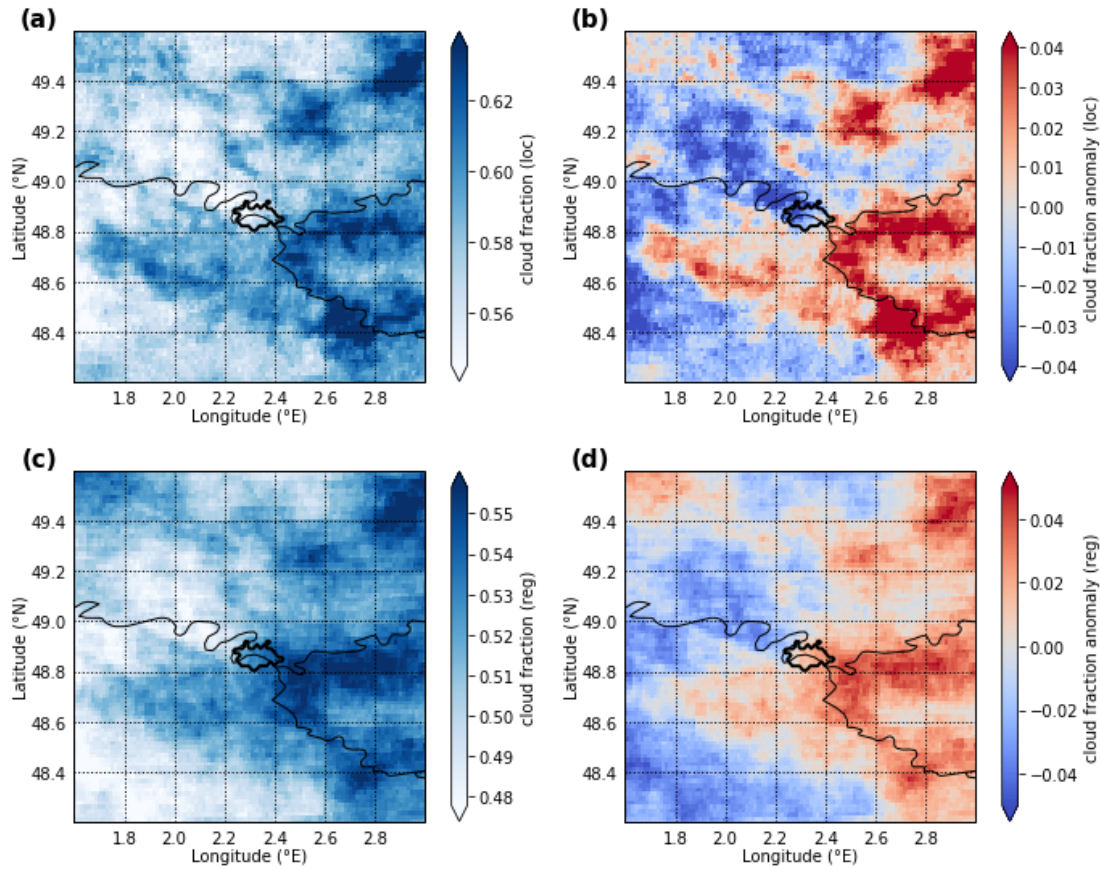
Figure 3. Cloud fraction CF_{loc} and CF_{reg} (a, c) as well as their spatial anomalies $CF_{loc}anomaly$ and $CF_{reg}anomaly$ (b, d) over the region of interest for the month of November during the years 2004-2019. Larger rivers are visualized as black lines, the continuous urban area of Paris as a black contour.

base (Wærsted et al., 2019; Williams et al., 2015; Underwood and Hansen, 2008; Haeffelin et al., 2005). This negative cloud fraction anomaly is likely a signal of the occurrences of fog holes (Gautam and Singh, 2018) over the urban center of Paris. It is interesting to note that the negative anomaly extends to the west of the city, suggesting that in the high pressure situations considered here, winds ~~maybe predominantly~~ may be predominating from easterly directions. The effect of wind direction on

5   the cloud fraction anomaly is considered as minor due the focus of conditions with low wind speeds only.

Positive cloud fraction anomalies show clear links to topography especially over river valleys where fog is naturally favoured (Bendix, 2002). This is visible in the Northeast (Oise), the Northwest (Epte, Avelon, Thérain) and East and Southeast (Marne, Seine) of the study area.The high-resolution cloud fraction anomaly map thus highlights the ability of the regional approach to detect small-scale cloud patterns that can be linked to land cover characteristics and the topography.

10   For a systematic assessment of the land-surface influences on spatial cloud fraction anomalies in the study region, cloud fraction anomalies are grouped by land cover types in Fig. 7. The land cover class "continuous urban" shows an average negative
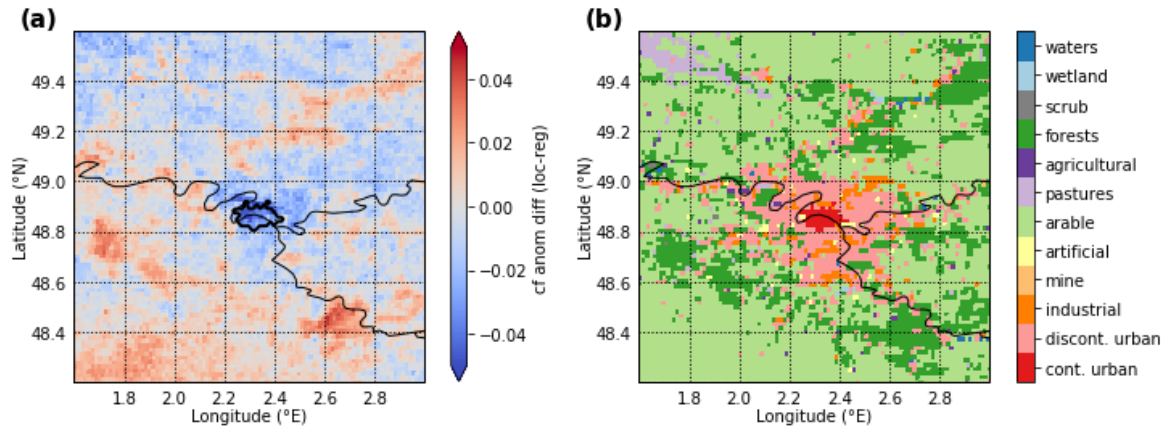
Figure 4. a) Difference of cloud fraction anomalies ($CF_{loc}anomaly$-$CF_{reg}anomaly$) from Local and Regional Empirical Cloud Detection Approaches. b) Main Corine Land Cover classes L2 and L3 for the year 2012. Larger rivers are visualized as black lines.
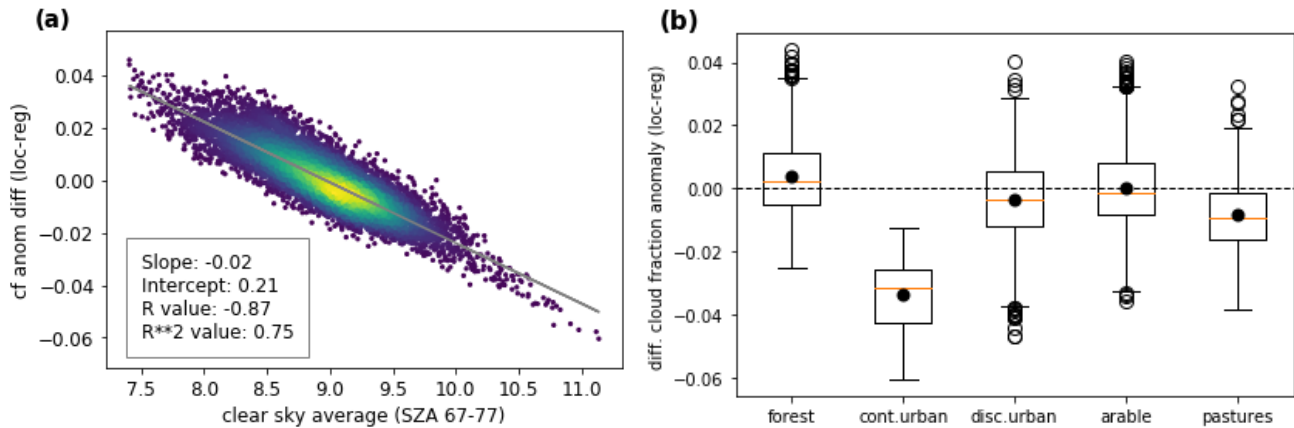
Figure 5. a) Linear regression: Difference of cloud fraction anomalies (LECDA-RECDA) vs. clear sky reflectance (average of SZA bins 67-77). Bright colors represent a higher probability density of the data points using Gaussian kernels. b) Difference of cloud fraction anomalies ($CF_{loc}anomaly$-$CF_{reg}anomaly$) from Local and Regional Empirical Cloud Detection Approaches grouped by 5 main Corine Land Cover types: forest, continuous urban fabric, discontinuous urban fabric, arable land and pastures. Median and mean of each class are presented as orange horizontal lines and black dots, respectively. Whiskers are equal to 1.5 times IQR beyond the first and third quartiles. Outliers are shown as black circles.

cloud fraction anomaly of 6 % that can be interpreted as a lower bound of the magnitude of the city's effect on fog occurrence, as Paris is located in a river valley and should thus have favourable conditions for fog formation. Over pastures an increase by 4 % is notable, however, this is likely linked to the location of the pastures in the Avelon river valley in the NW of the study area. Smaller anomalies of the remaining land cover classes are partly explained by the combined influence of land cover type,

terrain height and the dynamic of the fog hole evolution itself. The way this cloud pattern is linked to the urban heat island of Paris, land surface characteristics and further meteorological variables will require further analysis beyond the scope of this paper.
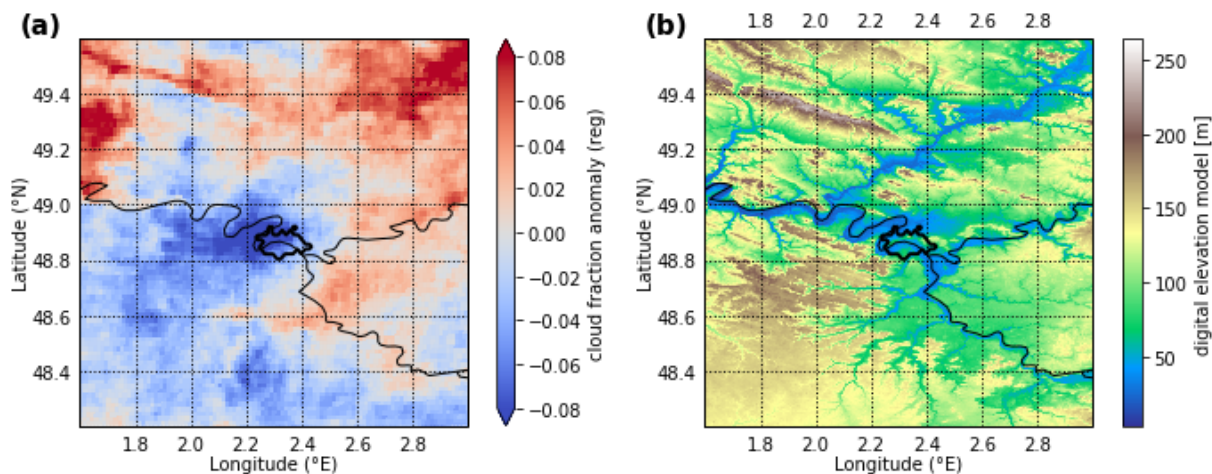
Figure 6. a) Cloud fraction CF_{reg} anomaly (difference between the cloud fraction of the pixel and the average cloud fraction of the region) from the Regional Empirical Cloud Detection Approach constrained by the following meteorological conditions: low wind speed (<3m/s 3 ms^{-1}), low blh (<300m 300 m), high msl (>1020hPa 1020 hPa). b) European Digital Elevation Model with a spatial resolution of 25m. The Seine river is visualized as black line.
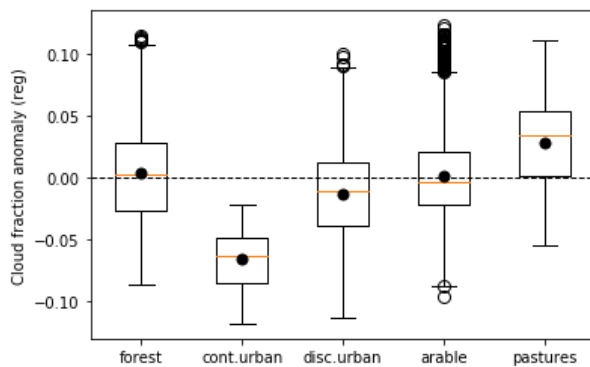
Figure 7. Cloud fraction CF_{reg} anomaly from the Regional Empirical Cloud Detection Approach constrained by meteorological conditions (Fig. 6) and grouped by 5 main Corine Land Cover types: forest, continuous urban fabric, discontinuous urban fabric, arable land and pastures. Details as in Fig. 5.

4.3 Validation with ground-based cloud fraction

The validation of the two cloud masks CM_{loc} and CM_{reg} with respect to the Cloudnet data seeks to compare both approaches with ground-truth data using standard measures of performance. However, it is not intended to prove the better applicability of one over another for the analysis of land surface effects on boundary layer clouds as this was shown in Figure 4a and 5a.

In Tab. 1 it is shown that both approaches perform similar in terms of overall quality, but achieve this with very different performance characteristics. In general, CM_{loc} outperforms CM_{reg} with a probability of detection (POD) of 0.77 compared to 0.69. However, the false alarm ratio (FAR) of 0.0 for CM_{reg} falls below that for CM_{loc} of 0.04. These results are expected as T_{loc} is generally lower than T_{reg} (except where $T_{loc} = \max(T_{loc})$), leading to a general underestimation of clouds in CM_{reg} with the benefit of reducing false alarms. It is notable that a main difference in validating both approaches originates from the definition of a more cloud-conservative threshold vs. a more clear-conservative threshold. The Cloudnet data validate both approaches over a location with a surface reflectance of 8.65 (see Fig. 2) that is close to the scene average of 8.96. It is expected that the validation of the local and regional approach will give the same or nearly identical results over the bright surfaces of the city, while over darker forest regions, the local approach is likely to have a much higher POD, with little influence on the FAR.

For assessing the general performance of both approaches (CM_{reg} and CM_{loc}) in the context of various thresholds, POD and FAR are computed along the range of HRV values and exemplary for different SZA (Fig. 8). The resulting pseudo-ROC (receiver operating characteristic) curves can be used to determine the ideal threshold for cloud detection, which typically can be found in the upper left-hand corner with a minimal FAR and a maximal POD. The main outcome is that CM_{reg} for the SZA bins 67(–69) and 69(–71) is close to being an optimal tradeoff, achieving a low FAR while loosing only little detection skill (POD) when compared to CM_{loc} and alternative HRV thresholds. The proposed regional approach, however, shows a lower POD compared to the local approach that can be attributed to the general cloud underestimation that is true for all the different surface reflectances equally. ~~The ROC curves further show a degradation of performance in SZA bins > 71 in both approaches. This SZA dependence of surface albedo and thus clear sky reflection is expected due to the gradual approximation of cloud and clear sky reflectance in the morning and evening twilight and known by other cloud mask algorithms as well, e.g. EUMETSAT (2019). Due to the similar signature of clouds and clear skies the number of false alarms is increased while the POD decreases. The degradation of performance towards lower SZA (not shown) is explained by the limited amount of data that is available for the lowest SZA bins. In general, the month of November, though relevant for this study, has only a limited number of daytime hours and thus available data per SZA slot~~ Further, for the application suggested in this study the ROC cannot provide a measure to evaluate the performance of LECDA compared to RECDA as done in section 4.2.

The ground-based validation of the satellite-retrieved cloud masks is difficult as differences can be due to the specific measurement methods, including the different scales and the way cloud net data is aggregated to match the satellite data (cf. (Greuell and Roebeling, 2009)).

The performance of RECDA may vary depending on the location-specific characteristics of the land cover type, the satellite viewing geometry and the domain size. It is assumed that the performance of RECDA similar to LECDA will decrease in cases where the bimodal distribution broadens and flattens due to the effect of aerosols, thin liquid clouds, subpixel clouds or cloud edges contributing to the overlap area. It is expected that e.g. the presence of thin liquid clouds can lead to a surface-dependent bias in RECDA that is assumed to be minor compared to the influence of the surface signal in LECDA. Based on this study we can recommend RECDA for the proposed application and suggested domain size with a clear sky reflectance between 8 and 10 % (see Figure 2.). RECDA will not provide reliable results over regions with clear sky reflectances varying between 8 and 50% as e.g. over agricultural and desert regions. In addition, the gradual approximation of cloud and clear sky reflectance in the morning and evening twilight contributes to the flattening of the distribution curve and will degrade the ability of both algorithms to separate clouds and clear sky. The SZA dependence of the clear sky reflection is known by other cloud mask algorithms as well, e.g. EUMETSAT (2019).

Table 1. Cloudnet validation results for the cloud masks using the local (CM_{loc}) and the regional (CM_{reg}) threshold. Validation measures: probability of detection (POD), false alarm ratio (FAR), percentage correct (PC), critical success index (CSI), bias score (BS) and Heidke skill score (HSS).

Class	POD	FAR	PC	CSI	BS	HSS
CM_{loc}	~~0.95~~ 0.77	~~0.25~~ 0.04	~~0.84~~ 0.86	~~0.72~~ 0.75	~~1.27~~ 0.8	~~0.69~~ 0.72
CM_{reg}	~~0.86~~ 0.69	~~0.18~~ 0.0	~~0.86~~ 0.83	~~0.72~~ 0.68	~~1.05~~ 0.69	~~0.71~~ 0.66

5 Conclusions

This study presents and compares the applicability of two empirical cloud masking approaches based on the single HRV channel of MSG SEVIRI for the high-resolution analysis of land surface effects on boundary layer clouds. The performance of the cloud mask based on the regional approach RECDA is compared to the local approach LECDA with respect to the influence of the underlying surface. Both cloud mask approaches and obtained cloud fractions are analysed over the $\sim$150$\times$150 km region centered on Paris, where holes in fog and low stratus are expected.

It is shown that cloud masks obtained from LECDA result in a relative underestimation of cloud occurrence over the bright urban surface of Paris (up to 5 %), while cloud occurrence is overestimated in relative terms over dark surfaces as e.g. forests (up to 5 %) when compared to RECDA. This leads to the conclusion that studies using such locally-optimized approaches have to be attentive when associating different land surface classes to cloud occurrence.

The application of the regional approach RECDA to compute cloud masks in contrast to a pixel-based surface-dependent threshold is shown to be advantageous as local biases due to differences of surface reflectance can be mostly prevented. As suggested in the study's context the ~~surface-independent~~ cloud mask obtained from empirically based regional thresholds is
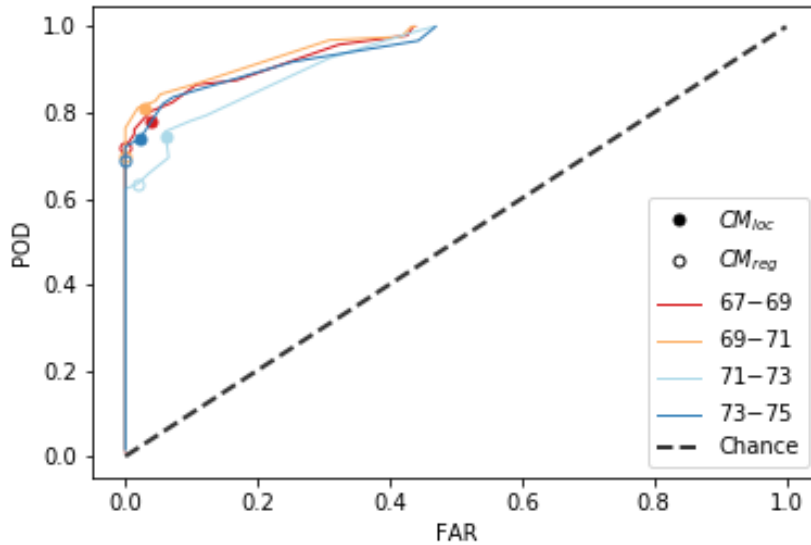
Figure 8. Relationship of POD and FAR for local and regional cloud masks (CM_{loc} and CM_{reg}; filled and transparent circles) and for a range of different HRV threshold values and four Solar Zenith Angle (SZA) bins (colored lines). HRV threshold values increase from top to bottom of the plot, from HRV minimum to HRV maximum per SZA. The pseudo-ROC curve shows the best performance of the HRV threshold values used for cloud masking in the upper left corner, while accuracy decreases towards the dashed diagonal.

more robust towards variability of the surface reflectance and show a reduced FAR. Based on this approach a reduction of cloud cover filtered for typical fog/low stratus conditions up to 10 % could have been observed for the month of November over ~~and to the west of~~ the urban area of Paris and west of it. This spatial cloud pattern is associated with the urban surface that probably affects the boundary layer through gradual heat release by the urban surface. In addition to this prominent decrease

5   small-scale cloud enhancements attributed to river valleys, topography and forests regions prove the plausibility of the cloud mask. It should be noted that the cloud masks obtained with RECDA are likely underestimating cloud occurrence so that their application is limited to the quantification and analysis of regional cloud cover anomalies.

This study shows the great potential of ~~a~~ an HRV-derived robust ~~,~~ cloud mask that is designed to be more land-surface ~~independent cloud mask~~ independent than locally optimized approaches, and thus ideal for the satellite-based quantification of

10   small-scale interactions of the urban surface and boundary layer clouds in the range of ∼1-2 km. The relatively simple and independent implementation and application of this approach for regional and urban analyses is expected to have a diverse potential on an Europe-wide scale that need to be tested for a varying domain size and temporally varying surface reflectance. Future research may also include the expansion and validation of this approach to regions that have experienced man-made or naturally induced land cover changes. Understanding the impact of vegetation changes on clouds due to e.g. heat stress is

Table A1. Cloudnet target classification

Class	Definition
0	Clear sky
1	Cloud liquid droplets only
2	Drizzle or rain
3	Drizzle or rain coexisting with cloud liquid droplets
4	Ice particles
5	Ice coexisting with supercooled liquid droplets
6	Melting ice particles
7	Melting ice particles coexisting with cloud liquid droplets
8	Aerosol particles, no cloud or precipitation
9	Insects, no cloud or precipitation
10	Aerosol coexisting with insects, no cloud or precipitation

a relevant aspect in a changing climate and can be addressed by combining this cloud masking approach with future satellite missions such as FLEX (FLuorescence EXplorer).

Code and data availability.

Author contributions. JF fully developed the concept and methodology and wrote the software, obtained and analyzed the data sets, conducted the original research and wrote the manuscript. JF, HA and JC discussed and improved the algorithm. EP preprocessed the CLC data. RR contributed to the validation of the algorithm. HA contributed to the interpretation of the results. JF, HA, JC, EP and RR reviewed and edited the manuscript.

Competing interests. No competing interests are present

Acknowledgements. We acknowledge the use of CloudNet data which are part of the European Aerosol, Clouds and Trace Gases Research Infrastructure (ACTRIS) project. The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654109 and the Cloudnet project (European Union contract EVK2-2000-00611) for providing the Cloudnet classification product, which was produced by the Department of Meteorology, University of Reading using

measurements from the atmospheric observatory SIRTA at Palaiseau. ERA5 data (Hersbach et al., 2018) were obtained from the Copernicus Climate Change Service (C3S) Climate Data Store. The European Environment Agency (EEA) is acknowledged for the CORINE land cover (https://land.copernicus.eu/pan-european/corine-land-cover) and the EU-DEM (https://land.copernicus.eu/imagery-in-situ/eu-dem). The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. Eva Pauli has been financially supported by the Graduate Funding from the German States. We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

References

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., and Gumley, L. E.: Discriminating clear sky from clouds with MODIS, Journal of Geophysical Research Atmospheres, 103, 32 141–32 157, https://doi.org/10.1029/1998JD200032, 1998.

Andersen, H. and Cermak, J.: First fully diurnal fog and low cloud satellite detection reveals life cycle in the Namib, Atmospheric Measurement Techniques, 11, 5461–5470, https://doi.org/10.5194/amt-11-5461-2018, 2018.

Bendix, J.: A satellite-based climatology of fog and low-level stratus in Germany and adjacent areas, Atmospheric Research, 64, 3–18, https://doi.org/10.1016/S0169-8095(02)00075-3, 2002.

Bley, S. and Deneke, H.: A robust threshold-based cloud mask for the HRV channel of MSG SEVIRI, Atmospheric Measurement Techniques Discussions, 6, 2829–2855, https://doi.org/10.5194/amtd-6-2829-2013, 2013.

Cermak, J.: SOFOS – A New Satellite-based Operational Fog Observation Scheme, PhD Thesis, Philipps-Universität Marburg, http://archiv.ub.uni-marburg.de/diss/z2006/0149/, 2006.

Cermak, J. and Bendix, J.: A novel approach to fog/low stratus detection using Meteosat 8 data, Atmospheric Research, 87, 279–292, https://doi.org/10.1016/j.atmosres.2007.11.009, https://linkinghub.elsevier.com/retrieve/pii/S0169809507002074, 2008.

Chen, P. Y., Srinivasan, R., and Fedosejevs, G.: An automated cloud detection method for daily NOAA 16 advanced very high resolution radiometer data over Texas and Mexico, Journal of Geophysical Research: Atmospheres, 108, 1–9, https://doi.org/10.1029/2003jd003554, 2003.

Collier, C. G.: The impact of urban areas on weather, Quarterly Journal of the Royal Meteorological Society, 132, 1–25, https://doi.org/10.1256/qj.05.199, http://doi.wiley.com/10.1256/qj.05.199, 2006.

Deneke, H., Feijt, A., van Lammeren, A., and Simmer, C.: Validation of a physical retrieval scheme of solar surface irradiances from narrowband satellite radiances, Journal of Applied Meteorology, 44, 1453–1466, https://doi.org/10.1175/JAM2290.1, 2005.

Deneke, H. M. and Roebeling, R. A.: Downscaling of METEOSAT SEVIRI 0.6 and 0.8 μm channel radiances utilizing the high-resolution visible channel, Atmospheric Chemistry and Physics, 10, 9761–9772, https://doi.org/10.5194/acp-10-9761-2010, 2010.

Deneke, H. M., Knap, W. H., and Simmer, C.: Multiresolution analysis of the temporal variance and correlation of transmittance and reflectance of an atmospheric column, Journal of Geophysical Research Atmospheres, 114, https://doi.org/10.1029/2008JD011680, 2009.

Derrien, M., Gléau, H. L., and Raoul, M.: The use of the high resolution visible in SAFNWC/MSG cloud mask, Proceedings of the 2010 EUMETSAT Meteorological Satellite Conference, 20–24 September 2010, Córdoba, Spain, p. 57 pp., http://hal.archives-ouvertes.fr/meteo-00604325/, 2010.

Di Vittorio, A. V. and Emery, W. J.: An automated, dynamic threshold cloud-masking algorithm for daytime AVHRR images over land, IEEE Transactions on Geoscience and Remote Sensing, 40, 1682–1694, https://doi.org/10.1109/TGRS.2002.802455, 2002.

EEA: Copernicus Land Monitoring Service - Reference Data: EU-DEM, https://land.copernicus.eu/user-corner/publications/eu-dem-flyer/at_download/file, 2017.

EEA: CORINE Land Cover - User Manual, Copernicus Land Monitoring Service, https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata%0Ahttps://land.copernicus.eu/pan-european/corine-land-cover/clc2018%0Ahttps://land.copernicus.eu/pan-european/corine-land-cover/clc2018%0Ahttps://land.copernicus.eu/pan-europea, 2021.

EUMETSAT: EUMETSAT CM SAF Algorithm Theoretical Basis Document (ATBD) Meteosat Solar Surface Radiation and effective Cloud Albedo Climate Data Records - Heliosat (SARAH-2), p. 31, https://doi.org/10.5676/EUM_SAF_CM/SARAH/V002_01, 2019.

Fan, J., Wang, Y., Rosenfeld, D., Liu, X., Fan, J., Wang, Y., Rosenfeld, D., and Liu, X.: Review of Aerosol-Cloud Interactions: Mechanisms, Significance, and Challenges, Journal of the Atmospheric Sciences, 73, 4221–4252, https://doi.org/10.1175/JAS-D-16-0037.1, http://journals.ametsoc.org/doi/10.1175/JAS-D-16-0037.1, 2016.

Gautam, R. and Singh, M. K.: Urban Heat Island Over Delhi Punches Holes in Widespread Fog in the Indo-Gangetic Plains, Geophysical Research Letters, 45, 1114–1121, https://doi.org/10.1002/2017GL076794, 2018.

Greuell, W. and Roebeling, R. A.: Toward a standard procedure for validation of satellite-derived cloud liquid water path: A study with SEVIRI data, Journal of Applied Meteorology and Climatology, 48, 1575–1590, https://doi.org/10.1175/2009JAMC2112.1, 2009.

Haeffelin, M., Barthès, L., Bock, O., Boitel, C., Bony, S., Bouniol, D., Chepfer, H., Chiriaco, M., Cuesta, J., Delanoë, J., Drobinski, P., Dufresne, J. L., Flamant, C., Grall, M., Hodzic, A., Hourdin, F., Lapouge, F., Lemaître, Y., Mathieu, A., Morille, Y., Naud, C., Noël, V., O'Hirok, W., Pelon, J., Pietras, C., Protat, A., Romand, B., Scialom, G., and Vautard, R.: SIRTA, a ground-based atmospheric observatory for cloud and aerosol research, Annales Geophysicae, 23, 253–275, https://doi.org/10.5194/angeo-23-253-2005, 2005.

Han, J.-Y., Baik, J.-J., and Lee, H.: Urban impacts on precipitation, Asia-Pacific Journal of Atmospheric Sciences, 50, 17–30, https://doi.org/10.1007/s13143-014-0016-7, http://link.springer.com/10.1007/s13143-014-0016-7, 2014.

Henken, C. C., Schmeits, M. J., Deneke, H., and Roebeling, R. A.: Using MSG-SEVIRI cloud physical properties and weather radar observations for the detection of Cb/TCu clouds, Journal of Applied Meteorology and Climatology, 50, 1587–1600, https://doi.org/10.1175/2011JAMC2601.1, 2011.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.adbb2d47, 2018.

Hutchison, K. D., Roskovensky, J. K., Jackson, J. M., Heidinger, A. K., Kopp, T. J., Pavolonis, M. J., and Frey, R.: Automated cloud detection and classification of data collected by the Visible Infrared Imager Radiometer Suite (VIIRS), International Journal of Remote Sensing, 26, 4681–4706, https://doi.org/10.1080/01431160500196786, 2005.

Illingworth, A. J., Hogan, R. J., O'Connor, E. J., Bouniol, D., Brooks, M. E., Delanoë, J., Donovan, D. P., Eastment, J. D., Gaussiat, N., Goddard, J. W., Haeffelin, M., Klein Baltink, H., Krasnov, O. A., Pelon, J., Piriou, J. M., Protat, A., Russchenberg, H. W., Seifert, A., Tompkins, A. M., van Zadelhoff, G. J., Vinit, F., Willen, U., Wilson, D. R., and Wrench, C. L.: Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-based observations, Bulletin of the American Meteorological Society, 88, 883–898, https://doi.org/10.1175/BAMS-88-6-883, 2007.

Ipe, A., Clerbaux, N., Bertrand, C., Dewitte, S., and Gonzalez, L.: Pixel-scale composite top-of-the-atmosphere clear-sky reflectances for Meteosat-7 visible data, Journal of Geophysical Research: Atmospheres, 108, 1–7, https://doi.org/10.1029/2002jd002771, 2003.

Jain, A., Duin, R. P. W., and Mao, J.: Statistical pattern recognition: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 4–37, https://doi.org/10.1109/34.824819, http://ieeexplore.ieee.org/ielx5/34/17859/00824819.pdf?tp=&arnumber=824819&isnumber=17859%5Cnhttp://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=824819, 2000.

Kim, M., Cermak, J., Andersen, H., Fuchs, J., and Stirnberg, R.: A new satellite-based retrieval of low-cloud liquid-water path using machine learning and meteosat seviri data, Remote Sensing, 12, 1–16, https://doi.org/10.3390/rs12213475, 2020.

Klüser, L., Rosenfeld, D., MacKe, A., and Holzer-Popp, T.: Observations of shallow convective clouds generated by solar heating of dark smoke plumes, Atmospheric Chemistry and Physics, 8, 2833–2840, https://doi.org/10.5194/acp-8-2833-2008, 2008.

Liang, X., Miao, S., Li, J., Bornstein, R., Zhang, X., Gao, Y., Chen, F., Cao, X., Cheng, Z., Clements, C., Dabberdt, W., Ding, A., Ding, D., Dou, J. J., Dou, J. X., Dou, Y., Grimmond, C. S., Gonzál, J. E., He, J., Huang, M., Huang, X., Ju, S., Li, Q., NiYogi, D., Quan, J., Sun,

J., Sun, J. Z., Yu, M., Zhang, J., Zhang, Y., Zhao, X., Zheng, Z., and Zhou, M.: Surf: Understanding and predicting urban convection and haze, Bulletin of the American Meteorological Society, 99, 1391–1413, https://doi.org/10.1175/BAMS-D-16-0178.1, 2018.

Matthews, E. and Rossow, W. B.: Regional and Seasonal Variations of Surface Reflectance from Satellite Observations at 0.6 μm, Journal of Climate and Applied Meteorology, 26, 170–202, https://doi.org/10.1175/1520-0450(1987)026<0170:RASVOS>2.0.CO;2, http://journals.ametsoc.org/doi/10.1175/1520-0450(1987)026%3C0170:RASVOS%3E2.0.CO;2, 1987.

Minnis, P. and Harrison, E. F.: Diurnal variability of regional cloud and clear-sky radiative parameters derived from GOES data. Part III: November 1978 radiative parameters., Journal of Climate Applied Meteorology, 23, 1032–1051, https://doi.org/10.1175/1520-0450(1984)023<1032:DVORCA>2.0.CO;2, 1984.

Nilo, S. T., Romano, F., Cermak, J., Cimini, D., Ricciardelli, E., Cersosimo, A., Di Paola, F., Gallucci, D., Gentile, S., Geraldi, E., Larosa, S., Ripepi, E., and Viggiano, M.: Fog detection based on Meteosat Second Generation-Spinning enhanced visible and infrared imager high resolution visible channel, Remote Sensing, 10, 1–20, https://doi.org/10.3390/rs10040541, 2018.

Pauli, E., Cermak, J., and Teuling, A. J.: Enhanced nighttime fog and low stratus occurrence over the Landes forest, France, Geophysical Research Letters (in review).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, https://doi.org/10.1007/s13398-014-0173-7.2, http://dl.acm.org/citation.cfm?id=2078195%5Cnhttp://arxiv.org/abs/1201.0490http://arxiv.org/abs/1201.0490, 2011.

Rossow, W. B. and Garder, L. C.: Cloud Detection Using Satellite Measurements of Infrared and Visible Radiances for ISCCP, Journal of Climate, 6, 2341–2369, https://doi.org/10.1175/1520-0442(1993)006<2341:CDUSMO>2.0.CO;2, http://journals.ametsoc.org/doi/10.1175/1520-0442(1993)006%3C2341:CDUSMO%3E2.0.CO;2, 1993.

Saunders, R. W. and Kriebel, K. T.: An improved method for detecting clear sky and cloudy radiances from AVHRR data, International Journal of Remote Sensing, 9, 123–150, https://doi.org/10.1080/01431168808954841, 1988.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., and Ratier, A.: An Introduction to METEOSAT Second Generation (MSG), Bulletin of the American Meteorological Society, pp. 977–992, https://doi.org/10.1175/1520-0477(2002)083<0977, 2002.

Schulz, H. M., Thies, B., Cermak, J., and Bendix, J.: 1 km fog and low stratus detection using pan-sharpened MSG SEVIRI data, Atmospheric Measurement Techniques, 5, 2469–2480, https://doi.org/10.5194/amt-5-2469-2012, 2012.

Schutgens, N. A. and Roebeling, R. A.: Validating the validation: the influence of liquid water distribution in clouds on the intercomparison of satellite and surface observations, Journal of Atmospheric and Oceanic Technology, 26, 1457–1474, https://doi.org/10.1175/2009JTECHA1226.1, 2009.

Shepherd, J. M.: A Review of Current Investigations of Urban-Induced Rainfall and Recommendations for the Future, Earth Interactions, 9, 1–27, https://doi.org/10.1175/EI156.1, http://journals.ametsoc.org/doi/abs/10.1175/EI156.1, 2005.

Stowe, L. L., Davis, P. A., and Mcclain, E. P.: Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the advanced very high resolution radiometer, Journal of Atmospheric and Oceanic Technology, 16, 656–681, https://doi.org/10.1175/1520-0426(1999)016<0656:SBAIEO>2.0.CO;2, 1999.

Teuling, A. J., Taylor, C. M., Meirink, J. F., Melsen, L. A., Miralles, D. G., van Heerwaarden, C. C., Vautard, R., Stegehuis, A. I., Nabuurs, G.-J., and de Arellano, J. V.-G.: Observational evidence for cloud cover enhancement over western European forests, Nature Communications, 8, 14 065, https://doi.org/10.1038/ncomms14065, http://www.nature.com/articles/ncomms14065, 2017.

Theeuwes, N. E., Barlow, J. F., Teuling, A. J., Grimmond, C. S. B., and Kotthaus, S.: Persistent cloud cover over mega-cities linked to surface heat release, npj Climate and Atmospheric Science, 2, 15, https://doi.org/10.1038/s41612-019-0072-x, http://dx.doi.org/10.1038/s41612-019-0072-xhttp://www.nature.com/articles/s41612-019-0072-x, 2019.

Tøttrup, C.: EU-DEM Statistical Validation Report August 2014, European Environment Agency (EEA), p. 27, 2014.

5 Underwood, S. J. and Hansen, C.: Investigating urban clear islands in fog and low stratus clouds in the San Joaquin Valley of California, Physical Geography, 29, 442–454, https://doi.org/10.2747/0272-3646.29.5.442, https://www.tandfonline.com/doi/full/10.2747/0272-3646.29.5.442, 2008.

Varentsov, M., Wouters, H., Platonov, V., and Konstantinov, P.: Megacity-Induced Mesoclimatic Effects in the Lower Atmosphere: A Modeling Study for Multiple Summers over Moscow, Russia, Atmosphere, 9, 50, https://doi.org/10.3390/atmos9020050, http://www.mdpi.com/10 2073-4433/9/2/50, 2018.

Veefkind, J. P., Haan, J. F. D., Sneep, M., and Levelt, P. F.: Improvements to the OMI O2–O2 operational cloud algorithm and comparisons with ground-based radar–lidar observations, Atmos. Meas. Tech., 9, 6035–6049, https://doi.org/10.5194/amt-9-6035-2016, https://amt.copernicus.org/articles/9/6035/2016/, 2016.

Wærsted, E. G., Haeffelin, M., Steeneveld, G., and Dupont, J.: Understanding the dissipation of continental fog by analysing the 15 LWP budget using idealized LES and in situ observations, Quarterly Journal of the Royal Meteorological Society, 145, 784–804, https://doi.org/10.1002/qj.3465, https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3465, 2019.

Williams, A. P., Schwartz, R. E., Iacobellis, S., Seager, R., Cook, B. I., Still, C. J., Husak, G., and Michaelsen, J.: Urbanization causes increased cloud base height and decreased fog in coastal Southern California, Geophysical Research Letters, 42, 1527–1536, https://doi.org/10.1002/2015GL063266, https://onlinelibrary.wiley.com/doi/abs/10.1002/2015GL063266, 2015.

20 Yang, Y., Di Girolamo, L., and Mazzoni, D.: Selection of the automated thresholding algorithm for the Multi-angle Imaging SpectroRadiometer Radiometric Camera-by-Camera Cloud Mask over land, Remote Sensing of Environment, 107, 159–171, https://doi.org/10.1016/j.rse.2006.05.020, 2007.

Zhong, S., Qian, Y., Zhao, C., Leung, R., Wang, H., Yang, B., Fan, J., Yan, H., Yang, X. Q., and Liu, D.: Urbanization-induced urban heat island and aerosol effects on climate extremes in the Yangtze River Delta region of China, Atmospheric Chemistry and Physics, 17, 25 5439–5457, https://doi.org/10.5194/acp-17-5439-2017, https://www.atmos-chem-phys.net/17/5439/2017/, 2017.