CLOUD DETECTION FOR NIGHT-TIME PANCHROMATIC VISIBLE AND NEAR-INFRARED SATELLITE IMAGERY

L. Joachim[1], T. Storch[2,*]

[1] Institute for Photogrammetry (IFP), University of Stuttgart, Geschwister-Scholl-Str. 24D, 70174 Stuttgart, Germany -
lena.joachim@ifp.uni-stuttgart.de
[2] Earth Observation Center (EOC), German Aerospace Center (DLR), Münchener Str. 20, 82234 Weßling, Germany -
tobias.storch@dlr.de

Commission II, ICWG II/III

KEY WORDS: Cloud Detection; Day-Night-Band; Moon Illumination; Night-Time Satellite Imagery; Random Forest; Urban Areas

ABSTRACT:

Cloud detection for night-time panchromatic visible and near-infrared (VNIR) satellite imagery is typically performed based on synchronized observations in the thermal infrared (TIR). To be independent of TIR and to improve existing algorithms, we realize and analyze cloud detection based on VNIR only, here NPP/VIIRS/DNB observations. Using Random Forest for classifying cloud vs. clear and focusing on urban areas, we illustrate the importance of features describing a) the scattering by clouds especially over urban areas with their inhomogeneous light emissions and b) the normalized differences between Earth's surface and cloud albedo especially in presence of Moon illumination. The analyses substantiate the influences of a) the training site and scene selections and b) the consideration of single scene or multi-temporal scene features on the results for the test sites. As test sites, diverse urban areas and the challenging land covers ocean, desert, and snow are considered. Accuracies of up to 85% are achieved for urban test sites.

1. INTRODUCTION

The footprint of human activities on Earth is particularly visible at night, because human life and work at night usually involves artificial lighting (Levin et al., 2019). This is clearly visible in night-time satellite imagery in the visible and near-infrared (VNIR) spectral range from 0.4 µm to 1.1 µm. Such data contains wealth of information that is not so explicitly derivable from any other remote sensing product. On this basis, not only populated areas and its infrastructure are globally mapped and characterized and the expansion of urbanization is observed, but the illumination of a region also allows conclusions to be drawn about the behavior of the population (Miller et al., 2013). For example, socio-economic and economic factors influence the type and extent of artificial lighting, and different temporal lighting patterns indicate cultural differences. Analyses of night-time artificial lighting are also used to derive statements regarding its effects on the environment, for example in the form of light pollution affecting wildlife and human health, astronomical observations, and energy consumption.

So far, however, there are only a few satellite-based sensors that provide data for such analyses. Since 1976 DMSP/OLS (Imhoff et al., 1997) generates panchromatic VNIR imagery with a daily global coverage at 2.77 km spatial resolution, but without radiometric calibration. Since 2011 the Day-Night-Band (DNB) of NPP/VIIRS (Lee et al., 2006) generates panchromatic VNIR imagery with a daily global coverage at 0.75 km spatial resolution and with radiometric calibration. Beside these operational missions providing their global data free and open, some non-commercial cubesat missions improve for example the spatial resolution such as LJ1-01 (Jiang et al., 2018) to 0.13 km or the spectral resolution such as AC-5 (Pack et al.,

2017) to three bands. Furthermore, some commercial missions such as JL1-3B (Zheng et al., 2018) provide better spatial or spectral resolutions, but focus on day-time acquisitions. And similar to ISS astronaut photography (Kotarba et al., 2016) all these images are typically not quality-controlled and acquired for dedicated areas-of-interest only and are thereby not consistently global. To fill this gap, Elvidge et al. (2007) proposed an optimized nocturnal global satellite mission with high spatial and high spectral resolution for high-precise analyses of night-time artificial lighting.

The major prerequisite for analyses of properties of the Earth's surface based on optical data is reliable cloud detection to exclude those pixels from further considerations, where the surface is obscured by clouds. At night this is critical, because clouds are typically not visible in the VNIR spectral range due to the missing direct illumination by the Sun. There is only the limited reflected illumination by the Sun via the Moon and the limited artificial illumination of the Earth's surface. Therefore, operational global cloud detection algorithms used to generate cloud masks for existing night-time satellite imagery typically do not rely on the VNIR observations, but on the synchronized TIR (thermal infrared) observations. In this spectral range clouds are usually well-detectable as they are usually colder than the Earth's surface. And as radiation is emitted by the clouds or Earth's surface, it is also observable at night. Typically, observations in the long-wave infrared range from 8 µm to 14 µm (sensitive to low temperatures between 193 K and 362 K) supported by observations in the mid-wave infrared range from 3 µm to 5 µm (sensitive to high temperatures between 362 K and 966 K) are used for night-time cloud detection. Thus, for example, the cloud mask VCM for DNB at night is generated using several constant threshold value tests in

---

* Corresponding author

these two spectral ranges taking land cover types and other effects into account for the pixel-by-pixel differentiation between cloudy and clear skies (Hutchison et al., 2005). The enterprise cloud mask (ECM) (Heidinger et al., 2016) for DNB deviates from this approach by realizing a learning of the threshold values and by integrating tests in the VNIR spectral range at night, if the Moon is visible and illuminated. Furthermore, under such conditions some experimental approaches are based on the VNIR spectral range, but they relate to the detection of special cloud types (differences in the reflectivity of haze and low-lying stratus clouds) (Hu et al., 2017) or aerosols (contrast between illuminated and unilluminated pixels, which is reduced by the influence of aerosols) (Johnson et al., 2013) only.

We study how a feature-based algorithm for semantic segmentation of clouds has to be robustly realized using exclusively night-time observations in the VNIR spectral range (and information on the Moon) with two major objectives: First, to evaluate, if TIR bands are avoidable concerning cloud detection for night-time optical satellite missions focusing on VNIR bands. In the latter bands, artificial lighting is dominantly emitted. Adding TIR bands only for cloud detection impacts the entire mission. Second, to evaluate, if considering observations in the VNIR spectral range more intensely improves existing algorithms for night-time cloud detection. The focus is on the achieved quality of the global night-time cloud masks based on VNIR spectral range only, especially for areas with artificial lighting caused by human activities.

We use; as input: DNB products as they are globally available and well-calibrated nocturnal data in the VNIR spectral range with a daily revisit, which allows considering time series; as ground truth: ECM products as they are globally available and state-of-the-art high-accurate during night-time/day-time with 85%/94% accuracy over ocean and 88%/90% over land for clouds with an optical thickness ≥1. Chapter 2 details this data together with the considered sites, states the considered quality measures for the cloud masks, and establishes a model for the observations. Chapter 3 realizes the algorithm detailing the classification, considered features, and performed optimizations. Chapter 4 analyses the achieved results for different training and test sites and scenes (and radiometric sensitivities). Finally, Chapter 5 concludes with an overview of further investigations to be performed on the way to global cloud detection methods for night-time VNIR satellite imagery.

2. METHODOLOGY

2.1 DNB

DNB (Lee et al., 2006) is one of the 22 channels of the radiometer VIIRS on NPP, which is placed in a Sun-synchronous near-circular polar orbit at 824 km with a Local Time on Descending Node at 10:30 hours. VIIRS scans the Earth's surface according to the whiskbroom principle with a swath width of 3060 km, achieving twice daily (at noon and at night) coverage of the Earth. DNB products have a spatial resolution uniformly sampled to 0.75 km, a spectral range from 500 nm to 900 nm, and a dynamic range of seven orders of magnitude; by means of different amplification levels radiances from 3×10^{-9} to 2×10^{-2} $Wcm^{-2}sr^{-1}$ are detectable to allow acquiring images during day and night with high sensitivity. Here, calibrated and georeferenced Sensor Data Record (SDR) products (Cao et al., 2017) are used. They take a terrain model

into account and thereby achieve accuracies of about half a pixel in nadir and about one pixel in off-nadir. The direct georeferencing of urban areas covered by clouds is partially even more inaccurate due to a height error caused by the scattering of artificial lights by the cloud. We consider for each site and night, namely with Sun zenith angles larger than 96°, the scene with minimal Moon zenith angle.

2.2 ECM

ECM products (Heidinger et al., 2016) are cloud masks for VIIRS data generated by a Naïve Bayes approach based on various spectral observations and additional data like land cover maps. At night, the ECM algorithm is mainly using observations of the thermal infrared, especially if there is no lunar illumination. The ECM products contain cloud probabilities per pixel from which a binary cloud mask identifying cloudy and clear skies is derived. With ECM a conservative estimate of the cloudiness is obtained, namely the precision of cloudy skies is optimized. These products serve as ground truth, here. However, a closer look to the used ECM products exhibits that some of them contain rectangular artifacts, namely several rectangular areas have a significantly different cloud probability and class than the surrounding area as illustrated in Figure 1.
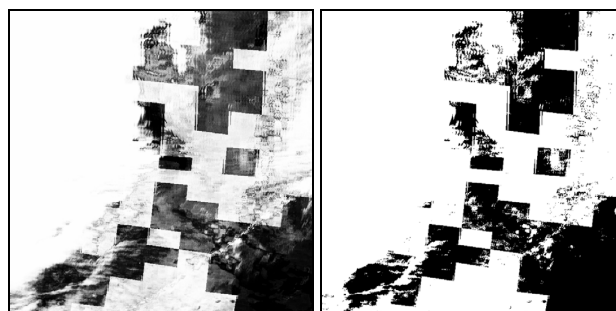

Figure 1. ECM product for 540 km × 540 km. left: cloud probabilities, right: cloud classes; white: cloud, black: clear

This classification is not corresponding to real cloud conditions and is not identified in the quality information. It is expected that these artefacts are caused by the data used to characterize the land cover types, which affects the selected threshold values. We do not consider the DNB product of a site, if the corresponding ECM product visibly has these artefacts.

2.3 Sites

The training and test sites selected for the considered cloud detection and its analysis cover some characteristic land cover types. The size of 150 × 150 pixels for each site corresponds to the swath width for a satellite mission suggested by Elvidge et al. (2007). Since the focus is on cloud detection for urban areas, the majority of the sites cover such regions in various forms as illustrated in Figure 2. To cover the specific challenges of night-time cloud detection, sites covering specific land cover types with limited artificial lighting are also considered as illustrated in Figure 3.

- Munich, Germany: European metropolis, which is densely populated and therefore strongly illuminated, but the hinterland is sparsely populated. It represents a variety of typical forms of settlement and lighting patterns.
- Stuttgart, Germany: similar to Munich, Germany

- Milan, Italy: European urban area, which is strongly illuminated also because of the dense population of the area around the metropolis itself.
- Brussels, Belgium: European urban area, which is heavily illuminated also because of the large cities around the metropolis itself and the illuminated motorways.
- New Orleans, LA, USA: USA area, which is irregularly illuminated because of considerable differences in land cover types such as lakes and rivers around and along the metropolis itself.
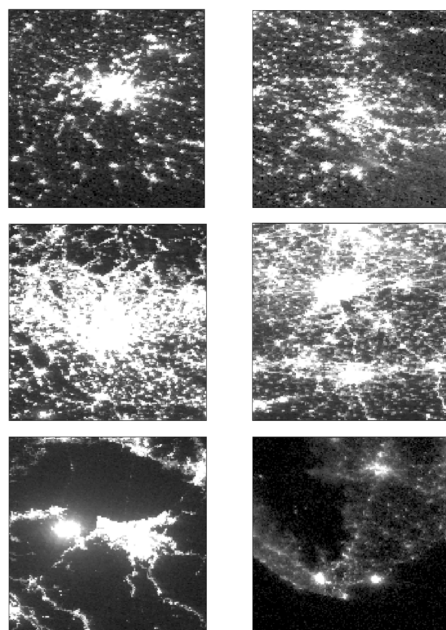

Figure 2. Cloud- & Moon-illumination-free ex. of Munich (top left), Stuttgart (top right), Milan (middle left), Brussels (middle right), New Orleans (bottom left) & Nagercoil (bottom right)
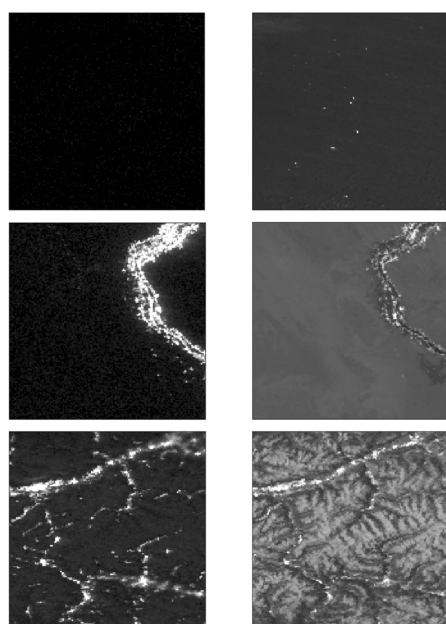

Figure 3. Cloud free ex. of Open Ocean (top left), Oil Platforms (top right), Desert (middle) & Snow (bottom); Moon-illumination-free (left) & Moon-illuminated (right)

- Nagercoil, India: Indian area, which covers different population densities around the metropolis itself and borders on the ocean.
- Open Ocean: Indian Ocean south of Nagercoil, which appears as dark area.
- Oil Platforms: Gulf of Mexico south of New Orleans, covering several deep-water offshore oil platforms, which are continuously illuminated and appear as fix punctual artificial lightings on dark areas.
- Desert: Sahara, which exhibits a much higher surface albedo than the Nile shore with its agricultural, building and water environments.
- Snow: Alps, which exhibits a much higher surface albedo especially in winter.

2.4 Quality

The semantic image segmentation separates the two classes cloudy and clear skies. The resulting structure of the confusion matrix describing the classification result of all pixels with the derived quantities Recall and Precision is illustrated in Table 1.

		Classification Result		
		cloudy	clear	Recall
Ground Truth (ECM)	cloudy	TP	FN	TP/ (FN+TP)
	clear	FP	TN	TN/ (FP+TN)
	Precision	TP/ (FP+TP)	TN/ (FN+TN)	

Table 1. Confusion matrix structure

To consider Precision and Recall together, they are used to form the harmonic mean $F_1=2\times Recall\times Precision/(Recall+Precision)$ for each class separately. Furthermore, the Overall Accuracy $OA=(TP+TN)/(FP+FN+TP+TN)$ and the Balanced Accuracy $BA=[TP/(FN+TP)+TN/(FP+TN)]/2$, which is used to make the assessment of the classifier less dependent on the class distribution present in the data, are considered (Ting, 2010).

2.5 Model

As basis for the interpretation of night-time satellite imagery in the VNIR spectral range, a global model for such observations, simplified to focus on the major effects concerning cloud detection, is introduced. It covers the

- Earth's surface artificial lighting (Elvidge et al., 2017) with its light emissions and Moon illuminance (Miller et al., 2009) – the only globally available source of nocturnal illumination (depending on the zenith angle and phase of the Moon) – with their influence on the appearance of the surface and clouds,
- Atmosphere and especially clouds – with their scattering, transmission and reflection depending on the cloud type and especially its albedo and optical thickness, and
- Different land cover types – with their different reflection properties in the images.

A radiative transfer model for these different elements is illustrated in Figure 4. The observed at-sensor or Top-Of-Atmosphere (TOA) radiance is the sum of all radiation reaching the sensor from different illumination sources via different paths. For example, no statement can be made for a single observed value as to whether it was caused by a dimly lit city in

cloudless visibility or a brightly lit city in cloudy visibility. In addition to these major effects, there are other effects that are neglected in this model, for example under certain conditions at night the refractions of light from the Sun, auroras, and lightning flashes influence the observed radiance.
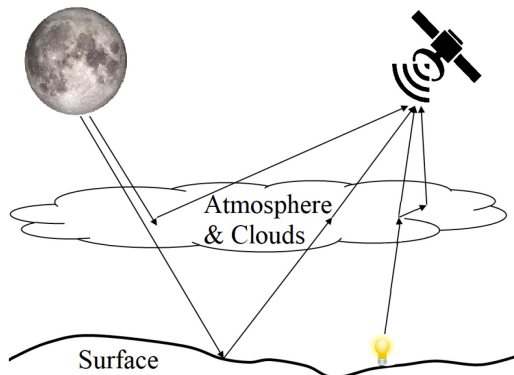

Figure 4. Model for night-time satellite imagery

3. ALGORITHM

3.1 Classification

With Random Forest (Breimann, 2001), a standard machine learning procedure is chosen for the classification of the pixels as cloudy or clear. It is a supervised and feature-based classification method based on the randomized learning of an ensemble of decision trees. Thus, it enables the integration of knowledge from the observation model in the feature design but does not solely rely on the model as it implements a learning of tests based on training data. An advantage is its robustness against noisy data. This is essential for our task, as our training data ECM is not a perfect ground truth. The hyperparameters of the Random Forest were estimated using random search (Bergstra, Bengio, 2012) and optimized using grid search. For our task, a Random Forest with 70 trees, unlimited tree depth and a minimum of 100 data points per leaf performed best. In the training phase, a randomly selected subset of features, here the square-root of the total number of features, is considered at each node. As classification result, the class for which the average of the predicted class probabilities over all trees becomes maximum is chosen. This also gives a good indication on the reliability of the assignment.

3.2 Features

The features shall describe the influences of clouds on the night-time images as characteristically as possible. Thus, features are needed to characterize the Earth's surface, cloud albedo and optical thickness as well as the effects caused by scattering by clouds. Since only observations in the VNIR spectral range and estimations of Moon illuminations shall be used, all characteristics have to be derivable from this limited amount of information. Unless otherwise stated, we consider a neighborhood of 25 pixels, namely a 5×5 window.

3.2.1 Single Scenes: The **cloud albedo and optical thickness** generally influence the measured radiance L_{DNB} of a pixel, which is therefore a possible feature. Since clouds are usually extensive, the alternative is to use the mean radiance μL_{DNB} in the neighborhood of the pixel. However, these observations are strongly influenced by the present Moon illumination L_m (Miller et al., 2009). Therefore, the feature

$\delta L_{norm}=(L_m-L_{DNB})/(L_m+L_{DNB})$, the normalized difference of the radiance of the moon and the observation, is used as a relative measure for the radiance of the pixel. If $L_m=0$, namely for invisible or new moon, and if $L_{DNB}=0$, δL_{norm} has a constant value of -1 or +1, respectively. Thus, also $\delta L=L_m-L_{DNB}$ is considered as a feature.

The effect of **scattering by clouds** influences the contrast and texture of the image (Elvidge et al., 2017), which is describable by features based on the neighborhood. The characteristic for the contrast or homogeneity of an image is the variance $\sigma^2 L_{DNB}=\sum_{n\in N}(L_{DNB,n}-\mu L_{DNB})^2/|N|$ describing the square deviation of the observed radiances $L_{DNB,n}$ from the mean value in the neighborhood N. The smaller the variance, the more homogeneous is the neighborhood. For example, with cloudless skies a small variance occurs over open oceans and a high variance occurs in artificially illuminated areas. The influence of clouds on the images is contradictory. In case of homogeneous surfaces and illumination by the Moon, clouds lead to increases in variance, because the scattering is more inhomogeneous, and in case of inhomogeneous surfaces such as artificially illuminated areas, clouds smoothen and thus reduce the variance. To describe this effect of clouds on artificially illuminated scenes, which are characterized by strong edges during clear skies, an edge filter such as the modified Laplace, represented by this filter mask, is considered:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

In contrast to the standard Laplace operator, this filter takes into account not only the horizontal and vertical second derivatives, but also the diagonal directions. The Laplace operator was selected instead of the also usual Sobel filter (first derivative), because it is also suitable for the detection of blobs, namely single brighter or darker pixels. These appear in night images of the considered resolution in the form of single strongly illuminated objects like offshore platforms. Additionally, features which describe the texture of the neighborhood are incorporated, namely some of the so-called Haralick features like Contrast and Energy (Haralick et al., 1973).

3.2.2 Multi-Temporal Scenes: As single scenes do only provide a limited amount of information, additional features describing temporal changes based on the daily time series of DNB images are considered. These features are based on comparisons of the considered scene with previous images (here, at most 30) of the time series and are calculated for every image of the time series. Since the cloud masks of all scenes are assumed to be unknown, there is no cloudless reference available as a basis for the comparison. Therefore, the mean value of a single scene feature f in the corresponding previous images can be considered as reference and $\Delta f(N_t)=f(N_t)-\sum_{1\leq i\leq 30}f(N_{t-i})/30$ can be used as feature (Jedlovec, 2009).

Furthermore, instead of comparing features, the images can be compared directly. As clouds induce many changes in the images, a cloudy image usually has a weak correlation with other images of its time series. Thus, the maximum correlation of the considered pixel and its neighborhood N_t with the corresponding neighborhoods N_{t-i} in the previous images, namely $\rho(N_t)=\max(\text{co-}\sigma(N_t,N_{t-i})/\sigma(N_t)\sigma(N_{t-i}) \mid 1\leq i\leq 30)$, where σ is the variance and co-σ the co-variance, is considered as another feature (Lyapustin et al., 2008). For these investigations all images of a site have to be mapped on the same grid, therefore a sufficiently accurate georeferencing is essential.

3.3 Optimization

Since the use of all proposed features does not necessarily lead to the best classification results, especially because some of them are very similar, the best feature combination is determined by a series of tests. For this task, the first 100 valid images of 2018 of Munich are used as training data and of Stuttgart as test data; the same dataset was used to define the hyperparameters of the Random Forest. To reduce the search space, the best combination of single scene features is determined first and afterwards they are added or replaced by the multi-temporal features. To further reduce the number of combinations, single scene features were grouped based on their correlation. Then all possible combinations of features were tested automatically in several runs for classification with the defined Random Forest, using only one feature from each group at a time. In several runs it is tested to completely omit single groups as well. The feature importance estimated during training is considered.

Overall, Contrast (0) and δL_{norm} (1) are always among the most important features. Along with μL_{DNB} (2) concerning cloud albedo and optical thickness as well as $\sigma^2 L_{DNB}$ (3), Energy (4), and Laplace (5) concerning scattering by clouds, they prove to be the best combination. Here, Energy and Laplace only lead to minor improvements.

Based on this combination, multi-temporal features are included by adding them to the single scene feature vector or replacing their single scene equivalents. The feature ρ is considered as an optional feature. The features ρ (6) and $\Delta\sigma^2 L_{DNB}$ (7) as the only temporal equivalent significantly improve the classification results; $\Delta\sigma^2 L_{DNB}$ either by supplementing or replacing $\sigma^2 L_{DNB}$. The improvements by adding one feature after another are illustrated in Figure 5. As cloud conservative ECM products are used, the method itself is as well cloud conservative.
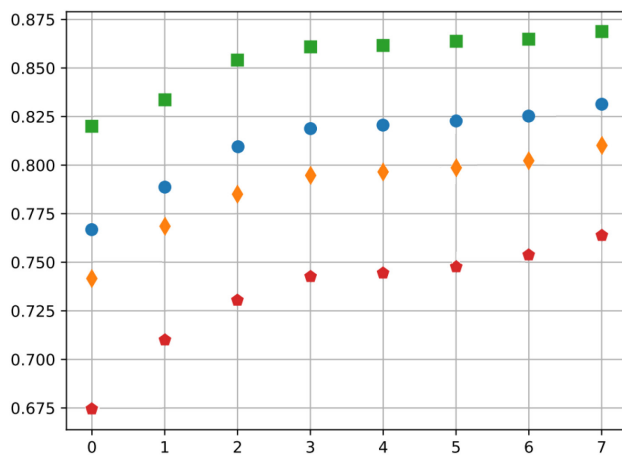

Figure 5. OA (blue), BA (orange), F_1 cloud (green), F_1 clear (red); number i indicates use of set of features $\{0,…,i-1,i\}$.

4. RESULTS

We consider the first 100 valid images of 2018 of Munich as training data as well as the first 30 valid images of 2018 of Stuttgart, Brussels, New Orleans for populated areas and Oil Platforms, Desert, Snow for specific land cover types as test data. Thus, the different conditions concerning Moon illumination are covered. As illustrated in Figure 6, the defined algorithm achieves an OA of 70% and a BA of 67% for all test

sites as detailed in Table 2. F_1 of 0.77 for cloudy and 0.57 for clear skies are obtained. The results of the cloud class are better than those of the clear class which is mainly caused by the high recall of the cloud class, whereas the precisions are similar. This corresponds to the expectations, because the cloud class is intentionally optimized.
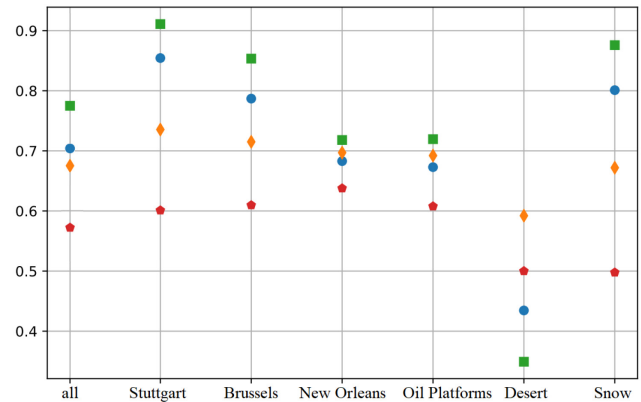

Figure 6. OA (blue), BA (orange), F_1 cloud (green), F_1 clear (red) for all test sites and single test sites

		Classification Result		
		cloudy	clear	Recall
Ground Truth	cloudy	2,053,770	324,269	0.86
	clear	868,699	803,262	0.48
	Precision	0.70	0.71	

Table 2. Confusion matrix for all test sites

4.1 Test Sites

Figure 6 illustrates the performances of the classifier for the different test sites. Overall, the three urban test sites perform best in terms of BA, which is expected due to the exclusive training with an urban site and the focus in the realization of the algorithm on such data. By far the worst results are obtained for Desert. Analyzing the results separated by F_1 for cloudy and clear skies, there are even greater differences between the considered test sites, which is an effect of the different class distributions in the data. For example, Stuttgart and Snow have small ratios of clear sky pixels, an advantage for the cloud conservative algorithm to achieve a high F_1 for cloudy skies compared to, for example, New Orleans with a balanced class distribution. For Stuttgart and Brussels the results are better than the results for all test sites, especially with F_1 of at least 0.85 for cloudy skies. Thus, for test sites with strong light emissions, which are similar to the training site, the method works well.

For New Orleans and Oil Platforms similar results are obtained. Their F_1 of 0.7 for cloudy skies is worse than for Stuttgart and Brussels, although similar results are expected for New Orleans. However, the detailed performance for New Orleans and Oil Platforms is quite different. While the classifier still tends to classify too many pixels as cloudy skies (high recall with low precision of the cloud class) for New Orleans, for Oil Platforms the opposite occurs. Clouds over oceans are invisible without illumination of the Moon. Also, the blurring effects of light emissions from the surface by clouds do not occur. Therefore, cloudy and clear skies are not distinguishable for the classifier. This is confirmed as the best results for Oil Platforms are obtained with Moon illumination.

For Desert the interaction of Recall and Precision is essential. Although 84% of the clouds are detected, only 22% of the pixels in the resulting cloudy class are correctly classified. For Snow both Recall and Precision of the cloudy class are good.

4.2 Test Scenes

Not only between the different test sites, but also between the scenes of one test site there are large differences in the quality of cloud detection. Therefore, good and bad results for some test sites are analyzed to derive the reasons for these variations. For Stuttgart, the overall best classification is achieved and the best results – similar for Brussels and New Orleans – are obtained for heavily cloud covered images, where the structure of the metropolis is blurred, as illustrated in the first two rows of Figure 7. This corresponds to properties characterized by some of the features. Considering the worst results for Stuttgart, they sometimes agree better with the visual impression of the images than the corresponding ECM product, as illustrated in Figure 7. This is similar for other images and other test sites. For example, for the second worst result (row 3) a cloud is clearly visible at the lower right edge of the DNB product, which was correctly classified but is not part of the ECM product. However, for thin clouds leading to a limited blurring as for the worst result (row 4), the decisions of the classification are uncertain.

For Desert the in total worst result is achieved. It explains the behavior concerning illumination by the Moon. As illustrated in Figure 7, in the worst results (rows 7 and 8) all pixels are classified as clouds with exception of the artificially illuminated bank of the Nile, although the images are not cloud covered. For the second best result for Desert (row 6) this is less often the case. Between these cloudless situations the appearance of the surface considerably differs due to different illuminations by the Moon and the large surface albedo of sand. For the worst results, the scenes are illuminated by the moon, for the best not. When illuminated by the Moon, the vast desert resembles clouds which leads to the described classification errors.

Similar, in the worst results for Snow artificially illuminated surface structures are correctly classified as clear, but the Moon-illuminated and snow-covered mountains are incorrectly classified as cloudy, because they are confused with clouds. However, also the opposite effect occurs, where areas with high darkness and low inhomogeneity are incorrectly classified as cloudy. This is explained by the use of Munich as training site. For clear skies there are no extended dark areas in the training set, but rather a high inhomogeneity because intensities of neighboring pixels are often highly different due to highly variable artificial lighting in urban structures.

In general, the influence of the illumination by the Moon on the results can be analyzed by considering the relation between OA and estimated Moon illumination L_m for each scene. For all test sites except Desert, an OA of less than 40% only occurs in the absence of Moon illumination. Thus, in general Moon illumination mostly has a positive effect on the results. Considering the test sites individually, a dependence of OA and Moon illumination is recognizable especially for Oil Platforms and Desert, which confirms the observations described previously. While an increase of Moon illumination has a positive effect on the OA for scenes of the Oil Platform site, it has the opposite effect on Desert.
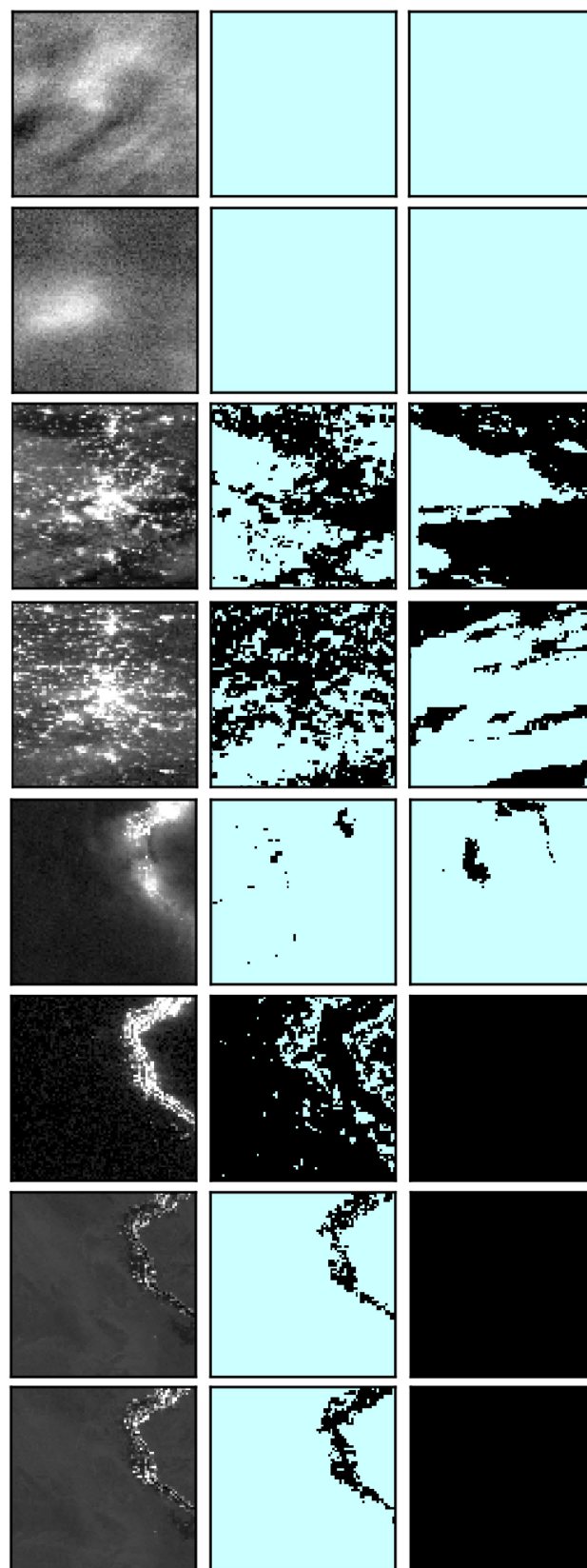

Figure 7. Best (rows 1, 2, 5, 6) and worst (rows 3, 4, 7, 8) scenes of Stuttgart (rows 1, 2, 3, 4) and Desert (rows 5, 6, 7, 8); left: DNB, right: ECM, middle: result of classification (blue: cloud, black: clear)

4.3 Single Scenes vs. Multi-Temporal Scenes

Let us compare the cloud classification results not using or using the multi-temporal beside the single scene features. Overall, additionally considering multi-temporal features led to minor improvements of the cloud detection for all test sites. However, an improvement for F_1 for clear skies of at least 0.04 is achieved only for Oil Platforms and Snow and for F_1 for cloudy skies by 0.09 for Oil Platforms only. All other improvements are below 0.02. The difference between these and the urban test sites, where the multi-temporal features have limited influences, is the significant difference in the presence of artificial lighting. Because of the higher inhomogeneity of lighting in urban areas, small inaccuracies in the geolocation lead to large deviations in the multi-temporal features. In this case the comparisons of time series are no longer valid. Thus, for urban areas an accurate co-registration is more important for the success of temporal considerations than for unpopulated areas with lower inhomogeneity. For example, consider the differences in the co-registration of Stuttgart in the second worst result, which is partly free of clouds, and in the second best result, which is covered by clouds, as illustrated in Figure 7. The cloud virtually lifts the city from surface to cloud elevation and shifts the geolocation of the city, especially depending on the tilting angle, terrain, and cloud elevation. Here, these effects result in the limited improvements of the classification for urban areas using multi-temporal features.

4.4 Training Data

Let us consider the choice of training data. The classifier was trained using only one urban area, namely 100 scenes of Munich. However, the aim for accurate classifications is to have a sufficient data set representative for all situations. We therefore analyze how the diversity or type of the training sites and the number of training scenes affects the quality of the cloud detection results. Figure 8 illustrates the BA of the semantic segmentation of the six test sites using different training data sets. As an extreme situation, the exclusive usage of the first 100 valid images of 2018 of Open Ocean (0_{100}), an area without artificial lighting, as training data leads to bad classification results as expected. The only exceptions are Oil Platforms, which also mainly consists of an ocean. But for Oil Platforms good results are also obtained, if only urban areas are used for training. Thus, the selection of the training sites and scenes is essential.

4.4.1   Adding Training Sites: Let us consider training the classifier not only with Munich (1_n), but with two further training sites, namely adding Milan and Nagercoil (2_n) and further adding Open Ocean (3_n), where the first valid images of 2018 are used. These sets are selected such that the diversity of types of land cover is increasing. With regard to the diversity and type of training sites, there is a tendency for better results achieved with less different sites considering an in total equivalent number of training scenes. In this case, the time series per site gets shorter by adding sites, which makes the multi-temporal features less meaningful. Furthermore, not only the land cover types represented in the training data have influences on the results, but also the class distribution. For example, if an image covered by clouds has to be classified, having only few examples of cloudiness in the training data results in lower robustness, as the cloudy class is only influenced by few points; independent of the land cover type. Therefore, namely because of few examples of cloudiness in the training data compared to the test data, adding Open Ocean

to the training sites worsens the results for almost all test sites – not necessarily because the area is not artificially illuminated.

4.4.2   Adding Training Scenes: Let us consider the classifier is trained with 200 and 300 scenes in addition to 100 scenes for the different selected sets of training sites. As expected, the quality of the results increases with the number of considered images. Furthermore, the quality of the results for the clear class is typically more influenced by the variation of the training data than that of the cloudy class. And it is advantageous if the training sites are similar to the test sites – both in the land cover type and in class distribution. For example, consider training data set (1) with 100 scenes and (2) with 300 scenes, which fully contains (1). The latter leads to the expected improvements for all test sites. However, the extent of these effects depends on the test site. For example, for Stuttgart the results partially even worsen, when the number of training scenes is increased. Therefore, a careful selection of the training data is essential. To obtain optimized results, the training sites have to be adapted to the characteristics of the test sites to which the method is applied. Further improvements are achieved by increasing the number of scenes.
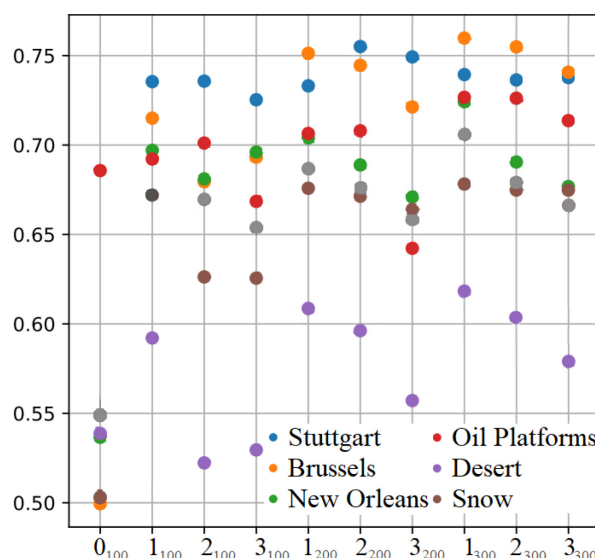

Figure 8. BA for all test sites (grey) and single test sites; number i_n indicates n training scenes of training sites (i) are considered in equal shares

4.5 Radiometric Sensitivity

Finally, the influence of the quality of the night-time images is investigated. Therefore, changing the radiometric sensitivity from the specified 3×10^{-9} to 3×10^{-8} and 3×10^{-7} $Wcm^{-2}sr^{-1}$ is simulated. The BA for all test sites changes from 67% to 57% and 52% as expected.

5.   CONCLUSIONS

We realized and analyzed an algorithm based on Random Forest for cloud detection in night-time panchromatic visible and near-infrared satellite imagery focusing on urban areas. The classification was trained and tested on DNB imagery using ECM as reference, considered Moon illumination and especially no information based on thermal infrared data. Features on contrast – describing the scattering by clouds especially for the inhomogeneous artificial Earth's surface light emissions – and

normalized difference of Moon illumination and measurement – describing the Earth's surface compared to the cloud albedo and optical thickness especially in presence of Moon illumination – were of major relevance. Overall accuracies of up to 85% for urban areas were obtained. These investigations lead to improvements of methods for existing missions as well as support realizations and analyses of algorithms for future missions. To operationalize the method, more training and test sites and scenes have to be considered as well as the features on single and multi-temporal scenes have to be detailed and refined. Instead of binary cloud masks also cloud probability maps and more detailed information such as cloud albedo and optical thickness shall be derived to support next processing chains. Furthermore, an evaluation on the consideration of further external operational services such as on land cover types based on global, for example, static maps or Sentinel products, or on cloud information based on geostationary weather satellites has to be performed. Concerning future missions for night-time VNIR satellite imagery, advantages of multi-spectral data for deriving atmospheric and cloud parameters – also with dedicated spectral bands – have to be investigated.

REFERENCES

Bergstra, J., Bengio Y., 2012. Random Search for Hyper-Parameter Optimization. *The Journal of Machine Learning Research,* 13, 281–305.

Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5-32.

Cao, C., Xiong, X., Wolfe, R., DeLuccia, F., Liu, Q., Blonski, S., Lin, G., Nishihama, M., Pogorzala, D., Oudrari, H., Hillger, D., 2017. Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Record (SDR) User's Guide. Techn. Report Version 1.3. U.S. Department of Commerce, National Oceanic and Atmospheric Administration. *https://ncc.nesdis.noaa.gov/ documents/documentation/viirs-users-guide-tech-report-142a-v1.3.pdf* (3 April 2020).

Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C., Ghosh, T., 2017. VIIRS night-time lights. *International Journal of Remote Sensing*, 38(21), 5860-5879.

Elvidge, C.D., Cinzano, P., Pettit, D.R., Arvesen, J., Sutton, P., Small, C., Nemani, R. Longcore, T. Rich, C., Safran, J., Ebener, S., 2007. The Nightsat mission concept. *International Journal of Remote Sensing*, 28(12), 2645-2670.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(3), 610-621.

Heidinger, A., Botambekov, D., Walther, A., 2016. A Naive Bayesian Cloud Mask Delivered to NOAA Enterprise, Algorithm Theoretical Basis Document. Techn. Report Version 1.1. NOAA NESDIS Center for Satellite Applications und Research.*https://www.star.nesdis.noaa.gov/jpss/documents/ATB D/ATBD_EPS_Cloud_Mask_v1.1.pdf* (3 April 2020).

Hu, S., Ma, S., Yan, W., Jiang, J., Huang, Y., 2017. A new multichannel threshold algorithm based on radiative transfer characteristics for detecting fog/low stratus using night-time NPP/VIIRS data. *International Journal of Remote Sensing*, 38(21), 5919-5933.

Hutchison, K.D., Roskovensky, J.K., Jackson, J.M., Heidinger, A.K., Kopp, T.J., Pavolonis, M.J., Frey, R., 2005. Automated cloud detection and classification of data collected by the Visible Infrared Imager Radiometer Suite (VIIRS). *International Journal of Remote Sensing*, 26(21), 4681-4706.

Imhoff, M.L., Lawrence, W.T., Stutzer, D.C., Elvidge, C.D., 1997. A technique for using composite DMSP/OLS "City Lights" satellite data to map urban area. *Remote Sensing of Environment*, 61, 361-370.

Jedlovec, G., 2009. Automated Detection of Clouds in Satellite Imagery. In: *Advances in Geoscience and Remote Sensing*, InTech, Rijeka, Croatia, 303-3016.

Jiang, W., He, G., Long, T., Guo, H., Yin, R., Leng, W., Liu, H., Wang, G., 2018. Potentiality of Using Luojia 1-01 Nighttime Light Imagery to Investigate Artificial Light Pollution. *Sensors*, 18(9), 2900.

Johnson, R.S., Zhang, J., Hyer, E.J., Miller, S.D., Reid, J.S., 2013. Preliminary investigations toward nighttime aerosol optical depth retrievals from the VIIRS Day/Night Band. *Atmospheric Measurement Techniques*, 6(5), 1245-1255.

Kotarba, A.Z., Aleksandrowicz, S., 2016. Impervious surface detection with nighttime photography from the International Space Station. *Remote Sensing of Environment*, 176, 295-307.

Lee, T.E., Miller, S.D., Turk, F.J., Schueler, C., Julian, R., Deyo, S., Dills, P., Wang, S., 2006. The NPOESS VIIRS Day/Night Visible Sensor. *Bulletin of the American Meteorological Society*, 87(2), 191-200.

Levin, N., Kyba, C., Zhang, Q., 2019. Remote Sensing of Night Lights–Beyond DMSP. *Remote Sensing*, 11(12), 1472.

Lyapustin, A., Wang, Y.; Frey, R., 2008. An automatic cloud mask algorithm based on time series of MODIS measurements. *Journal of Geophysical Research*, 113(D16), 1-15.

Miller, S., Straka, W., Mills, S., Elvidge, C., Lee, T., Solbrig, J., Walther, A., Heidinger, A., Weiss, S., 2013. Illuminating the Capabilities of the Suomi National Polar-Orbiting Partnership (NPP) Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band. *Remote Sensing*, 5(12), 6717-6766.

Miller, S.D., Turner, R.E., 2009. A Dynamic Lunar Spectral Irradiance Data Set for NPOESS/VIIRS Day/Night Band Nighttime Environmental Applications. *IEEE Transactions on Geoscience and Remote Sensing*, 47(7), 2316-2329.

Pack, D., Hardy, B., Longcore, T., 2017. Studying the Earth at Night from CubeSats. In: *Proc. 31st Annual AIAA/USU Conference on Small Satellites*, Logan, UT, USA, 1-10.

Ting, K.M., 2010, Confusion Matrix/Precision and Recall. In: *Encyclopedia of Machine Learning*, Springer, Boston, MA, USA, 209/781.

Zheng, Q., Weng, Q., Huang, L., Wang, K., Deng, J., Jiang, R., Ye, Z., Gan, M., 2018. A new source of multi-spectral high spatial resolution night-time light imagery—JL1-3B. *Remote Sensing of Environment*, 215, 300-312.