UKIS-CSMASK: A PYTHON PACKAGE FOR MULTI-SENSOR CLOUD AND CLOUD SHADOW SEGMENTATION

M. Wieland[1,*], F. Fichtner[1], S. Martinis[1]

[1] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Münchener Straße 20, 82234 Weßling-Oberpfaffenhofen, Germany - (marc.wieland, florian.fichtner, sandro.martinis)@dlr.de

KEY WORDS: Semantic segmentation, Cloud, Cloud shadow, Sentinel-2, Landsat, Python.

ABSTRACT:

Cloud and cloud shadow segmentation is a crucial pre-processing step for any application that uses multi-spectral satellite images. In particular, time-critical disaster applications, require accurate and immediate cloud and cloud shadow masks while being able to adapt to possibly large variations caused by different sensor characteristics, scene properties or atmospheric conditions. This study introduces the newly developed open-source Python package *ukis-csmask* for cloud and cloud shadow segmentation in multi-spectral satellite images. Segmentation with *ukis-csmask* is performed with a pre-trained Convolutional Neural Network based on a U-Net architecture. It works directly on Level-1C data, eliminating the need for prior atmospheric correction. Images need to be in top of atmosphere reflectance and include at least the Blue, Green, Red, NIR, SWIR1 and SWIR2 spectral bands. We provide a performance evaluation on a recent benchmark dataset for cloud and cloud shadow segmentation and proof the generalization ability of our method across multiple satellites (Landsat-5, Landsat-7, Landsat-8, Landsat-9 and Sentinel-2). We also show the influence of augmentation and image bands on the segmentation performance and compare it to the widely used Fmask algorithm and a Random Forest classifier. Compared to previous work in this direction, our study focuses on multi-sensor generalization ability, simplicity and efficiency and provides a ready-to-use software package that has been thoroughly tested.

1. INTRODUCTION

Cloud and cloud shadow segmentation is a crucial pre-processing step for any application that uses multi-spectral satellite images. In particular, time-critical disaster applications, require accurate and immediate cloud and cloud shadow masks while being able to adapt to possibly large variations caused by different sensor characteristics, scene properties or atmospheric conditions.

Significant work has been undertaken to detect and segment clouds and cloud shadows in multi-spectral satellite images (Skakun et al., 2022). Zhu et al. (2015) introduce the Function of mask (Fmask) algorithm that uses rules based on physical cloud properties to identify potential cloud pixels from temperature, spectral variability, and brightness in Landsat and Sentinel-2 data. As part of the Sen2Cor Level-2A processor for Sentinel-2 a scene classification map that includes cloud and cloud shadow classes is computed (Main-Knorn et al., 2017). The algorithm is based on a series of threshold tests on the top-of-atmosphere reflectance from spectral bands and spectral band indices. The MAJA processor for atmospheric correction is based on complex rule-sets which exploit time-series of Sentinel-2 data to detect clouds and cloud shadows. It is largely based on the assumption that surface reflectance in the absence of clouds are stable with time while the presence of a cloud or a cloud shadow introduces a quick variation of the reflectance (Hagolle et al., 2010). Hollstein et al. (2016) test Classical Bayes, Decision Trees, Support Vector Machine and Stochastic Gradient Descent classifiers to detect clouds and cloud shadows amongst other classes in Sentinel-2 images.

Convolutional Neural Networks (CNNs) gradually appear in recent studies on cloud segmentation (Chen et al., 2018; Francis et al., 2019; Domnich et al., 2021) and show superior accuracy, generalization ability and inference speed compared to rule-based and classical machine learning approaches. Zhaoxiang et al. (2018) use wavelet compressed images with a CNN based on the U-Net architecture for cloud detection on-board small satellites. Ozkan et al. (2018) adapt a deep pyramid network to produce cloud masks from satellite images with noisy annotations. Jeppesen et al. (2019) introduce the Remote Sensing Network (RS-Net), a deep learning model for detection of clouds in Landsat-8 images that is based on the U-Net architecture. López-Puigdollers et al. (2021) carry out an extensive benchmarking study of deep learning methods for cloud detection in Landsat-8 and Sentinel-2 images. Mohajerani and Saeedi (2021) introduce Cloud-Net+ that is trained with a filtered Jaccard Loss and under consideration of a novel sunlight direction-aware data augmentation. They test their method on four Landsat-8 datasets, including their own 35-/95-cloud datasets, which consist of image subsets with four spectral bands ("Blue", "Green", "Red", "NIR") and corresponding cloud annotations. Similar to the L8-Biome (Scaramuzza et al., 2016) dataset, cloud shadows are not consistently annotated. Despite the promising results for cloud detection, only few studies and datasets specifically consider cloud shadows. Jiao et al. (2020) present a refined U-Net which extracts coarse cloud and cloud shadow regions in image patches and then refines their boundaries using a dense conditional random field. Hughes and Kennedy (2019) use a pretrained CNN with VGG16 encoder to separate clouds, cloud shadows, water and snow in Landsat 8 images. They train and test on the freely available Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) dataset (USGS, 2016). A recently published dataset that considers clouds and cloud shadows with a focus on Sentinel-2 is introduced by Francis et al. (2020). Their Sentinel-2 Cloud Mask Catalogue

---

* Corresponding author

consists of 513 subscenes with semi-automatically derived annotations that were randomly sampled from the 2018 Level-1C Sentinel-2 archive.

Besides the growing number of studies and datasets that target the use of CNNs, we can observe a general lack of easy to use, simple and fast tools for reliable cloud and cloud shadow segmentation across different sensors. Besides the complex rule-based solutions provided by Fmask, Sen2Cor and MAJA, none of the aforementioned studies that use CNNs provide their solutions as software packages and only few studies release their experimental code (e.g., Jeppesen et al., 2019). The s2cloudless algorithm is to the best of our knowledge the only machine learning approach that has been published as a ready-to-use and open-source software package in this context (https://github.com/sentinel-hub/sentinel2-cloud-detector). It uses gradient boosting trees for cloud classification on a single Sentinel-2 scene. Cloud shadows are not considered and transferability to other sensors is not specifically addressed.

Based on the above mentioned observations, we present the open-source Python package "UKIS Cloud Shadow MASK" (*ukis-csmask*) that masks clouds and cloud shadows in Landsat-5, Landsat-7, Landsat-8, Landsat-9 and Sentinel-2 images (https://github.com/dlr-eoc/ukis-csmask). Its objective is to provide a fast and simple to use tool that is able to mask cloud and cloud shadow pixels in single date images from a large variety of multi-spectral satellite sensors without the need for calibration or manual interference. This study builds up on the work presented in (Wieland et al., 2019) and focusses on the newly developed Python package *ukis-csmask*, a performance evaluation on a recent benchmark dataset and a transferability experiment to images of the Landsat-9 satellite.

2. DATA

In this study, we use three datasets for training, validation and testing of our method, namely the SPARCS dataset for Landsat-8 (USGS, 2016), a custom reference dataset for Landsat and Sentinel-2 (Wieland et al., 2019) and the Sentinel-2 Cloud Mask Catalogue (Francis et al., 2020) (Figure 1).

2.1 SPARCS Landsat-8 dataset

The freely available Spatial Procedures for Automated Removal of Cloud and Shadow (SPARCS) dataset (USGS, 2016) contains 80 samples of globally sampled Landsat-8 Level-1 data and corresponding manually annotated and quality checked thematic masks. The masks contain the classes cloud, cloud shadow, snow / ice, water, flooded and land. Each sample covers a subset of a Landsat-8 scene with 1,000 x 1,000 pixels (resampled to 30 m). Masks and are provided in PNG format and images are delivered with 11 spectral bands in GeoTIFF format with associated quality assessment band and metadata. In this study, we use only spectral bands that are available across different satellite sensors to ensure a high degree of transferability of the trained models. Specifically, we use Landsat-8 bands 2 (Red), 3 (Green), 4 (Blue), 5 (NIR), 6 (SWIR1) and 7 (SWIR2). We transform Digital Numbers (DN) to Top of Atmosphere (TOA) reflectance using the conversion factors provided in the scene's metadata and convert the masks into GeoTIFF format. We carry out two sets of reclassifications of the thematic masks: 1.) we derive a dataset with five classes (clouds, cloud shadows, water, snow / ice, land) by combining classes water and flooded; 2.) we derive a dataset with three

classes (clouds, cloud shadows, clear sky) by merging the land, water and snow / ice classes into the clear sky pixel class.
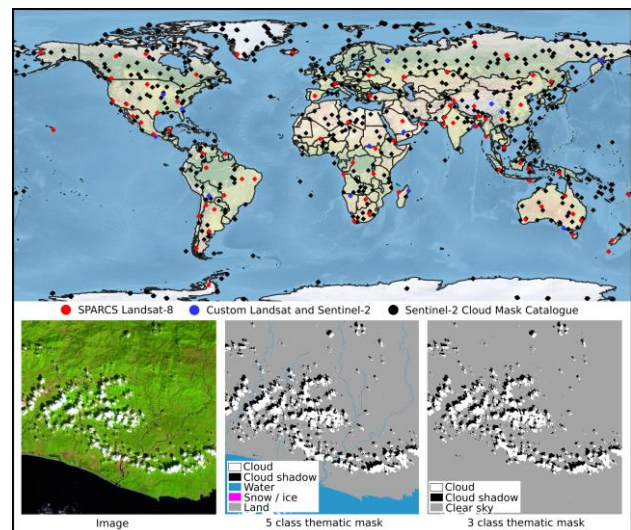

Figure 1. Spatial distribution of datasets used in this study. Samples consist of a satellite image and thematic masks.

2.2 Custom reference dataset for Landsat and Sentinel-2

We compile a custom reference dataset from Landsat-8, Landsat-7, Landsat-5 and Sentinel-2 images. A stratified random sample of the Landsat Worldwide Reference System (WRS2) path/row locations with a minimum distance constraint of 370 km (twice the swath width of a Landsat scene) is used to identify 14 globally distributed sample locations. For each location we acquire images of all four satellites, while keeping acquisition times across satellites as close as possible to each other to strengthen the significance of a cross-sensor comparison. Acquisition times across sample locations cover different seasons and the minimum cloud-cover percentage is set to 5 % in order to guarantee a certain degree of cloud-cover per sample. We create 56 1,024 x 1,024 pixels subscenes (resampled to 30 m), stack the Red, Green, Blue, NIR, SWIR1 and SWIR2 image bands together and convert DN to TOA reflectance. Thematic masks are generated by manual visual image interpretation and annotation. Several iterations of quality checks and adjustments by multiple operators are carried out to refine the masks. Like for the SPARCS Landsat-8 dataset, we compile two sets of reclassifications of the thematic masks: 1.) five classes (clouds, cloud shadows, water, snow / ice, land) and 2.) three classes (clouds, cloud shadows, clear sky).

2.3 Sentinel-2 Cloud Mask Catalogue

The freely available Sentinel-2 Cloud Mask Catalogue dataset consists of 513 1,022 x 1,022 pixels subscenes (resampled to 20 m) of Sentinel-2 Level-1C images that were sampled randomly from the 2018 archive. Thematic masks cover three classes (clouds, cloud shadows, clear sky), were annotated with a semi-automatic procedure and quality checked by several operators (Francis et al., 2020). In addition to the thematic annotation masks, metadata is provided with information about properties of each subscene, such as surface type, cloud type, cloud height, cloud thickness and cloud extent. 89 subscenes are reported to have no cloud shadow annotations due to ambiguous

boundaries and overlaps with terrain shadows. We therefore discarded these samples from our experiments.

3. METHOD

Cloud and cloud shadow segmentation with *ukis-csmask* is performed with a pre-trained CNN based on the U-Net architecture. U-Net is capable to learn from small datasets while achieving state-of-the-art results on semantic segmentation tasks. Weights are initialized randomly as described in (He et al., 2015). During training, we optimize the weights using an adaptive moment estimation (Adam) algorithm (Kingma and Lei, 2015) with default hyper-parameters $l = 0.001$, $\beta1 = 0.9$ and $\beta2 = 0.999$. Even though Adam already computes adaptive learning rates to reduce the impact of tuning the hyper-parameters on convergence, it has been shown that additional learning rate adaptation can improve convergence speed and accuracy (Smith, 2017). Therefore, we use a step decay scheduler to further adapt the learning rate during training. As loss function we use weighted cross-entropy loss. To account for class imbalance, we use the inverse occurrence of a class as a weight to penalize the class specific loss and therefore prevent the model from focusing too much on the majority class. We train the network in batches of 10 until convergence and track overall accuracy, Cohen's Kappa coefficient, micro-averaged F1 score and Intersection over Union (IoU) metrics.

ukis-csmask works directly on Level-1C data, eliminating the need for prior atmospheric correction. Images need to be in top of atmosphere reflectance and include at least the Blue, Green, Red, NIR, SWIR1 and SWIR2 spectral bands. The final model that is being deployed by *ukis-csmask* has been trained on the SPARCS dataset (USGS, 2016) and has been refined by further training on a custom reference dataset of images from Landsat-8, Landsat-7, Landsat-5 and Sentinel-2, that have been sampled globally under consideration of land-cover, seasonality and type of cloud-cover (Wieland et al., 2019). The training dataset is augmented with random contrast and brightness. Factors are randomly applied within predefined ranges and all augmentations are applied with equal probability. Due to the spatial resolution of the input images and the natural variety of shapes of the classes of interest, we did not consider geometrical augmentations. We standardize the input image feature space to zero mean and unit variance with mean and standard deviation being computed on the training dataset and applied to the validation and testing datasets. The training set is shuffled once between every training epoch.

Since higher prediction errors towards the image borders may be observed when using CNNs, *ukis-csmask* expands the input image with mirror-padding before inference. The expanded image is split it into overlapping tiles for prediction and the prediction tiles are blended with a tapered cosine window function and un-padded to reconstruct the input image's x-y-shape (Wieland and Martinis, 2019). We use Tensorflow as deep learning framework and converted the final model to the Open Neural Network Exchange (ONNX) format for deployment in *ukis-csmask*. To run inference with GPU support, CUDA runtime libraries are required to be installed on the system.

As a benchmark method, we train a Random Forest classifier on the same training datasets as our method. We use the Scikit-learn ("Scikit-learn.," 2022) implementation, which fits a number of C4.5 decision tree classifiers (Breiman, 2001) on sub-samples of the training data and uses averaging to control over-fitting and improve the accuracy of the predictions. Additionally, we apply the rule-based Fmask algorithm (Zhu et al., 2015) to the test images using the Python Fmask package ("Python Fmask," 2022). In contrast to the machine learning methods, the rule-sets that underlie Fmask are sensor specific and make use of all available spectral bands including thermal and cirrus bands if available.

All experiments are carried out on a standard desktop PC with Intel Xeon W-2235 CPU @ 3.80 GHz, 6 cores, 60 GB RAM and an NVIDIA Quadro RTX 4000 GPU.

4. RESULTS

In the following we present three sets of experiments. At first, we evaluate the influence of augmentation and input bands on the segmentation performance and compare the results with Fmask and a Random Forest classifier. For these experiments we use the SPARCS dataset for training and the custom multi-sensor reference dataset for testing (Section 4.1). Based on the outcomes of these experiments we select the best-performing setup and train a final model for use in *ukis-csmask* using both SPARCS and our custom reference dataset for training. This final model is then evaluated against an independent test dataset provided by Francis et al. (2020) (Section 4.2). Finally, we apply *ukis-csmask* on Landsat-9 images. Since no benchmark datasets are yet available for Landsat-9 cloud and cloud shadow segmentation, this is simply an application experiment with qualitative interpretation of the results (Section 4.3).

4.1 Model selection and comparison with Random Forest and Fmask

Table 1 provides a description of the different training setups used for U-Net (UNET) and Random Forest (RF). UNETL8B4A for example refers to a U-Net model that has been trained on four Landsat-8 spectral bands with contrast / brightness augmentation.

Model	Train data	Bands	Augm.	Class
L8B6	SPARCS Landsat-8	Red, Green, Blue, NIR, SWIR1, SWIR2	False	5
L8B4	SPARCS Landsat-8	Red, Green, Blue, NIR	False	5
L8B4A	SPARCS Landsat-8	Red, Green, Blue, NIR	True	5
L8B6A	SPARCS Landsat-8	Red, Green, Blue, NIR, SWIR1, SWIR2	True	5

Table 1. Training setups for U-Net and Random Forest models.

Figure 2 shows a summary of the results for each tested training setup grouped by satellite. It can be seen that UNETL8B6A (U-Net trained on six Landsat-8 spectral bands with augmentation) outperforms the other U-Net models. On all satellites the model trained with augmentation shows superior results. The only exception is Landsat-7, for which augmentation seems to not significantly affect the performance of the prediction. Highest accuracies are achieved on Landsat-8. This is not surprising because the machine learning models have been trained on data of the same satellite. Overall, the UNETL8B6A shows good generalization ability across satellites with only minor accuracy differences being observed.
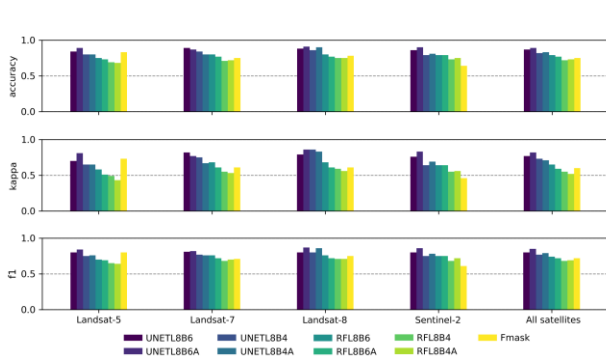
Figure 2. Results for each model grouped by satellite.

UNETL8B6A consistently outperforms Fmask and Random Forest which is confirmed by a qualitative comparison of the results across different satellites and locations (Figure 3). Moreover, the results indicate that models trained with only four spectral bands (Red, Green Blue and NIR) can already produce good results. Adding SWIR1 and SWIR2 spectral bands can, however, further improve the accuracy. Also applying contrast and brightness augmentations seems to be beneficial for the segmentation performance.
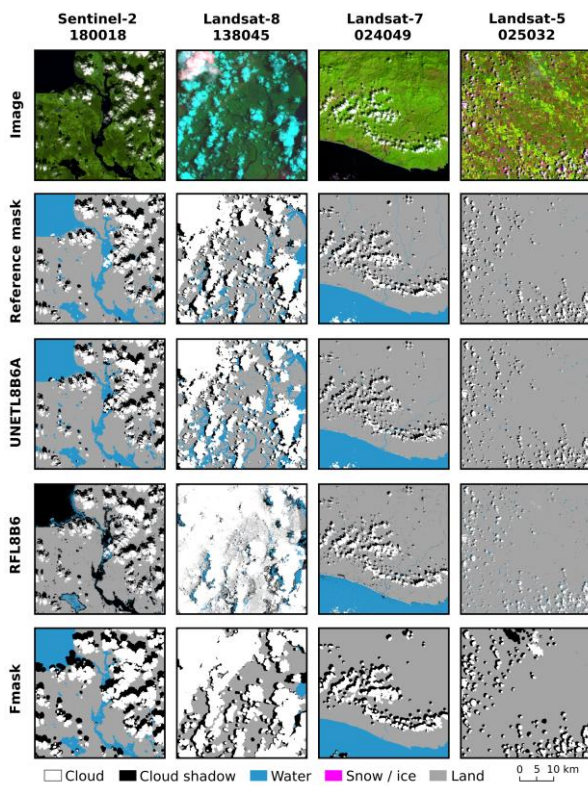

Figure 3. Selected results for the best performing U-Net and Random Forest models as well as Fmask.

4.2 Evaluation on the Sentinel-2 Cloud Mask Catalogue

We applied *ukis-csmask* on the 513 subscenes of the Sentinel-2 Cloud Mask Catalogue dataset. Figure 4 shows a selection of predictions for different environmental conditions (surface type, cloud height, cloud thickness, cloud extent).
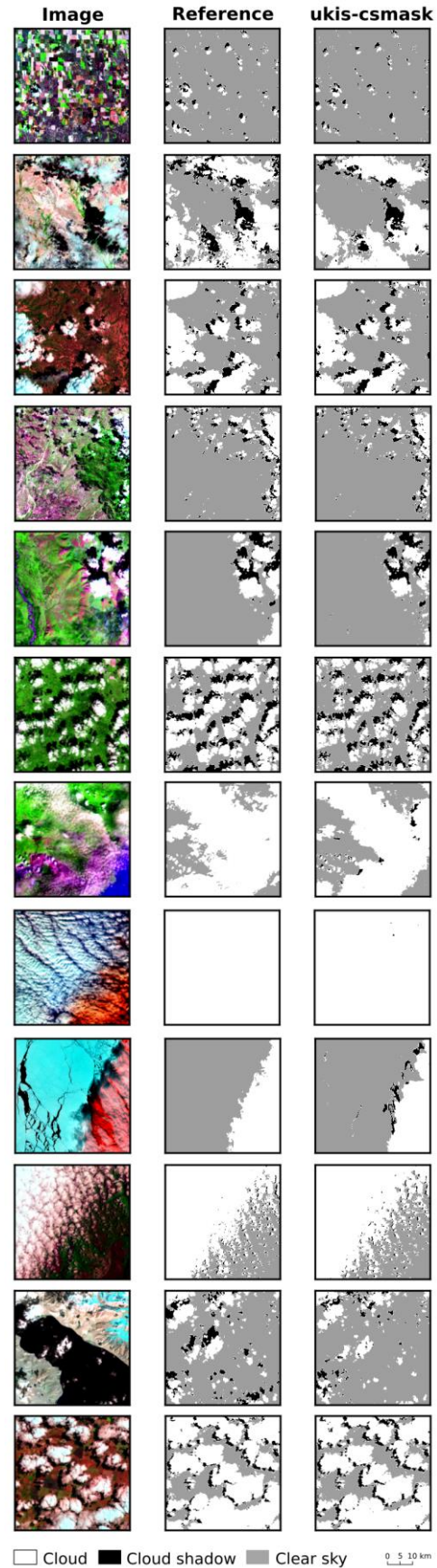

Figure 4. Selected *ukis-csmask* predictions on the Sentinel-2 cloud mask catalogue dataset.

Over all subscenes we achieve a mean overall accuracy of 0.80 (with 0.32 standard deviation), mean IoU of 0.77 (with 0.32 standard deviation) and mean F1 score of 0.81 (with 0.31 standard deviation). It can be seen that despite highly varying surface types with spectral signals that partially overlapping those of clouds and cloud shadows (e.g., snow / ice, open water bodies) relatively stable prediction results can be achieved. Variations in cloud type, thickness and extent seem to have a slightly larger influence on the segmentation performance. Cloud misclassifications are visible for thin cirrus clouds and some confusion between water bodies and cloud shadows as well as terrain shadows and cloud shadows can be observed in few of the tested subscenes.

4.3 Application to Landsat-9 images

Figure 5 shows prediction results of *ukis-csmask* on three Landsat-9 Level-1C images acquired over Australia, Greece and Colombia. Despite diverse surface and cloud types, the prediction results delineate well the visible clouds and cloud shadows. Similar to the performance on Sentinel-2 and the previous Landsat satellites, critical surface types like snow / ice or dark water bodies are separated consistently well from clouds and cloud shadows in all tested images. This first application experiment with qualitative interpretation of the results indicates that *ukis-csmask* can successfully be applied to Landsat-9 images.
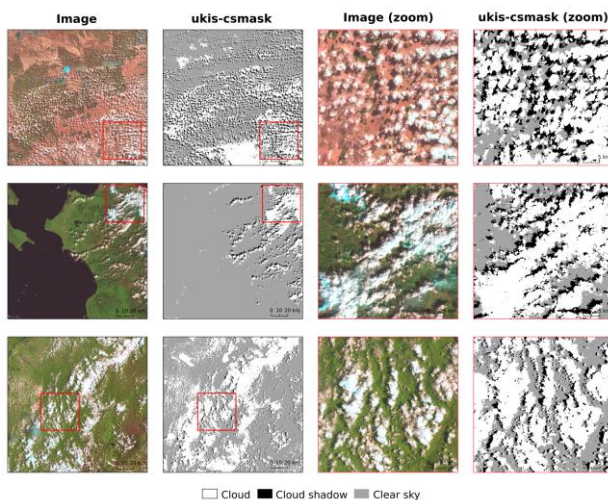

Figure 5. *ukis-csmask* predictions on Landsat-9 images acquired over Australia (first row), Greece (second row) and Colombia (third row).

5. DISCUSSION

With this study we could show that a pre-trained U-Net model that uses as input only four spectral bands (Red, Green, Blue and NIR) can already produce competitive prediction results. This is particularly relevant when sensors with limited spectral resolution (e.g., Planet Scope) are targeted by an application or when rapid and resource effective processing is required (Zhaoxiang et al., 2018). However, adding SWIR bands to the input feature space could improve the results on all tested sensors. Aerosols have a stronger effect on shorter wavelengths and atmospheric transparency is higher in the SWIR wavelengths. Therefore, adding SWIR bands to the input feature space can improve the distinction between clouds, snow

and ice. Rule-based methods, such as Fmask, also consider the Thermal Infrared (TIR) bands if available from a sensor. We specifically decided against using the TIR bands to increase the number of satellites that we can potentially support with *ukis-csmask*. Applying contrast and brightness augmentation on the training data helped the network to learn a certain degree of invariance to changes in the target domain (e.g., atmospheric conditions, seasonality or sun elevation).

A general issue with remote sensing reference datasets and specifically datasets related to cloud and cloud shadow segmentation concerns the fractal geometry and fuzzy reflectance characteristics of soft class boundaries. The border between cloud and clear sky as well as the fuzzy gradient of shadow outlines may lead to annotation inconsistencies when trying to define a hard class boundary for the reference masks. Therefore, annotation of reference masks is at least partially subjective and may introduce a bias in the reference dataset and hence affect the performance results. In consequence, the results should be interpreted as relative to a human operator rather than ground-truth.

Compared to other studies, we do not apply post-processing steps and focus on multi-sensor generalization ability. Accordingly, we test our results against independently acquired datasets from multiple sensors with overlapping spectral bands. In addition to our initial work in this direction (Wieland et al., 2019), we extended the experiments with a performance evaluation against the freely available Sentinel-2 Cloud Mask Catalogue (Francis et al., 2019) to provide results that are reproducible and comparable. Moreover, we tested *ukis-csmask* for the first time on Landsat-9 Level-1C images. Satellite-specific reference datasets for Landsat-9 are not yet available, so that we could only do a first subjective evaluation of the results. First predictions look promising, but more effort would be required to establish relevant benchmark datasets for cloud and cloud shadow segmentation on Landsat-9 images and hence provide quantitative and comparable results for this satellite.

6. CONCLUSIONS

The findings of this study indicate that a single CNN model is capable to generalize across a variety of different satellite sensors, geographies and environmental conditions without the need for atmospheric correction or more sophisticated domain adaptation techniques other than augmentation. We developed the outcomes of our scientific studies into a ready-to-use open-source Python package to enable transparency and reproducibility. Compared to other solutions, *ukis-csmask* allows us to quickly and accurately identify cloud and cloud shadow pixels in satellite images of various multi-spectral sensors, which is fundamental for unbiased down-stream analysis. This becomes particularly relevant in emergency applications where time is critical and all available data sources need to be utilized to provide timely information products about an evolving disaster.

REFERENCES

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Chen, Y., Fan, R., Bilal, M., Yang, X., Wang, J., Li, W., 2018. Multilevel Cloud Detection for High-Resolution Remote Sensing Imagery Using Multiple Convolutional Neural Networks. ISPRS International Journal of Geo-Information 7, 181. https://doi.org/10.3390/ijgi7050181

Domnich, M., Sünter, I., Trofimov, H., Wold, O., Harun, F., Kostiukhin, A., Järveoja, M., Veske, M., Tamm, T., Voormansik, K., Olesk, A., Boccia, V., Longepe, N., Cadau, E.G., 2021. KappaMask: AI-Based Cloudmask Processor for Sentinel-2. Remote Sensing 13, 4100. https://doi.org/10.3390/rs13204100

Francis, A., Mrziglod, J., Sidiropoulos, P., Muller, J.-P., 2020. Sentinel-2 Cloud Mask Catalogue. https://doi.org/10.5281/zenodo.4172871

Francis, A., Sidiropoulos, P., Muller, J.-P., 2019. CloudFCN: Accurate and Robust Cloud Detection for Satellite Imagery with Deep Learning. Remote Sensing 11, 2312. https://doi.org/10.3390/rs11192312

Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENµS, LANDSAT and SENTINEL-2 images. Remote Sensing of Environment 114, 1747–1755. https://doi.org/10.1016/j.rse.2010.03.002

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. arXiv:1502.01852 [cs].

Hollstein, A., Segl, K., Guanter, L., Brell, M., Enesco, M., 2016. Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. Remote Sensing 8, 666. https://doi.org/10.3390/rs8080666

Hughes, M.J., Kennedy, R., 2019. High-Quality Cloud Masking of Landsat 8 Imagery Using Convolutional Neural Networks. Remote Sensing 11, 2591. https://doi.org/10.3390/rs11212591

Jeppesen, J.H., Jacobsen, R.H., Inceoglu, F., Toftegaard, T.S., 2019. A cloud detection algorithm for satellite imagery based on deep learning. Remote Sensing of Environment 229, 247–259. https://doi.org/10.1016/j.rse.2019.03.039

Jiao, L., Huo, L., Hu, C., Tang, P., 2020. Refined UNet V2: End-to-End Patch-Wise Network for Noise-Free Cloud and Shadow Segmentation. Remote Sensing 12, 3530. https://doi.org/10.3390/rs12213530

Kingma, D.P., Lei, J., 2015. Adam: A Method for Stochastic Optimization. ArXiv:1412.6980v9 15.

López-Puigdollers, D., Mateo-García, G., Gómez-Chova, L., 2021. Benchmarking Deep Learning Models for Cloud Detection in Landsat-8 and Sentinel-2 Images. Remote Sensing 13, 992. https://doi.org/10.3390/rs13050992

Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. Sen2Cor for Sentinel-2, in:

Bruzzone, L., Bovolo, F., Benediktsson, J.A. (Eds.), SPIE Image and Signal Processing for Remote Sensing XXIII. Warsaw, Poland, p. 3. https://doi.org/10.1117/12.2278218

Mohajerani, S., Saeedi, P., 2021. Cloud and Cloud Shadow Segmentation for Remote Sensing Imagery via Filtered Jaccard Loss Function and Parametric Augmentation. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 14, 4254–4266. https://doi.org/10.1109/JSTARS.2021.3070786

Ozkan, S., Efendioglu, M., Demirpolat, C., 2018. Cloud detection from RGB rcolor remote sensing images with deep pyramid networks. ArXiv:1801.08706v1 4.

Python Fmask [WWW Document], 2022. URL http://pythonfmask.org/en/latest/ (accessed 5.1.22).

Scaramuzza, P., Foga, S., Dwyer, J., 2016. L8 Biome Cloud Validation Masks. https://doi.org/10.5066/F7251GDH

Scikit-learn. [WWW Document], 2022. URL http://scikit-learn.org (accessed 5.1.22).

Skakun, S., Wevers, J., Brockmann, C., Doxani, G., Aleksandrov, M., Batič, M., Frantz, D., Gascon, F., Gómez-Chova, L., Hagolle, O., López-Puigdollers, D., Louis, J., Lubej, M., Mateo-García, G., Osman, J., Peressutti, D., Pflug, B., Puc, J., Richter, R., Roger, J.-C., Scaramuzza, P., Vermote, E., Vesel, N., Zupanc, A., Žust, L., 2022. Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2. Remote Sensing of Environment 274, 112990. https://doi.org/10.1016/j.rse.2022.112990

Smith, L.N., 2017. Cyclical Learning Rates for Training Neural Networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). Presented at the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, Santa Rosa, CA, USA, pp. 464–472. https://doi.org/10.1109/WACV.2017.58

USGS, 2016. L8 SPARCS Cloud Validation Masks. [WWW Document]. URL https://landsat.usgs.gov/sparcs (accessed 7.12.21).

Wieland, M., Li, Y., Martinis, S., 2019. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. Remote Sensing of Environment 230, 1–12.

Wieland, M., Martinis, S., 2019. A modular processing chain for automated flood monitoring from multi-spectral satellite data. Remote Sensing 11, 2330.

Zhaoxiang, Z., Iwasaki, A., Guodong, X., Jianing, S., 2018. Small Satellite Cloud Detection Based On Deep Learning and Image Compression. Preprints. https://doi.org/10.20944/preprints201802.0103.v1

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment 159, 269–277. https://doi.org/10.1016/j.rse.2014.12.014