

Artificial neural networks for cloud masking of Sentinel-2 ocean images with noise and sunglint

Viktoria Kristollari & Vassilia Karathanassi

To cite this article: Viktoria Kristollari & Vassilia Karathanassi (2020) Artificial neural networks for cloud masking of Sentinel-2 ocean images with noise and sunglint, International Journal of Remote Sensing, 41:11, 4102-4135, DOI: [10.1080/01431161.2020.1714776](https://doi.org/10.1080/01431161.2020.1714776)

To link to this article: <https://doi.org/10.1080/01431161.2020.1714776>



Published online: 26 Jan 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



Artificial neural networks for cloud masking of Sentinel-2 ocean images with noise and sunglint

Viktoria Kristollari and Vassilia Karathanassi 

Lab of Remote Sensing, School of Rural and Surveying Engineering, National Technical University of Athens, Athens, Greece

ABSTRACT

Cloudy regions in optical satellite images prevent the extraction of valuable information by image processing techniques. Several threshold, multi-temporal and machine learning approaches have been developed for the separation of clouds in land and ocean applications, but this task still remains a challenge. Concerning deep water marine applications, the main difficulties are imposed in regions with high noise levels and sunglint. In this study, artificial neural networks (ANNs) with different configurations are evaluated for the detection of clouds in Sentinel-2 images depicting deep water regions with several noise levels. The ANNs are trained on a manual public dataset and on a manual dataset created for the needs of this study, which authors intend to make publicly available. Results are compared with the cloud masks produced by three state-of-the-art algorithms: Fmask, MAJA, and Sen2Cor. It was shown that the ANNs trained on the second dataset perform very favourably, in contrast to the ANNs trained on the first dataset that fails to adequately represent the spectra of the noisy Sentinel-2 images. This study further reinforces the value of the 'cirrus' band and indicates the bands that mitigate the influence of noisy spectra, by defining and examining an index that characterizes the importance of the bands according to the weights produced by the ANNs. Finally, the possibility of improving results by making predictions using the feature scaling parameters of the test set instead of those of the training set is also investigated in cases where the test set cannot be adequately represented by the training set.

ARTICLE HISTORY

Received 15 May 2019
Accepted 11 November 2019

1. Introduction

In optical satellite images, the presence of clouds is a crucial obstacle in land and ocean studies performed by image analysis tasks. Thus, the exclusion of clouds from the data is an important step which needs to be implemented prior to atmospheric correction. Two common assumptions that are employed in various cloud detection algorithms are that clouds are characterized by higher reflectance and lower brightness temperature than other types of surfaces (Platnick et al. 2003; Shin, Pollard, and Muller 1996; Zhu, Wang, and Woodcock 2015; Zhu and Woodcock 2012). Based on the aforesaid assumptions, most of current cloud detection methods extract the clouds from the imagery through ruled

CONTACT Viktoria Kristollari  vkristoll@central.ntua.gr  Lab of Remote Sensing, School of Rural and Surveying Engineering, National Technical University of Athens, Heron Polytechniou 9, Zographos, Athens 15780, Greece

based classification which applies a set of thresholds (both static and dynamic) of reflectance and brightness temperature. Threshold-based cloud detection is usually platform-specific and strongly linked to the geographical area and date of data collection (Ackerman et al. 2008; Banks and Mélin 2015). Most well-known threshold methods are ACCA (Automatic Cloud Cover Assessment) (Irish 2000) and Fmask (Function of mask) (Zhu, Wang, and Woodcock 2015; Zhu and Woodcock 2012) which have been designed for Landsat imagery (Foga et al. 2017). Multi-temporal methods have also been applied extensively by researchers and are based on the idea that abrupt changes in image time series are mainly caused by the presence of clouds since other types of surfaces follow smooth variations (Hagolle et al. 2010; Candra, Phinn, and Scarth 2016; Ahmad and Quegan 2014; Karvonen 2014; Lin et al. 2015; Mateo-García et al. 2018). A well-known multi-temporal cloud masking algorithm is MAJA (Hagolle et al. 2017) designed for Sentinel-2 images.

Several VISNIR (Visible-Near Infrared) and SWIR (Shortwave Infrared) wavelengths have been selected by researchers for cloud detection applications since a variety of VISNIR and SWIR bands carries useful information. Channel 2 (0.725–1.10 μm) of AVHRR (Advanced Very-High-Resolution Radiometer) is considered to provide high contrast between clouds and water (Gallegos, Hawkins, and Cheng 2008). The reflectance ratio of $R_{0.87}/R_{0.66}$ μm is used in MODIS (Moderate Resolution Imaging Spectroradiometer) data along with 0.936- μm band for low cloud detection, while the SWIR band at 1,380 nm is used for the detection of high clouds (cirrus) (Ackerman et al. 1998). The visible threshold test of the MSG/SEVIRI (Meteosat Second Generation/Spinning Enhanced Visible and Infrared Imager) cloud mask is applied on the 0.8- μm band over the seas and on the 0.6- μm band over coasts (Hocking, Francis, and Saunders 2011). Reflectance in the blue was used in the multi-temporal research of Hagolle et al. (2010) in Formosat-2 and Landsat 5, 7 images, and in Proba-V cloud detection (Iannone et al. 2017). In Landsat 8, bands 3 (0.525–0.600 μm) and 4 (0.630–0.680 μm) were selected by Candra, Phinn, and Scarth (2016) for the distinction between cloud and non-cloud. Finally, Wang and Shi (2006) proposed a SWIR threshold at 1,240 nm, 1,640 nm and 2,130 nm for cloud masking in turbid waters instead of the 865/869 nm used in MODIS and SeaWiFS (Sea-Viewing Wide Field-of-View Sensor) respectively, which is considered more suitable for open oceans.

Conventional machine learning as well as deep learning techniques have also been introduced for cloud masking of imagery collected by a wide variety of platforms and have indicated successful results. Bai et al. (2016) trained an SVM-RBF (Support Vector Machine-Radial Basis Function) classification model on fused multiple features of cloud and non-cloud regions of GaoFen-1 and GaoFen-2 images. Baseski and Cenaras (2015) trained a linear kernel-based SVM on images acquired from the commercial multispectral satellites: Geosy, Ikonos, and WorldView 2. Yuan and Hu (2015) employed the bag-of-words (BOW) model to construct compact features from dense local SIFT (Scale-Invariant Feature Transform) features extracted from RapidEye and Landsat imagery. Hughes and Hayes (2014) determined cloud and cloud shadow by training a total of 15 artificial neural network (ANN) configurations and explored the inclusion of spatial information through the tassell-cap transformation in Landsat 7 scenes. Strandgren et al. (2017) used the MSG/SEVIRI imager and detected cirrus clouds by utilizing a set of four ANNs trained on thermal observations and auxiliary data. Le Goff et al. (2017) compared a convolutional neural network architecture (CNN) with five ANNs applied to different spectral and spatial

features extracted from Spot 6 images. Taravat et al. (2015) combined the most significant band ratios and multi-layer perceptrons (MLPs) to differentiate clouds from a background in Landsat ETM+ and MSG/SEVIRI data. Weng, Kong, and Xia (2016) used deep extreme learning machine to detect cloud cover fraction and distinguished thick from a thin cloud in HJ-1A/B satellite images. Sholar (2017) trained pixel-level decision tree classifiers on the Hollstein et al. (2016) database and fed a deconvolutional network with the labelled results by the use of the Alexnet-FCN model in Sentinel-2 images. Mateo-Garcia, Gomez-Chova, and Camps-Valls (2017) studied patch-to-pixel and patch-to-patch CNN architectures for cloud masking of Proba-V multispectral images. Li et al. (2018) integrated multi-scale convolutional features in a network based on fully convolutional network (FCN) and Segnet which was trained on Gaofen-1 images. Xie et al. (2017) designed a CNN with two branches trained on Quickbird RGB patches of different sizes and distinguished thick from thin clouds. Finally, Zhang et al. (2018) proposed an ensemble method combining a lightweight U-Net with wavelet image compression for on-board cloud detection in small satellites.

From the above, it is easily implied that cloud detection methods use either spectral or spatial information, or their combination along with the inclusion of temporal information. In general, cloud detection methods usually suffer from thin cloud omission and bright non-cloud object commission. Sun glint and high noise levels constitute bright non-cloud objects. Sun glint is a transient anomaly that occurs when sunlight is reflected from the seawater surface directly into the down looking optical sensor (Gould, Arnone, and Sydor 2001; Lee et al. 2010). It is influenced by the position of the sun, viewing angle of the optical sensor, water refractive index, cloud cover, wind direction, and speed (Cox and Munk 1954; Mobley 1999; Zhang and Wang 2010). High noise levels in satellite images can appear as random ('salt and pepper') or periodic (vertical or oblique stripes) and can be optically recognized without difficulty. Directional reflectance effects caused by the configuration of the 12 detectors of the Multispectral Instrument (MSI) of Sentinel-2 imaging mission (ESA 2019) may also be considered as periodic noise with oblique wide stripes, whereas the Sentinel-2 cirrus band (1.375 μm and relatively low SNR(50)) additionally presents periodic noise with linear stripes.

Researchers have employed several methods to alleviate the high noise levels and sun glint for cloud masking applications. Concerning high noise levels, researchers usually use spatial information and post-processing methods. Tian et al. (1999) used textural properties since they tend to be less sensitive to detector noise to train probability NNs and SOMs (Self-Organized Maps) on GOES-8 (Geostationary Operational Environmental Satellite) images for cloud classification. On the same basis, Le Goff et al. (2017) also examined textural features to train CNNs on Spot 6 images, and Charantonis et al. (2009) experimented on using a spatial variation index to train SOMs on Landsat ETM+ images. Hughes and Hayes (2014) removed spatial noise with TVR (Total Variation Regularization) before training ANNs and post-processed the masks by applying the median filter. The median filter as a post-processing step was also applied by Zhai et al. (2018) whose method was examined on multispectral and hyperspectral sensors and was based on the use of spectral indices, while opening and closing operators were applied by Fisher (2014) on the output of their morphological method. Concerning sun glint, researchers use spectral and spatial information, as well as geometric. Hu (2011) developed an algorithm to discriminate sun glint from clouds based on its red characteristics by use of (469, 555,

1,240) nm MODIS bands. Garaba et al. (2012) masked measurements collected by unmanned/automated platforms and affected by sunglint by setting thresholds in the 700–950-nm range on the premise that open seawater is assumed to absorb all light in the NIR. Based on the high variability of clouds, Nicolas et al. (2005) (Polder-2 instrument) and Martins et al. (2002) (MODIS) proposed using a spatial variability threshold of reflectance at NIR, while Chen and Zhang (2015) (MODIS) produced better results by examining the spatial variability at SWIR. Roslan et al. (2014) attempted to mediate the sunglint effect by the use of image enhancement techniques on AVHRR images. Ricciardelli, Romano, and Cuomo (2008) used a combination of physical, statistical, and temporal approaches on SEVIRI images and managed not to overestimate cloudy pixels due to the sunglint. Ishida et al. (2018) trained SVMs on MODIS images and attempted to treat sunglint areas by use of the reflectance ratio of $R_{0.905}/R_{0.935}$ μm and a feature that combined $R_{0.87}$ μm , solar angle, and the satellite angle. Finally, Schröder et al. (2002) trained ANNs on textural features and gradient-filtered radiances on images collected by an airborne spectrographic imager. They observed that sunglint areas can be twice as bright as clouds of low brightness and used a single absorption-free wavelength (753 nm). They also decided not to include sun and viewing geometry as input parameters in order to avoid incorrect correlation between Sun zenith angle and cloudiness. It should be noted that several cloud mask products of satellites with low spatial resolution define sunglint affected areas geometrically. In more detail, the algorithm for the MODIS cloud masks defined the potential geometric sunglint region as being within 36 degrees of the specular direction and modified spectral tests on these areas (Platnick et al. 2003). The cloud masking algorithm used on GCOM-C (Global Change Observation Mission) satellite identified sunglint areas as those whose cone angle between the solar incident and the satellite direction is lower than 35 degrees (Nakajima et al. 2019). Finally, the cloud mask product of Himawari-8 satellite defined sunglint to be present in areas where the sun zenith angle is lower than 75 degrees (Imai and Yoshida 2016). They also took wind predictions into account.

The main advantage of deep learning techniques in comparison with conventional machine learning is their independency of the need for the extraction of human-engineered features which is a lengthy process. Another advantage concerns the fact that neural networks are reported as being less sensitive to noise in the training set (Bishop 1995; Yhann and Simpson 1995). This property is very important for cloud detection since the common practice when labelling pixels is visual observation which involves certain inaccuracy (Gallegos, Hawkins, and Cheng 2008; Hagolle et al. 2010; Yuan and Hu 2015; Zhu and Woodcock 2012; Hughes and Hayes 2014; Lin et al. 2015; Zhu, Wang, and Woodcock 2015; Bai et al. 2016; Foga et al. 2017; Iannone et al. 2017; Sholar 2017; Xie et al. 2017; Mateo-García et al. 2018). However, when the complexity of the architecture of the network is high, combined with the laborious process of the selection of the optimal hyperparameters, deep learning techniques can also prove to be time-consuming. Artificial neural networks by using multiple layers of perceptrons are characterized by simpler architectures but have proven to be a fast and very efficient method in a wide variety of applications. Thus, this article focuses on the use of multi-layer perceptrons (MLPs) for separating cloudy areas from deep water areas in Sentinel-2 images with high noise levels, directional reflectance effects, and sunglint, a task which is still a challenge. The study makes use only of spectral information and proposes

a simple and time-efficient method which produces satisfactory results. For its purpose: a) MLPs with different configurations are trained on two different databases: the public dataset produced manually by Hollstein et al. (2016) and a dataset based on the images used in this study, which will be also publicly provided, b) the possibility of improving results by making predictions using the feature scaling parameters of the test set instead of those of the training set is investigated in cases where the test set cannot be adequately represented by the training set and c) an index that characterizes the importance of the bands according to the weights produced by the MLPs is defined and examined. The results are compared with cloud masks produced by three state-of-the-art algorithms: Fmask, MAJA, and Sen2Cor (Richter, Louis, and Müller-Wilm 2012).

2. Materials and methods

2.1. Data description

Three datasets were used for analysis in this study. The first dataset consists of spectra extracted from the database created by Hollstein et al. (2016), the second dataset contains 79 Sentinel-2 satellite images analysed for the purpose of this study and the third dataset contains spectra extracted from the second dataset. In this paper, the first dataset is named 'Hollstein dataset', the second dataset is named 'S2 image dataset' and the third dataset is named 'S2 spectra dataset'. The Hollstein dataset and the S2 spectra dataset were used in the training and evaluation process, while the S2 image dataset was used in the visual inspection process. The datasets are described in detail below.

2.1.1. Hollstein dataset

The database created by Hollstein et al. (2016) is a manually created database with reflectance spectra collected around the globe from Sentinel-2 level 1C satellite images. To our knowledge, it is the only publicly available database of manually selected spectra from Sentinel-2 images and contains the classes: 'clear', 'cloud', 'shadow', 'snow', 'cirrus', and 'water'. The spectra were selected by use of spectral tools which included false-colour composites, image enhancements, and graphical visualization of spectra. It is also clarified that for the time being, publicly available manually created cloud masks for Sentinel-2 images do not exist. The location of the scenes where the spectra were collected is depicted in Figure 1. The data were collected in 2016 and 2017 with 20-m spatial resolution. The selected spectra are 5,647,725 and 3,152,273, respectively, and the database is stored in two separate .hd5 files. For this study, three classes were extracted from this database: 'cloud' (spectra from opaque clouds), 'cirrus' (spectra from cirrus and vapour trails) and 'water' (spectra from lakes, rivers, and seas). The number of spectra for each class is presented in Table 1. It is noted that the 'cloud' and 'cirrus' class were joined in one class (cloud) in the experiments where this dataset was used, since separating opaque and cirrus clouds is out of the scope of this study.

2.1.2. S2 image dataset

The second dataset used in this study contains 79 Sentinel-2A/2B level 1C images. These images refer to two tiles of the same orbit collected by the Sentinel-2 MultiSpectral Instrument in 2016 (four images), 2017 (40 images) and 2018 (35 images). The viewing



Figure 1. Location of scenes of the Hollstein et al. (2016) database.

Table 1. Spectra comprising the Hollstein dataset.

Class	Coverage	Number of spectra
Cloud	Opaque cloud	1,500,202
Cirrus	Cirrus and vapour trails	1,205,979
Water	Lakes, rivers, seas	1,435,003
Total		4,141,184

geometries of the Sentinel-2 detectors in these tiles range from 1° to 11° in zenith and from 21° to 316° in azimuth. The dates of collection covered all seasons of the year: 28 winter images (December, January, February), 24 spring images (March, April, May), 11 summer images (June, July, August) and 16 fall images (September, October, November). The collection time varied between 10:30 and 10:35 a.m. UTC. Depicting several noise levels and a wide variety of the percentage of cloud cover were the important factors during the selection of the dates. The noise analysed in this study refers to the random and periodic noise (mainly caused by directional reflectance effects) of the Sentinel-2 images. An example of the periodic noise caused by the detectors can be seen in [Figure 9\(a\)](#) and an example of the periodic noise of the cirrus band can be seen in [Figure 9\(c\)](#). A crucial factor for the selection of the study area was the availability of MAJA masks. These masks are highly accepted by the Remote-Sensing community and thus were considered significant for the evaluation process. It was decided to use the already available masks because running the binary code provided by the creators of the method requires high computational power. [Figure 3](#) depicts with red colour the scenes with available MAJA masks. In this Figure, it can be observed that concerning ocean applications, these masks are at present scarce. Depicting a high percentage of water was also considered during the selection of the tiles of the study area. Sentinel-2 images contain 13 bands, three with 60-m spatial resolution, four with 10-m spatial resolution, and six with 20-m spatial resolution. The wavelengths of the three spatial resolutions of the Sentinel-2 instruments are shown in [Table 2](#). Before analysis, these images were processed. The bands with spatial resolution 10 and 20 m were resampled to 60 m and then the images were cropped in order to remove the land and

Table 2. Wavelengths of the three spatial resolutions of the Sentinel-2 instruments.

Spatial resolution (m)	Band number	S2A	S2B
		Central wavelength (nm)	Central wavelength (nm)
10	2	496.6	492.1
	3	560.0	559
	4	664.5	665
	8	835.1	833
20	5	703.9	703.8
	6	740.2	739.1
	7	782.5	779.7
	8A	864.8	864
	11	1613.7	1610.4
60	12	2202.4	2185.7
	1	443.9	442.3
	9	945.0	943.2
	10	1373.5	1376.9

depict optically homogenous sea regions. The x-size (columns) of the cropped images was 1,830 pixels and the y-size (rows) was 1,130 pixels. [Figure 2](#) depicts the study area, the location of the Sentinel-2 tiles (white polygons (1,2)) and the cropped tiles (red polygons (3,4)).

2.1.3. S2 spectra dataset

This dataset includes spectra manually and randomly extracted from images of the S2 image dataset. In more detail, it includes:

a) Reflectance water spectra which were manually extracted by visual observation from 30 of the images of the S2 image dataset. These 30 images consisted of 8 winter images (December, January, February), 9 spring images (March, April, May), 5 summer images (June, July, August) and 4 fall images (September, October, November). These spectra were extracted from water areas with high noise levels and sunglint. [Figure 4](#) depicts some example scenes from which spectra were obtained through regions of interest



Figure 2. The study area, the Sentinel 2 tiles (white polygons (1, 2)), and the cropped tiles (red polygons (3, 4)).

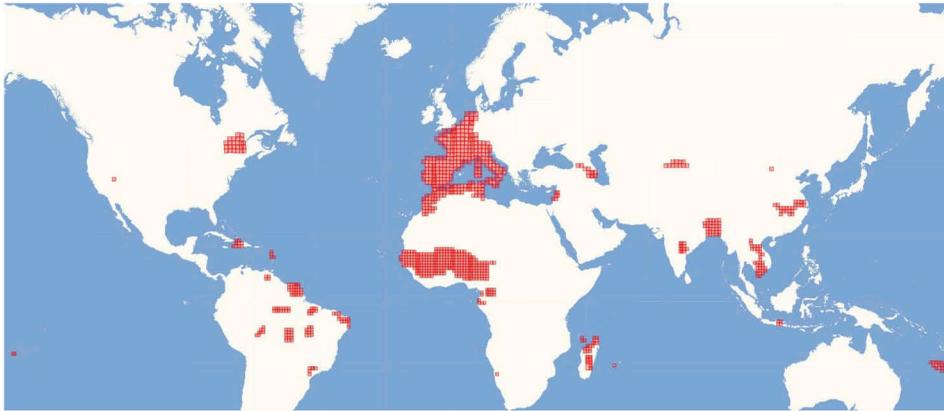


Figure 3. Scenes with available MAJA masks (red colour).

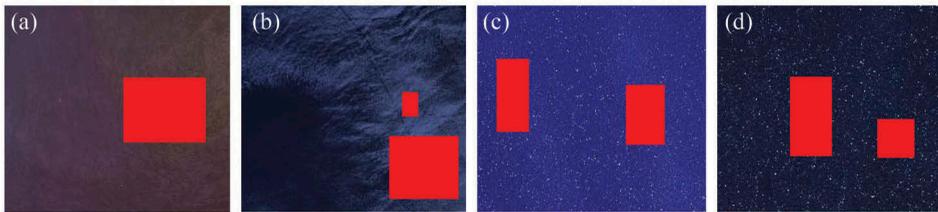


Figure 4. Sentinel-2 scenes with sunglint (*a,b*) and noise (*c,d*).

(rois). The spectra with high noise levels were extracted from regions that noise was visually recognized without difficulty, i.e. without the application of enhancement techniques (e.g. histogram stretching). The spectra with sunglint presence were discriminated from optically thin clouds by use of the cirrus band ($1.375 \mu\text{m}$) which is less affected by sunglint. The geometric pattern of sunglint was also taken into account. Public access will be provided to the database created by the manually extracted water spectra. Spectra of cloud and water without visually obvious presence of noise and sunglint were not manually extracted. To our opinion, these spectra would not be characterized by lower omission and commission errors than those produced by the third experiment mentioned above, due to the fact that commission and omission errors usually occur in areas where the observer cannot with certainty label the correct class of a pixel, because of high visual similarity (e.g. very thin clouds). In addition, in such a scenario, the observer would choose 'easier' cases in order to increase the confidence of labelling which would probably lead to a less effective training set.

b) Cloud spectra and water spectra which were randomly extracted from 34 images of the S2 image set (different from the 30 images mentioned above). These 34 images consisted of 15 winter images, 10 spring images, 7 summer images, and 6 fall images and the water areas were characterized by low noise and no sunglint presence. These spectra were selected from the cloud masks which were successfully derived from the implementation of the third experiment described in [sections 2.3.2.1.](#) and [3.1.3.](#) The number of manually and randomly extracted spectra is presented in [Table 3.](#) From each

Table 3. Spectra comprising the S2 spectra dataset.

Class	Number of spectra
Manually extracted water	2,133,324
Randomly extracted cloud	2,040,000
Randomly extracted water	2,040,000

of the 34 images, 60,000 spectra were obtained for cloud and water, respectively, which accounts for 6% of each image spectra ($120,000/(1,830 \times 1,130)$). This percentage of labelled areas corresponds to 4,080,000 spectra ($120,000 \times 34$).

2.2. Theoretical background

2.2.1. Multilayer perceptron neural network

ANNs consist of a number of neurons that exchange information in a similar manner as biological nerve cells transmit information via synapses in the human brain. An artificial neuron or perceptron (Rosenblatt 1958) forms the basis for designing artificial neural networks. A model of a perceptron is shown in Figure 5.

A neuron k can be described by the following pair of equations (Equations 1, 2):

$$u_k = \sum_{i=1}^N w_{k,i} x_i \quad (1)$$

$$y_k = \varphi(u_k + b_k) \quad (2)$$

where x_1, \dots, x_N are the input signals, $w_{k,1}, \dots, w_{k,N}$ are the synaptic weights of neuron k , b_k is the bias, $\varphi(\cdot)$ is the activation function, and y_k is the output signal of the neuron. The

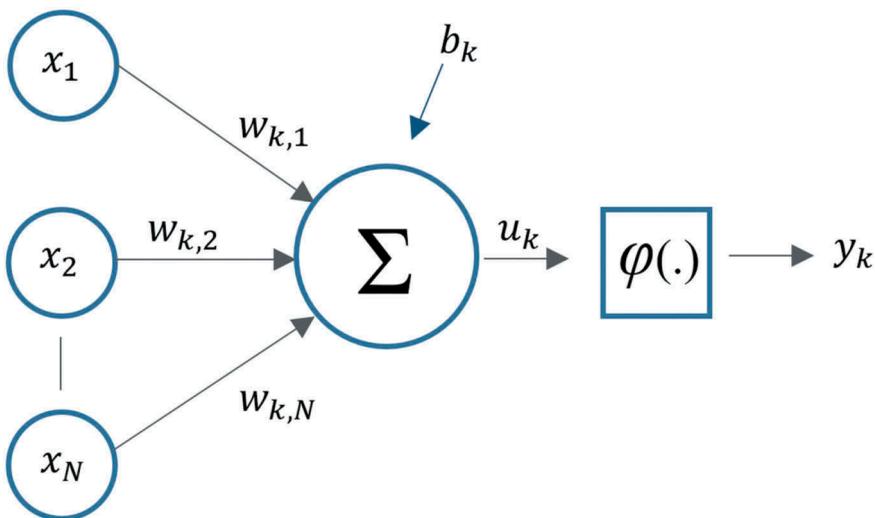


Figure 5. Model of a perceptron.

input signals in this study refer to the training spectra extracted from the Hollstein dataset and the S2 spectra dataset.

ANN architecture consists of three units: input layer, output layer, and several hidden layers. The number of the nodes of the input layer is determined by its input parameters, and the number of the nodes of the output layer is determined by its desired output. Neurons in successive layers are connected by weights which represent the importance of the connections in the network. MLP model is a feed-forward artificial neural network classifier. Each neuron receives inputs from the neurons in the previous layer and through a non-linear activation function converts them to input for the neurons in the next layer. MLP utilizes backpropagation for training the network. During the backward pass, the network's actual output is compared with the target output through an objective function (cost function (C)) (Equation 3) that needs to be minimized.

$$C = \sum_{i=0}^N \left(y_i - \sum_{j=0}^M x_{i,j} w_j \right)^2 \quad (3)$$

where y_0, \dots, y_N are the true output values, $x_{0,0}, \dots, x_{N,M}$ are the values of the neurons in the previous layer, w_0, \dots, w_M are the weights connecting the output layer with the previous layer, N is the sample size and M is the number of connections. The output values in this study refer to the class of the spectra. Output values over 0.5 were classified to the cloud class while output values below 0.5 were classified to the water class.

The error estimates are computed for the output units and the weights that connect the output units with the previous hidden layer are adjusted to reduce these errors. The error adjustment is propagated to the connections of the units in the hidden layers and the connections originating from the input units. The backpropagation process is typically implemented by the stochastic gradient descent method (Rumelhart, Hinton, and Williams 1986) which produces the updated weights for learning rate: α by calculating the partial derivatives of the cost function with respect to each weight (Equation 4).

$$w_j \leftarrow w_j - \alpha \frac{\partial C}{\partial w_j} \quad (4) \quad j \in [0, M]$$

where w_j , C , M are defined in Equation 3 and $\frac{\partial C}{\partial w_0}, \dots, \frac{\partial C}{\partial w_M}$ are partial derivatives.

In this study, Adaptive Moment Estimation (Adam) (Kingma and Ba 2014) was used for the implementation of the backpropagation process which is an optimization algorithm of the stochastic gradient descent.

2.2.2. Adaptive moment estimation

Adaptive Moment Estimation (Adam) is an optimization algorithm of the stochastic gradient descent method for the calculation of the weights during the backpropagation process. The method stores an exponentially decaying average of past squared gradients \mathbf{u}_t (Equation 6) and an exponentially decaying average of past gradients \mathbf{m}_t (Equation 5). The gradients \mathbf{g}_t denote the vector of partial derivatives of the objective function (cost function) at timestep t . \mathbf{m}_t and \mathbf{u}_t are estimates of the first moment (the mean) and the second moment (the uncentered variance) of the gradients, respectively. The zero bias of \mathbf{m}_t and \mathbf{u}_t is counteracted by computing bias-corrected first and second moment

estimates ($\hat{\mathbf{m}}_t, \hat{\mathbf{u}}_t$) (Equations 7, 8). These are used to update the parameters (weights (θ_t)) (Equation 9).

$$\mathbf{m}_t = \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \mathbf{g}_t \quad (5)$$

$$\mathbf{u}_t = \beta_2 \mathbf{u}_{t-1} + (1 - \beta_2) \mathbf{g}_t^2 \quad (6)$$

$$\hat{\mathbf{m}}_t = \frac{\mathbf{m}_t}{1 - \beta_1^t} \quad (7)$$

$$\hat{\mathbf{u}}_t = \frac{\mathbf{u}_t}{1 - \beta_2^t} \quad (8)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{\mathbf{u}}_t}} \hat{\mathbf{m}}_t \quad (9)$$

where β_1 and β_2 are exponential decay rates for the moment estimates and η is the learning rate.

2.2.3. Feature scaling

Feature scaling is a typical step of data pre-processing which is applied to independent variables or features in order to create a particular range of values. The implementation of this process impedes the dominance of the results by features of high magnitude and accelerates calculations. One of the methods widely used for feature scaling is standardization (or Z-score normalization) which is the process of rescaling the features so that they will have the properties of a Gaussian distribution with $\mu = 0$ (mean value) and $\sigma = 1$ (standard deviation). This process was applied in this study for rescaling the features (spectra values) of the training and test sets. The rescaled values of the features (z) were calculated by equation 10.

$$z = \frac{x - \mu}{\sigma} \quad (10)$$

where x are the initial values of the features.

2.3. Method description

2.3.1. Description of the ANNs

In this study, a total of four artificial neural network configurations were trained on the Hollstein dataset and on the S2 spectra dataset, i.e. 8 trainings were implemented in total. The four configurations were differentiated by the use of different algorithms that prevent overfitting.

The architecture of the ANNs consisted of one input layer, two hidden layers, and one output layer. The input layer contained 13 neurons (the total number of Sentinel-2 bands), each of the two hidden layers contained 20 neurons and the output layer contained one neuron since the classification is binary (cloud/water). It was decided to use the spectral information from all the bands of Sentinel-2 images since the literature exploits the VIS, NIR, and SWIR bands. In addition, the analysis of the importance of the different

wavelengths for the ANN was also a purpose of this study. The architecture and the number of neurons in the hidden layers were selected based on preliminary experiments conducted on the Hollstein dataset. The Rectified Linear Unit (ReLU) (Agarap 2018) (Equation 11) was utilized as an activation function in the two hidden layers. Its main advantages are computational simplicity, its linear behaviour, and its sparse representation capability since it can output true zero value. The sigmoid function (Nwankpa et al. 2018) was used as an activation function in the output layer (Equation 12). The graphs of the ReLU and the sigmoid function are presented in Figure 6. It should be stated that this figure follows the nomenclature of the Sentinel-2 products, i.e. the last band corresponds to number '12' (Table 2).

$$\varphi(x) = \max(0, x) \quad (11)$$

$$\varphi(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

where $x \equiv u_k$ as described in Equation 1.

Adam optimization (Equation 9) was selected for the backpropagation process with the default values of the Keras library (Chollet 2015) ($\eta = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$). In the first configuration, the ANN was trained without applying any algorithm that prevents overfitting. In the second configuration, the dropout method (Srivastava et al. 2014) was applied, which ignores neurons at random during the training phase. This method was applied with 0.3 value in both hidden layers, i.e. 30% of the neurons are ignored in each hidden layer. In the third and fourth configuration, the L1 (Equation 13) and L2 regularizations (Equation 14) (Park, Ho, and Chang 2018) which add a regularization term in the cost function (Equation 3) were, respectively, implemented in both hidden layers.

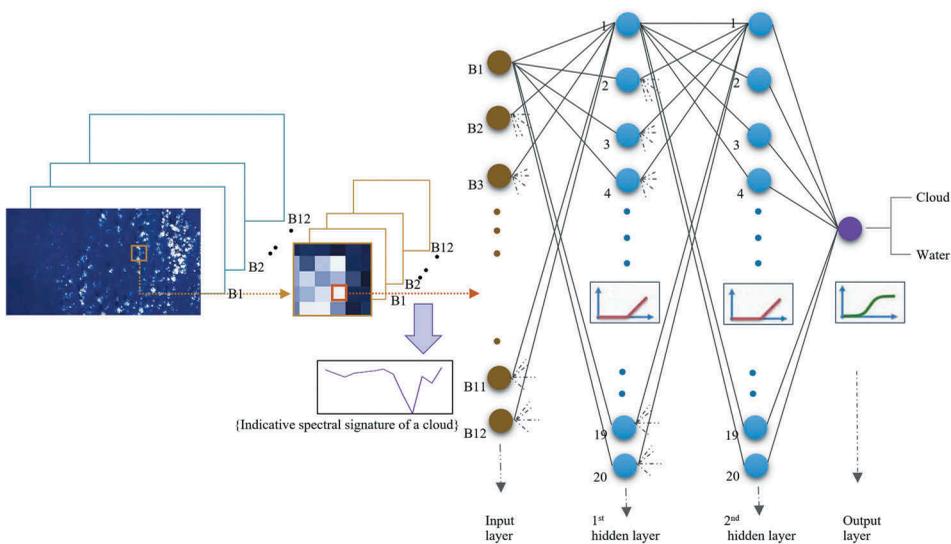


Figure 6. The proposed methodology.

$$C = \sum_{i=0}^N \left(y_i - \sum_{j=0}^M x_{ij} w_j \right)^2 + \lambda \sum_{j=0}^M |w_j| \quad (13)$$

$$C = \sum_{i=0}^N \left(y_i - \sum_{j=0}^M x_{ij} w_j \right)^2 + \lambda \sum_{j=0}^M w_j^2 \quad (14)$$

For the λ parameter, the value 0.001 was selected for the L1 regularization and the value 0.005 for the L2 regularization.

In all configurations, the ANNs were trained for 100 epochs, with batch size 1,024. The weights from all the epochs were stored and the weights that produced the higher accuracy in the training set were used for predictions on the test sets. Training was implemented on the Graphical Processing Unit (GPU): Nvidia GeForce GTX 960M and it lasted approximately 20 min for each of the 8 trainings. [Figure 6](#) presents the proposed methodology. The ANNs were trained by using the Keras library and the Tensorflow (Abadi et al. 2016) backend and were implemented in Python code. Tensorflow is an open-source software library for numerical computation developed by Google researchers. It uses a flexible data flow architecture that is suitable for parallel processing applications (e.g. neural networks). Keras is an open-source neural-network library written in Python and capable of running on top of Tensorflow. Creating neural-network models on Keras is simpler since emphasis was put on achieving user-friendliness.

2.3.2. Training the ANNs

2.3.2.1. Training on the Hollstein dataset. The Hollstein dataset was used in the training of all four configurations of the ANNs. The training set consisted of spectra extracted from the ‘water’ class, the ‘cloud’ class, and the ‘cirrus’ class. The number of labelled spectra for each class and their percentage which was calculated by use of the total number of spectra for each class of the Hollstein dataset ([Table 1](#)), is presented in [Table 4](#). The purpose of the choice of the number of training spectra for each class was the exploitation of a large number of the available labelled spectra, by simultaneously preserving a balance between the size of the classes ($N_{\text{cloud}} + N_{\text{cirrus}} \approx N_{\text{water}}$). Retaining an adequate number of spectra for the test set (different of those of the training set) was also important ($\geq 20\%$). As already mentioned the ‘cloud’ and ‘cirrus’ class were joined in one class during the training.

Three different experiments were implemented on the same training set ([Table 4](#)) which was rescaled using the average and standard deviation of the training set. The purpose of these experiments was to analyse the possibility of improving results by making predictions using the feature scaling parameters of the test set instead of those of the training set which is the usual practice. The motivation for this investigation was to maximize the exploitation of the Hollstein dataset since it contains a large number of

Table 4. Spectra comprising the Hollstein training set.

Class	A:Spectra of Hollstein training set	B:Spectra of Hollstein dataset	A as a proportion of total (%)
Water	1,000,000	1,435,003	67%
Cloud	500,000	1,500,202	33%
Cirrus	500,000	1,205,979	41%

publicly available spectra. In the first and second experiment, the spectra values of the test set were rescaled using the average value and standard deviation of the training set, while in the third experiment using the respective values of the test set. In addition, the experiments were differentiated by the test set used for the predictions. In the first experiment, the test set included spectra from the Hollstein dataset, in the second experiment it included spectra from the S2 spectra dataset and the S2 image dataset, and in the third experiment it included the S2 image dataset. The experiments are described in detail below and are summarized in Table 5. It is noted that 'z' in Table 5 symbolizes the input of the ANN. It is also noted that the term 'Predictions' on the titles of the subsections below refers to the testing process of the ANNs after they are trained. During this stage spectra not included in the training process are given as an input to the ANN and the output is evaluated.

First experiment: Predictions on the Hollstein dataset by using for the test set the feature scaling parameters of the training set

In the first experiment, the test set included spectra from the 'cloud' class, the 'cirrus' class, and the 'water' class, which were extracted from the Hollstein dataset and were different from the training set. The number of spectra for each class and their percentage calculated by use of the total number of spectra for each class used for the experiment is presented in Table 6. The spectra values of the test set were rescaled by applying the average value and standard deviation of the training set. For this experiment, the results were evaluated for all four configurations by evaluation metrics.

Second experiment: Predictions on the S2 spectra dataset and on the S2 image dataset by using for the test set the feature scaling parameters of the training set

In the second experiment, the test set included spectra from the S2 spectra dataset which as already mentioned (section 2.1.3) included cloud and water spectra randomly extracted by the successfully derived masks of the third experiment and manually extracted water signatures. The number of manually and randomly extracted spectra

Table 5. Summary of ANN experiments (training on Hollstein dataset).

Experiment	Training set	Test set	Training set feature scaling	Test set feature scaling
1 st	Hollstein training set	Hollstein test set	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$
2 nd	Hollstein training set	S2 spectra test set S2 image dataset	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$
3 rd	Hollstein training set	S2 image dataset	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$	$z = \frac{x - \mu_{\text{image}}}{\sigma_{\text{image}}}$

Table 6. Spectra comprising the Hollstein test set.

Class	A: Spectra of Hollstein test set	B: Spectra of Hollstein training set	A + B	A as a proportion of total (%)
Water	300,000	1,000,000	1,300,000	23%
Cloud	150,000	500,000	650,000	23%
Cirrus	150,000	500,000	650,000	23%

Table 7. Spectra comprising the S2 spectra test set.

Class	A:S2 spectra test set	B:S2 spectra training set	A + B	A as a proportion of total (%)
Manually extracted water	300,000	500,000	800,000	38%
Randomly extracted cloud	1,040,000	1,000,000	2,040,000	51%
Randomly extracted water	1,040,000	500,000	1,540,000	68%

and their percentage calculated by use of the total number of spectra for each class used for the experiment is presented in Table 7. In the same table, the number of spectra for the training set explained in section 2.3.2.2 is also presented for easier understanding. For the test set of the randomly extracted spectra, it was decided to use the total number of unused remaining spectra of the S2 spectra dataset after subtracting the spectra of the training set. The spectra values of the test set were rescaled by applying the average value and standard deviation of the training set. For this experiment, the results were evaluated for all four configurations by evaluation metrics. In addition, the ANN trained with the first configuration was used to predict the class (cloud/water) of the reflectance signatures for the 79 images of the S2 image dataset. The cloud masks produced by these predictions were evaluated by visual observation.

Third experiment: Predictions on the S2 image dataset by using for the test set the feature scaling parameters of the test set

In the third experiment, the test set consists the 79 images of the S2 image dataset. In this experiment, instead of rescaling the spectra values of the test set by applying the average value and standard deviation of the training set, the predictions on the test set were carried out by rescaling the values with the average value and standard deviation of the images. In more detail, when executing the predictions on the S2 image dataset, the (1,830x1,130) signatures of each image were rescaled according to the average value and standard deviation of this image, i.e. spectra of different images were differently rescaled. The cloud masks produced by these predictions were evaluated by visual observation. It is noted that these cloud masks were produced by the ANN trained with the first configuration. It should be also clarified that a test set consisting only of individual spectra (e.g. the test dataset used in the second experiment) cannot be used in this experiment since the key concept is feature scaling with the average value and standard deviation of the total number of spectra comprising a realistic cloud/water image.

2.3.2.2. Training on the S2 spectra dataset. Besides the Hollstein dataset, spectra from the S2 spectra dataset were also used in the training of all four configurations of the ANNs. This experiment is summarized in Table 8. The training set consisted of randomly extracted cloud signatures, randomly extracted water signatures and manually extracted water spectra. The number of manually and randomly extracted spectra and their percentage calculated by use of the total number of spectra for each class of the S2 spectra

Table 8. Summary of ANN experiment (training on S2 spectra dataset).

Training set	Test set	Training set feature scaling	Test set feature scaling
S2 spectra training set	S2 spectra test set S2 image dataset	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$	$z = \frac{x - \mu_{\text{training}}}{\sigma_{\text{training}}}$

Table 9. Spectra comprising the S2 spectra training set.

Class	A:S2 spectra training set	B:S2 spectra dataset	A as a proportion of total (%)
Manually extracted water	500,000	2,133,324	23%
Randomly extracted cloud	1,000,000	2,040,000	49%
Randomly extracted water	500,000	2,040,000	25%

dataset is presented in Table 9. It was decided that the number of training spectra for each class should be similar to the size of the Hollstein training set since it managed to produce satisfactory results in the first experiment (section 3.1.1.). Moreover, it was considered appropriate to use an equal size of manually extracted water (high noise levels and sunglint) and randomly extracted water (low noise levels/no sunglint presence). From the remaining unused signatures of the S2 spectra dataset, the S2 spectra test set mentioned in the second experiment was created. The spectra values of the test set were rescaled by applying the average value and standard deviation of the training set. The results were evaluated for all four configurations by evaluation metrics. In addition, the ANN trained with the first configuration was used to predict the class of the reflectance signatures for the 79 images of the S2 image dataset. The cloud masks produced by these predictions were evaluated by visual and quantitative comparison with the results produced by the algorithms of Fmask, MAJA, and Sen2Cor.

3. Results

3.1. Results produced by training on the Hollstein dataset

3.1.1. Predictions on the Hollstein dataset by using on the test set the feature scaling parameters of the training set

Accuracy (Equation 15), recall (producer's accuracy) (Equation 16), precision (user's accuracy) (Equation 17) and True Statistic Skill (TSS) (Equation 18) were calculated for the Hollstein training and test set. Recall corresponds to omission error (100%-omission error) while precision corresponds to commission error (100%-commission error). TSS was chosen instead of Cohen's kappa (the most popular measure for the evaluation of presence-absence predictions), since besides taking random agreement into account, it is also independent of prevalence (Allouche, Tsoar, and Kadmon 2006). It is calculated by use of sensitivity (recall) and specificity (True Negative Rate) (Equation 19) and measures interrater reliability (agreement of prediction model with ground truth). Table 10 presents the results of the predictions on the training set, while Table 11 presents the results of the predictions on the test set.

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (15)$$

Table 10. Evaluation metrics of the predictions on the Hollstein training set.

Configuration	TP	FP	FN	TN	Accuracy	Precision	Recall	TSS
1 st	999,856	82	144	999,918	0.9999	0.9999	0.9999	0.9998
2 nd	997,556	365	2,444	999,635	0.9986	0.9996	0.9976	0.9972
3 rd	998,968	44	1,032	999,956	0.9995	1.0000	0.9990	0.9989
4 th	997,973	473	2,027	999,527	0.9988	0.9995	0.9980	0.9975

Table 11. Evaluation metrics of the predictions on the Hollstein test set.

Configuration	TP	FP	FN	TN	Accuracy	Precision	Recall	TSS
1 st	299,963	19	37	299,981	0.9999	0.9999	0.9999	0.9998
2 nd	299,319	112	681	299,888	0.9987	0.9996	0.9977	0.9974
3 rd	299,737	14	263	299,986	0.9995	1.0000	0.9991	0.9991
4 th	299,427	148	573	299,852	0.9988	0.9995	0.9981	0.9976

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (17)$$

$$\text{TSS} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{(\text{TP} + \text{FN}) \times (\text{FP} + \text{TN})} = \text{sensitivity} + \text{specificity} - 1 \quad (18)$$

$$\text{specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (19)$$

where TP: true positives, TN: true negatives, FP: false positives and FN: false negatives.

It was observed that the evaluation metrics were very high for all four configurations, both for the training set and the test set.

3.1.2. Predictions on the S2 spectra dataset and on the S2 image dataset by using on the test set the feature scaling parameters of the training set

Accuracy, recall, precision, and TSS were calculated for the S2 spectra test set for all four configurations. From the evaluation metrics that are presented in Table 12, it was observed that a high number of water spectra was falsely classified as cloud (FN). The ANN trained with the first configuration was also used to predict the class of the reflectance signatures for the 79 images of the S2 image dataset. The cloud masks produced by these predictions were visually evaluated and in the majority of the images a large commission error was observed as expected by the evaluation metrics of Table 12 (Figure 7(e,f)). In more detail, the values of precision show a minimum commission error of 24%, which corresponds to the number of water pixels being incorrectly classified as cloud pixels. The values of recall show that the omission error is low, i.e. almost all the cloud pixels were correctly classified. Finally, the low TSS values confirm further the low reliability of the model.

These results led to the conclusion that the S2 image dataset cannot be adequately represented by the Hollstein dataset. As it can be noticed in Figure 1, the majority of the spectra has been collected from inland and coastal areas, while spectra from deep water

Table 12. Evaluation metrics of the predictions on the S2 spectra test set.

Configuration	TP	FP	FN	TN	Accuracy	Precision	Recall	TSS
1 st	1,009,418	340,241	30,582	999,759	0.8442	0.7479	0.9706	0.7167
2 nd	1,000,333	295,228	39,667	1,044,772	0.8593	0.7721	0.9619	0.7415
3 rd	1,003,491	320,519	36,509	1,019,481	0.8500	0.7579	0.9649	0.7257
4 th	1,003,853	320,751	36,147	1,019,249	0.8500	0.7579	0.9652	0.7259

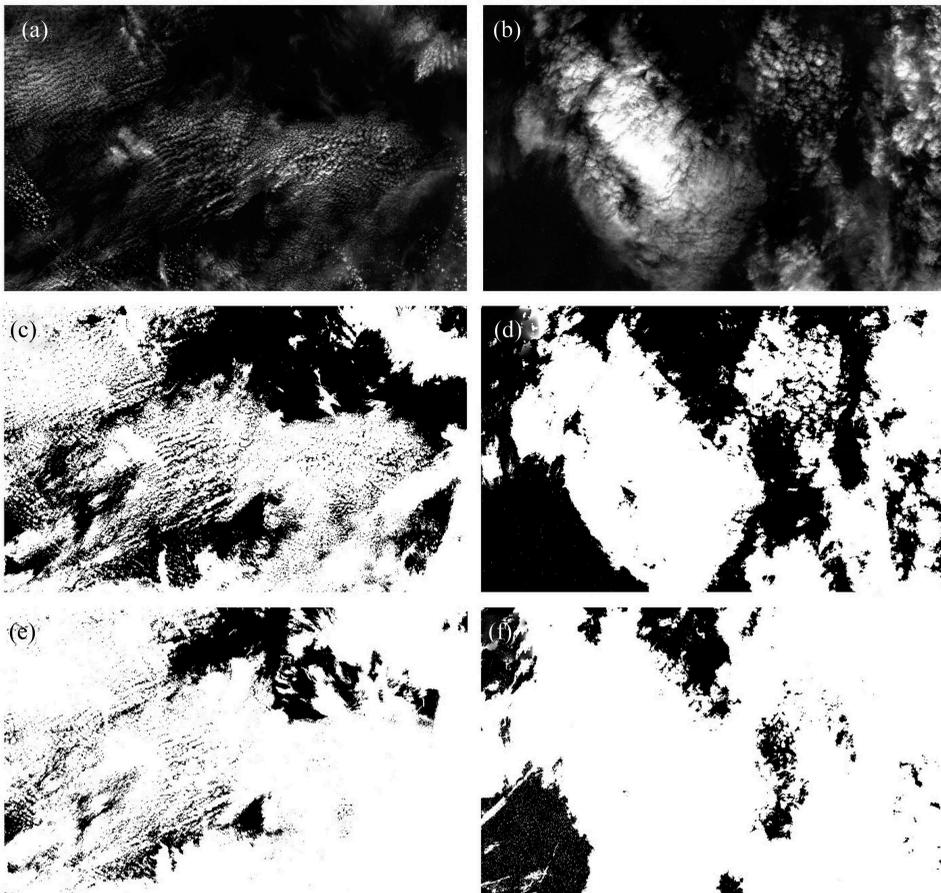


Figure 7. (a,b): 4-3-2 (RGB) natural colour composite, (c,d): cloud mask produced by using on the test set the feature scaling parameters of the test set, (e,f): cloud mask produced by using on the test set the feature scaling parameters of the training set. The size of all figures is $109.8 \times 67.8 \text{ km}^2$.

areas are scarce. As a result, it could be naturally concluded that water spectra with high noise levels and sunglint are scarce in the Hollstein dataset as well.

3.1.3. Predictions on the S2 image dataset by using for the test set the feature scaling parameters of the test set

The predictions on the S2 image set of the ANN trained with the first configuration were evaluated by visual observation and it was observed that for 34 images, the produced cloud masks were satisfactory (Figure 7(c,d)), while the cloud masks produced on the rest 45 images were characterized by very high commission error (Figure 8(c,d)). The successful results for the 34 images led to the conclusion that the feature scaling process applied on these images (use of the parameters of the images instead of those of the training set) created spectra with statistical parameters similar to those of the Hollstein dataset. As mentioned in section 2.1.3, spectra from these 34 cloud masks were randomly extracted and formed part of the S2 spectra dataset. Furthermore, it was observed that the majority of the 45 images (42/45) had high levels of oblique periodic noise in band 10 (cirrus band/ $1.375 \mu\text{m}$) (Figure 9(c,d)) in contrast

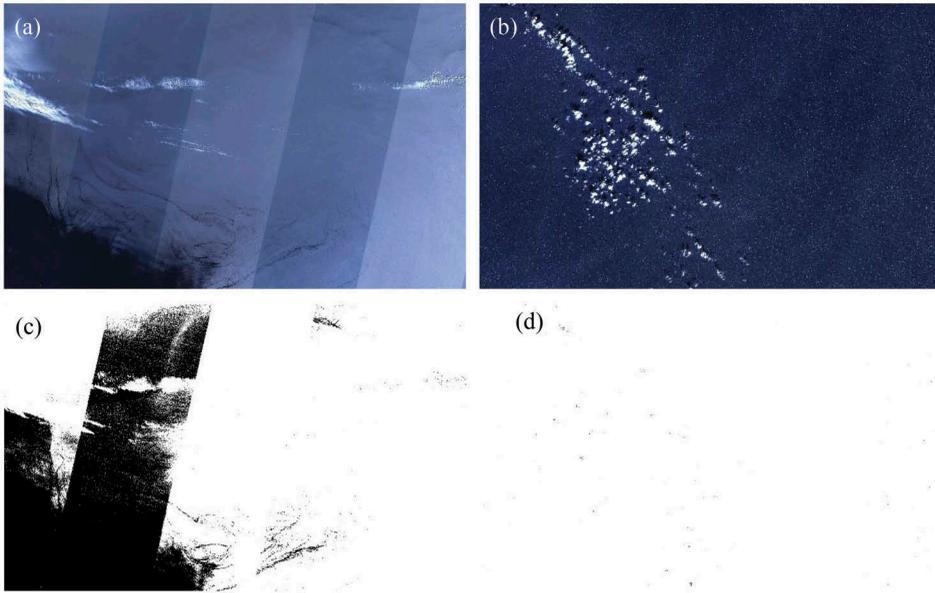


Figure 8. (a,b): 4-3-2 (RGB) natural colour composite, (c,d): cloud mask produced by using on the test set the feature scaling parameters of the test set. The size of all the figures is $109.8 \times 67.8 \text{ km}^2$.

with the majority of the 34 images (31/34) (Figure 10(c,d)) which either depicted very low levels of oblique periodic noise or none. The magnitude of band 10 (after the implementation of Fast Fourier Transformation) (Abu-Ein 2014) is presented in Figures 9(e,f) and 10(e,f). It was also observed that the vast majority of the noisy images had much lower average reflectance values in band 10 (Figure 11). It is noted that the 30 images from which the manually extracted spectra of the S2 dataset were extracted formed part of the 45 images mentioned above.

3.1.4. Observation of the weights of the first hidden layer

The weights of the first hidden layer for the four configurations were observed for the ANNs trained on the Hollstein dataset since they represent the importance of the bands for the ANN. Table 13 is created by calculating the importance of the bands which was defined as the sum of the absolute values of the 20 weights (equal to the number of neurons) corresponding to each of the 13 bands (Equation 20). This table shows for each configuration in descending order the importance of the bands.

$$\text{Im}_j = \sum_{i=1}^{20} |w_{i,j}| \quad (20) \quad j \in [1, 13]$$

where $w_{1,1}, \dots, w_{20,13}$ are the weights of the first hidden layer.

It was observed that band 11 ($1.6 \mu\text{m}$) which is primarily used for cloud separation in turbid waters was given high weights in all configurations. As far as the rest of the bands are concerned, the ranking of importance greatly varied since as described in the Introduction, for the detection of clouds, a variety of VISNIR and SWIR bands has proven to be useful.

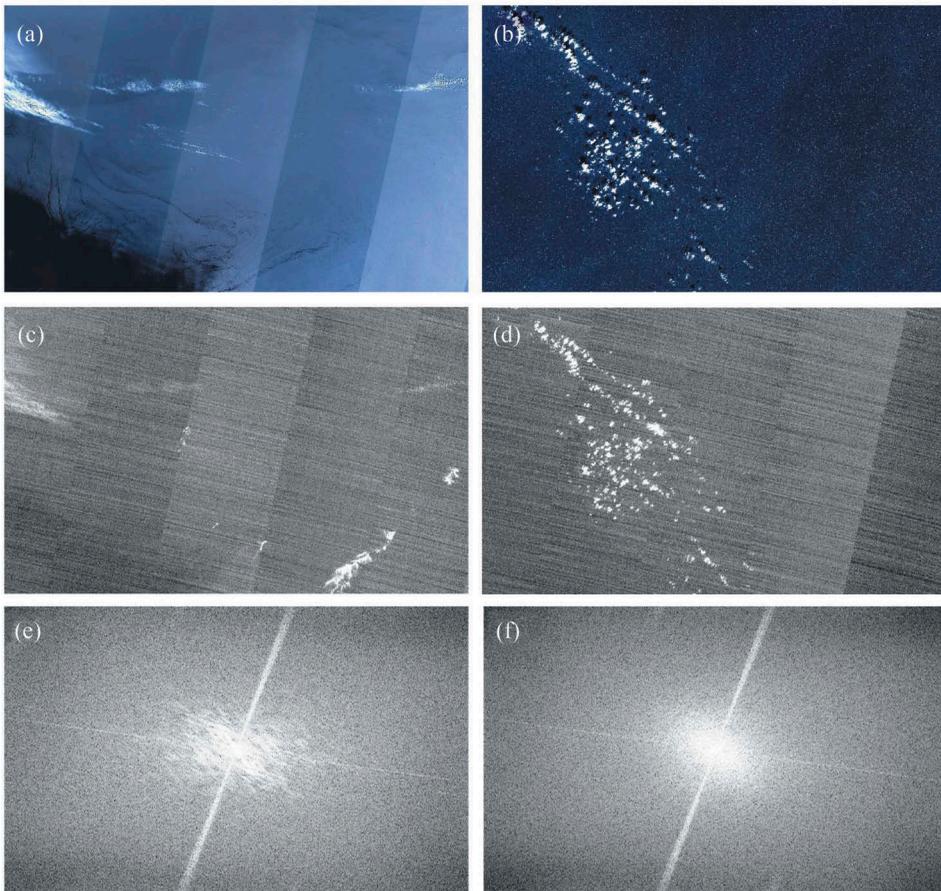


Figure 9. (a,b): 4-3-2 (RGB) natural colour composite, (c,d): cirrus band (1.375 μm), (e,f): magnitude of cirrus band. The size of all figures is $109.8 \times 67.8 \text{ km}^2$.

3.2. Results produced by training on the S2 spectra dataset

3.2.1. Predictions on the S2 spectra dataset and on the S2 image dataset by using on the test set the feature scaling parameters of the training set

This section presents the results of the predictions of the ANNs trained on the S2 spectra dataset. Evaluation metrics were calculated for the S2 spectra training set and the S2 spectra test set. Table 14 presents the results of the predictions on the training set, while Table 15 presents the results of the predictions on the test set.

It was observed that the evaluation metrics were high in all four configurations, both for the training set and the test set. Moreover, the ANN of the first configuration demonstrated the maximum values of accuracy, recall, precision, and TSS. The accuracy of this ANN on the test set was 92% and the TSS value was 0.86 which shows high model reliability. In addition, the recall value ($\sim 96\%$) shows that around 4% of cloud spectra were incorrectly classified to the water class (omission error), while the precision value ($\sim 88\%$) shows that 12% of water spectra were incorrectly classified to the cloud class (commission error).

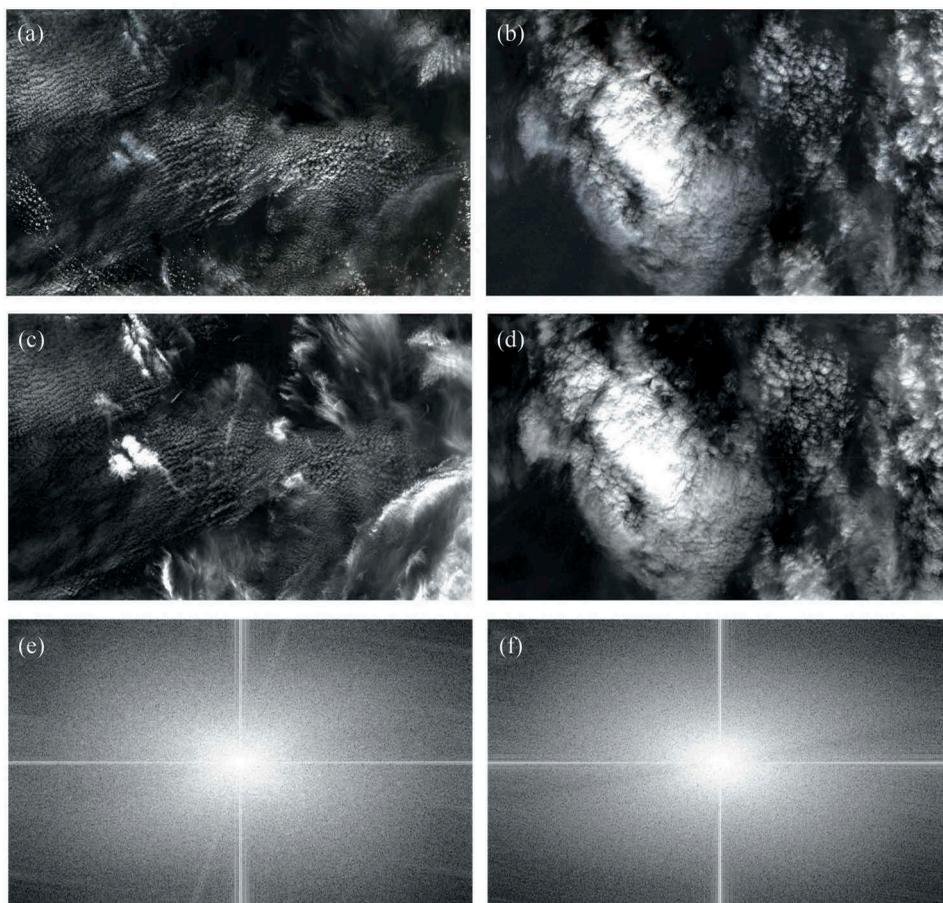


Figure 10. (a,b): 4-3-2 (RGB) natural colour composite, (c,d): cirrus band (1.375 μm), (e,f): magnitude of cirrus band. The size of all figures is $109.8 \times 67.8 \text{ km}^2$.

The ANN of the first configuration was also used to predict the class of the reflectance signatures for the 79 images of the S2 image dataset. The results (Figures 13(c,d), 14(c,d) and 15(c)) were at first evaluated by visual observation and were compared with the results produced by the algorithms of Fmask (Figures 13(e,f), 14(e,f) and 15(d)), MAJA (Figures 13(g,h), 14(c,d) and 15(e)) and Sen2Cor (Figures 13(i,j), 14(i,j) and 15(f)). In addition, quantitative evaluation was applied by the calculation of evaluation metrics which included a) accuracy, recall, precision and TSS scores for the total number of the spectra of the S2 spectra dataset (labelled pixels) for the ANN and the three state-of-the-art algorithms and b) the phi coefficient (measures the degree of association between 2 binary variables (Cramer 1946)) (Equation 21) between the masks produced by the ANN and the respective masks produced by the above-mentioned algorithms. It is noted that Figure 13 presents images with low noise levels and no sunglint presence while Figures 14 and 15 present more difficult cases (high noise levels and sunglint presence).

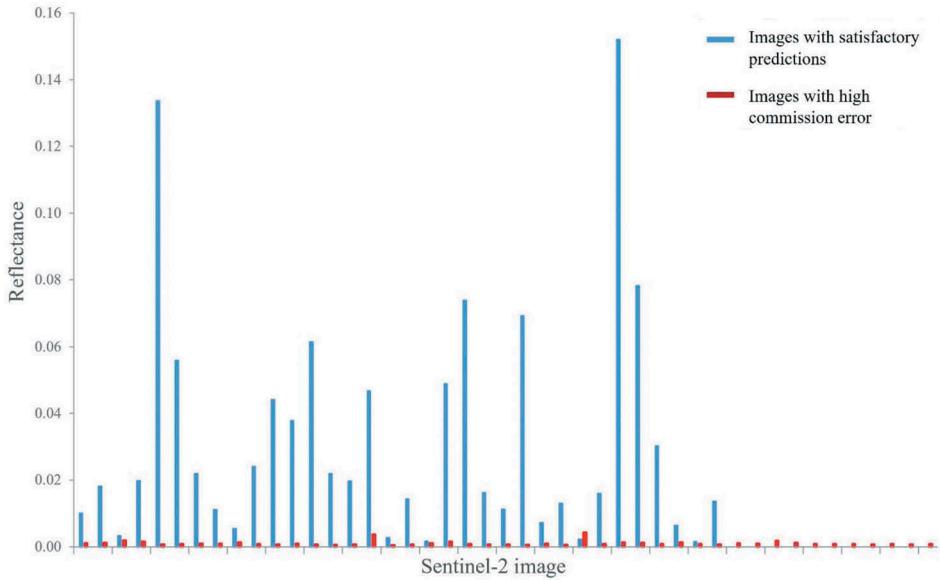


Figure 11. Average reflectance values of band 10 (1.375 μm) for the 79 Sentinel-2 images. (blue): 34 images with satisfactory predictions. (red): 45 images with high commission error.

Table 13. Importance of the Sentinel 2 bands for the ANNs trained on the Hollstein dataset.

1 st configuration		2 nd configuration		3 rd configuration		4 th configuration	
Bands	Importance	Bands	Importance	Bands	Importance	Bands	Importance
11	25.4059	11	18.7082	10	0.8123	11	0.4035
2	14.7908	2	8.8081	7	0.2332	12	0.1133
8A	13.1052	1	7.4302	11	0.2242	8	0.0949
1	8.4542	12	7.2015	8	0.1152	2	0.0867
8	7.4449	4	6.3049	1	0.0613	10	0.0784
10	7.0951	10	5.7476	6	0.0078	7	0.0691
3	6.9086	8	5.5324	5	0.0034	9	0.0606
12	5.08	8A	4.8357	2	0.0031	4	0.0595
9	4.7732	5	4.7502	9	0.003	6	0.0325
5	4.6562	3	4.7488	3	0.0028	5	0.0243
6	4.5948	7	4.3152	4	0.0027	1	0.0147
7	4.3817	9	4.1179	12	0.0023	8A	0.0095
4	4.3519	6	3.926	8A	0	3	0.0094

Table 14. Evaluation metrics of the predictions on the S2 spectra training set.

Configuration	TP	FP	FN	TN	Accuracy	Precision	Recall	TSS
1 st	960,126	66,741	39,874	933,259	0.9467	0.9350	0.9601	0.8934
2 nd	950,654	73,443	49,346	926,557	0.9386	0.9283	0.9507	0.8772
3 rd	937,267	68,792	62,733	931,208	0.9342	0.9316	0.9373	0.8685
4 th	934,203	84,061	65,797	915,939	0.9251	0.9174	0.9342	0.8501

$$\Phi = \frac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}} \quad (21)$$

Table 15. Evaluation metrics of the predictions on the S2 spectra test set.

Configuration	TP	FP	FN	TN	Accuracy	Precision	Recall	TSS
1 st	998,557	139,311	41,443	1,200,689	0.9241	0.8776	0.9602	0.8562
2 nd	988,743	153,633	51,257	1,186,367	0.9139	0.8655	0.9507	0.8361
3 rd	974,455	144,078	65,545	1,195,922	0.9119	0.8712	0.9370	0.8295
4 th	971,340	175,500	68,660	1,164,500	0.8974	0.8470	0.9340	0.8030

where A , D are the diagonal values of a 2×2 contingency table and B , C are the non-diagonal values.

Concerning the visual evaluation, the results of the ANN in all of the 79 Sentinel-2 images were considered to be very favourable compared to the above algorithms. ANN results were acceptable in all cases and unaffected by water areas with high noise levels and sunglint. In addition, the ANN proved to be robust since the results were homogenous and none of the 79 cases presented outlier classification output, i.e. classifying areas with opaque clouds as water (often observed in Sen2Cor masks) or classifying whole or large part of strips as cloud. It should be noted though that a small omission error was usually observed. The masks produced by the Sen2Cor algorithm demonstrated in general the least satisfactory results since they presented an overall omission error which in several cases was high. Fmask cloud masks showed better results than those of Sen2Cor but were characterized by commission error which in a few cases of water areas with high presence of sunglint was high (Figure 14(e)). MAJA masks presented in our opinion best results than Sen2Cor and Fmask, although it should be stated that a small commission error was usually observed.

The images depicted in Figures 13, 14 and 15 represent the results for the different types of cases of the study. For the case depicted in Figure 13(a), the ANN mask (Figure 13(c)) represented sufficiently the cloud presence of the image, the Fmask mask (Figure 13(e)) demonstrated high commission error, the MAJA mask (Figure 13(g)) showed high similarity with the ANN mask and Sen2Cor (Figure 13(i)) presented very high omission error. For the case depicted in Figure 13(b), the ANN mask (Figure 13(d)) showed the most acceptable results, the MAJA mask (Figure 13(h)) presented a commission error which was higher for the Fmask mask (Figure 13(f)) and Sen2Cor (Figure 13(j)) presented very high omission error. Regarding the cases with high noise levels and sunglint, for the case of Figure 14(a), the ANN mask (Figure 14(c)) was overall satisfactory since it was unaffected by sunglint and the oblique periodic noise but it omitted a few thin clouds, the Fmask mask (Figure 14(e)) incorrectly classified a large percentage of sunglint areas as clouds, the MAJA mask (Figure 14(g)) was similar with the ANN mask but omitted a higher cloud percentage and the Sen2Cor mask (Figure 14(i)) presented high omission error. For the case of Figure 14(b), the ANN mask (Figure 14(d)) was overall acceptable since it was unaffected by the random noise, but slightly underestimated the cloud presence, the Fmask mask (Figure 14(f)) presented a commission error which was higher for the MAJA mask (Figure 14(h)) and Sen2Cor (Figure 14(j)) presented an omission error. Finally, for the case of Figure 15, the ANN and MAJA masks (Figure 15(c,e)) were unaffected by sunglint, while Fmask (Figure 15(c)), incorrectly classified a large sunglint area as cloud. Sen2Cor (Figure 15(f)) also misclassified a few sunglint pixels to the cloud category. Besides the above observations, the ANN masks seem to be the ones that better represent the natural shape of the clouds, since MAJA masks present the appearance of globs, while Sen2Cor masks show linear structure.

Table 16. Accuracy, precision, recall and TSS scores for the S2 spectra dataset (Comparison of algorithms).

Method	TP	FP	FN	TN	Accuracy	Precision	Recall	TSS
ANN (1 st configuration)	1,958,683	273,747	81,317	3,899,577	0.9429	0.8774	0.9601	0.8945
Fmask	1,989,338	1,085,049	50,662	3,088,275	0.8172	0.6471	0.9752	0.7152
MAJA	1,973,030	703,422	66,970	3,469,902	0.8760	0.7372	0.9672	0.7986
Sen2Cor	1,383,951	456,874	656,049	3,716,450	0.8209	0.7518	0.6784	0.5689

For the quantitative evaluation, evaluation metrics were at first calculated for the total number of the spectra of the S2 spectra dataset (Table 16).

It was observed that the ANN showed the highest accuracy/TSS scores (~94%/0.89) followed by MAJA (~88%/0.80). The respective scores for Fmask (~82%/0.72) and Sen2Cor (~82%/0.57) were lower with Sen2Cor showing the minimum TSS value. Concerning recall values, except for Sen2Cor which showed high omission error (32%), the other algorithms produced low omission errors (~3%). Finally, regarding precision values, the ANN showed the highest value (~88%) corresponding to the lowest commission error of 12%. Sen2Cor and MAJA produced similar values (~75%/74%) while Fmask showed the lowest score (65%).

The phi coefficient was also calculated. Values over 0.4 are considered to show strong positive correlation while values over 0.7 show a very strong positive correlation. Figure 12 shows the phi coefficient values for the cloud masks of the S2 image dataset.

It was observed that the mean value of the phi coefficient between the ANN masks and Fmask masks was 0.58 with a standard deviation of 0.19. Regarding the comparison with the MAJA masks, the mean value of the phi coefficient was 0.65 with a standard deviation of 0.19. Finally, the mean value of the phi coefficient between the ANN masks and Sen2Cor masks was 0.44 with a standard deviation of 0.27. Thus, MAJA masks are more

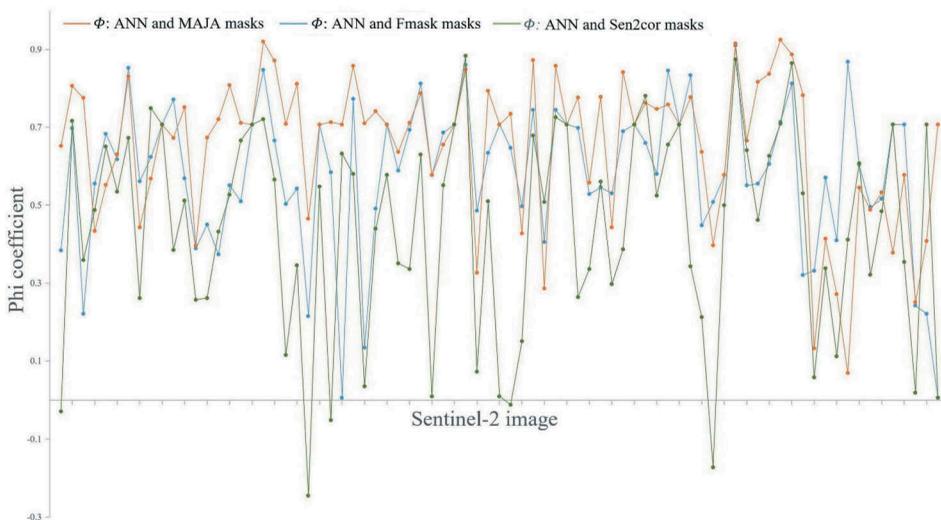


Figure 12. Phi coefficient for the cloud masks of the S2 image dataset. Brown: phi coefficient between ANN and MAJA masks, Blue: phi coefficient between ANN and Fmask masks, Green: phi coefficient between ANN and Sen2cor masks.

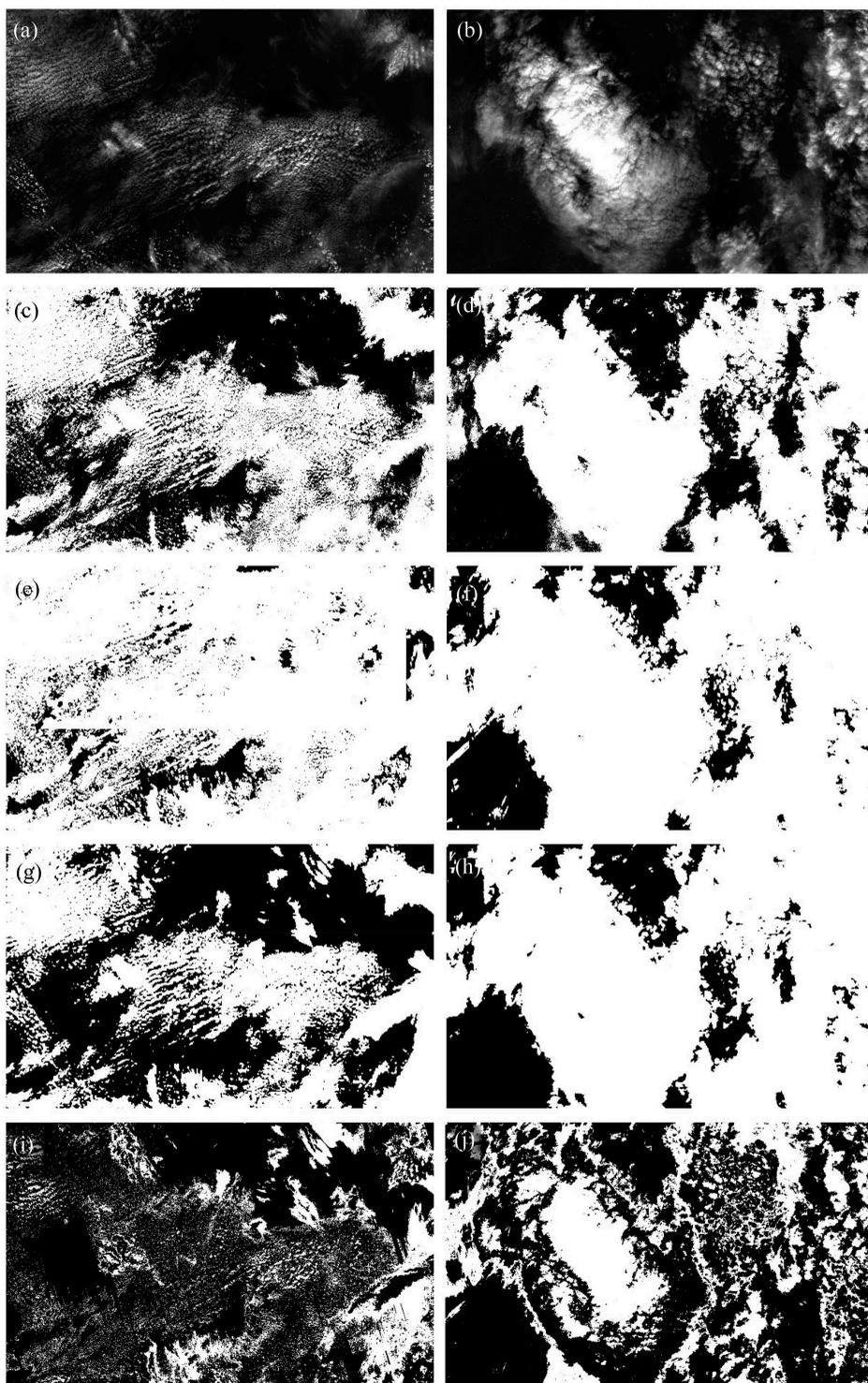


Figure 13. (a,b): 4-3-2 (RGB) natural colour composite, (c,d): ANN cloud mask, (e,f): Fmask cloud mask, (g,h): MAJA cloud mask, (i,j): Sen2Cor cloud mask. The size of all figures is $109.8 \times 67.8 \text{ km}^2$.

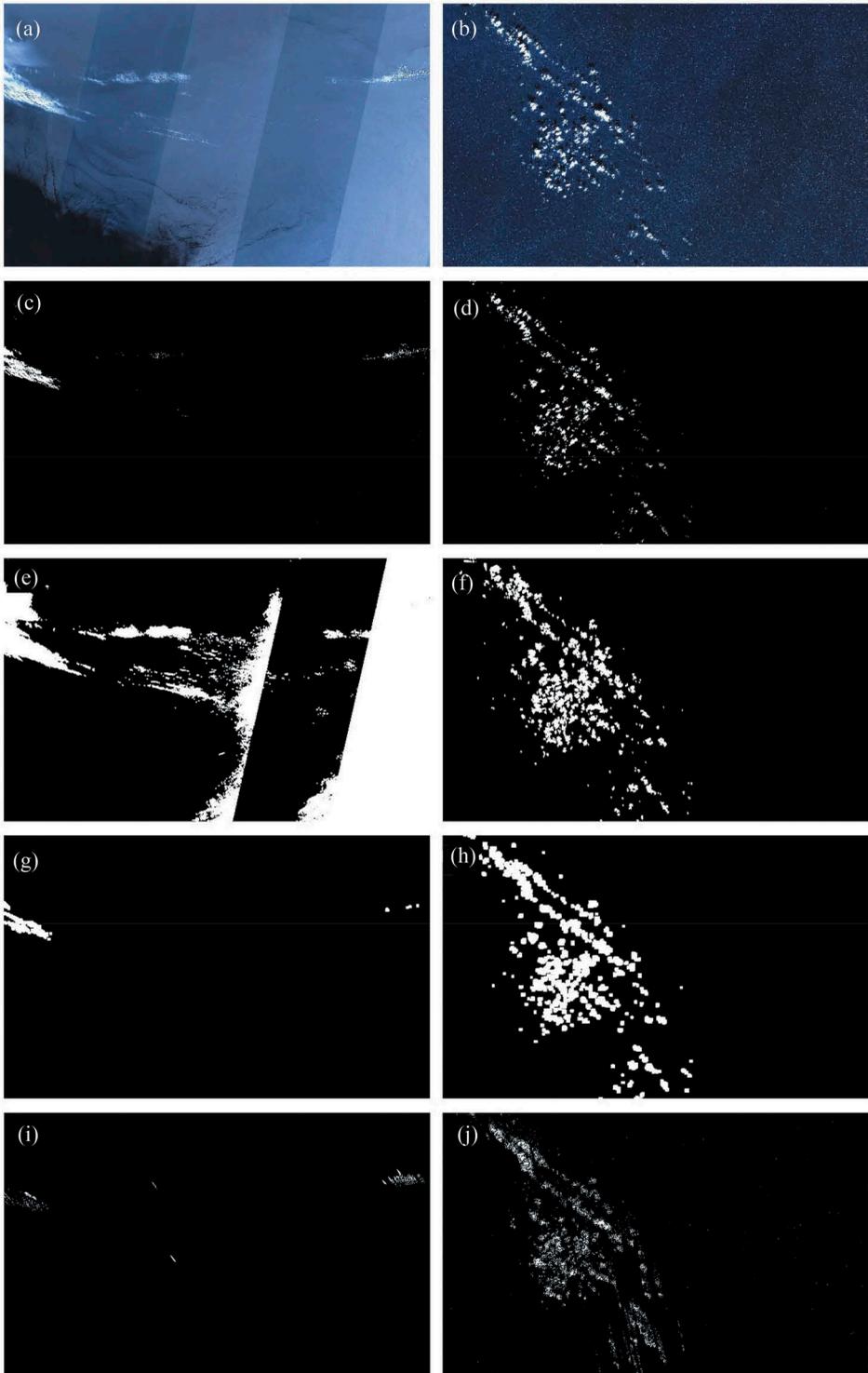


Figure 14. (a,b): 4-3-2 (RGB) natural colour composite, (c,d): ANN cloud mask, (e,f): Fmask cloud mask, (g,h): MAJA cloud mask, (i,j): Sen2Cor cloud mask. The size of all figures is $109.8 \times 67.8 \text{ km}^2$.

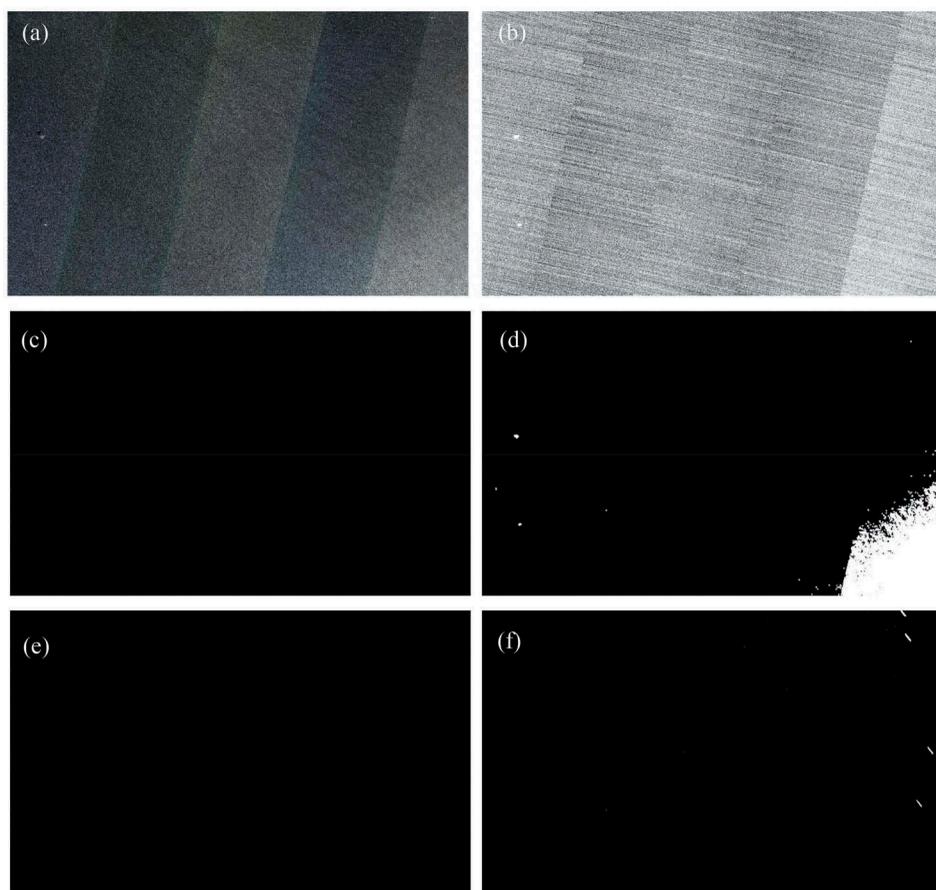


Figure 15. (a): 4-3-2 (RGB) natural colour composite, (b): magnitude of cirrus band, (c): ANN cloud mask, (d): Fmask cloud mask, (e): MAJA cloud mask, (f): Sen2Cor cloud mask. The size of all the figures is $109.8 \times 67.8 \text{ km}^2$.

Table 17. The phi coefficient values for the masks of the images depicted in Figure 13, 14, 15.

Cloud mask	Phi Fmask	Phi MAJA	Phi Sen2Cor
Figure 13(a)	+0.38	+0.65	-0.03
Figure 13(b)	+0.65	+0.73	-0.01
Figure 14(a)	+0.13	+0.71	+0.03
Figure 14(b)	+0.53	+0.44	+0.29
Figure 15(a)	+0.00	+0.71	+0.63

positively correlated with the ANN masks followed by Fmask and Sen2Cor. Table 17 shows the phi coefficient values between the ANN masks and the masks of the other algorithms for the masks depicted in Figures 13, 14 and 15.

It was observed that the ANN mask (Figure 13(c)) of Figure 13(a) showed strong positive correlation with the MAJA mask (Figure 13(g)), lower positive correlation with the Fmask mask (Figure 13(e)) and negative correlation with Sen2Cor mask (Figure 13(i)). In addition, the ANN mask (Figure 13(d)) of Figure 13(b) showed very strong positive correlation with the

Table 18. Importance of the Sentinel 2 bands for the ANNs trained on the S2 spectra dataset.

1 st configuration		2 nd configuration		3 rd configuration		4 th configuration	
Bands	Importance	Bands	Importance	Bands	Importance	Bands	Importance
1	14.5757	1	11.0918	10	1.4623	10	1.0388
9	11.3767	10	10.9864	1	1.2076	11	0.7415
11	10.2176	9	10.6795	11	0.54	1	0.7076
10	8.9378	11	10.1314	12	0.4906	9	0.582
12	8.7829	3	7.5562	9	0.1506	8A	0.4642
3	6.9809	12	6.7347	8	0.0944	12	0.4026
8A	6.9176	8	6.1217	3	0.0889	8	0.3834
5	5.4747	8A	5.3496	8A	0.0817	7	0.345
8	5.4138	4	4.1833	4	0.0747	2	0.3381
7	4.6918	5	4.1448	7	0.0084	3	0.3071
2	4.5922	2	4.0972	2	0.0077	6	0.305
6	3.9374	7	3.4356	5	0.0074	5	0.3047
4	3.718	6	1.7959	6	0.0066	4	0.2559

MAJA mask (Figure 13(h)), strong correlation with the Fmask mask (Figure 13(f)) and negative correlation with the Sen2Cor mask (Figure 13(j)). Concerning the images with high noise levels and sunglint, the ANN mask (Figure 14(c)) of Figure 14(a) showed strong positive correlation with the MAJA mask (Figure 14(g)) and very low positive correlation with the Fmask mask (Figure 14(e)) and the Sen2Cor mask (Figure 14(i)), while the ANN mask (Figure 14(d)) of Figure 14(b) showed strong positive correlation for Fmask (Figures 14(f)) and MAJA (Figure 14(h)) masks and lower for Sen2Cor mask. In addition, the ANN mask (Figure 15(c)) of Figure 15(a) showed strong positive correlation with the MAJA mask (Figure 15(e)) followed by the Sen2Cor mask (Figure 15(f)), and no correlation with the Fmask mask (Figure 15(d)).

From the above, it is concluded that the quantitative evaluation is in accordance with the visual evaluation. In more detail, the ANN masks present the highest accuracy/TSS scores and are more correlated with the MAJA masks which present the second best highest accuracy. In addition, the least satisfactory results are presented by Sen2Cor masks and Fmask presents the highest commission error. It should be also noted that an additional advantage of the ANN is that it is more time-efficient than Fmask, MAJA, and Sen2Cor since the mask can be created in seconds (inference time), while the other algorithms need at least 10 min (STEP 2016a, 2016b; CESBIO 2018).

3.2.2. Observation of the weights of the first hidden layer

The weights of the first hidden layer for the four configurations were also observed for the ANNs trained on the S2 spectra dataset since as already mentioned they represent the importance of the bands for the ANN. Table 18 is created in the same way as Table 13 and likewise shows for each configuration in descending order the importance of the bands (Equation 20).

It was observed that during the training with this dataset, band 11 (1.6 μm) is given high weights in all configurations, a behaviour which is similar to the one previously observed on the training on the Hollstein dataset. However, it was also observed that the cirrus band (1.375 μm) acquired high ranking, a behaviour that can be explained by the fact that this band is less affected by sunglint since it corresponds to a strong absorption band of water vapour. High clouds have a high chance of being visible in

this band in contrast to low clouds because incident and reflected light are highly absorbed. Bands 1 (coastal aerosol band/443 nm) and 9 (water vapour absorption band/940 nm) which are typically used for atmospheric correction purposes acquired high ranking as well. Compared to the ranking of the importance of the bands on the training on the Hollstein dataset, these changes directly indicate the bands that counteract the influence of the presence of spectra with high noise levels and sunglint in the training set.

4. Conclusions and discussion

In this study, ANNs were trained with four different configurations on a dataset extracted by Hollstein et al. (2016) database and on a dataset that was created by the extraction of spectra from 79 Sentinel-2 level 1C images that were used for this study. The second dataset adequately represented different types of water, namely, it included water with high noise levels and sunglint. The four configurations were differentiated by the use of different algorithms that prevent overfitting but it was observed that these algorithms slightly affected the performance of the ANNs. Since the configuration of the MSI Sentinel-2 imaging mission leads to a broad range of viewing geometries, the developed ANN was tested for directional reflectance effects only by the use of spectral information.

The ANNs trained on the Hollstein dataset were used in three experiments which were differentiated by the test set used for the predictions and by the feature scaling parameters of the test set. When the test set consisted of spectra from the Hollstein dataset, the evaluation metrics were over 0.99 in all configurations. However, the cloud masks of the Sentinel-2 images produced by using the ANN of the first configuration presented high commission errors when applying on the spectra of the images the feature scaling parameters of the training set. This behaviour leads to the conclusion that the dataset of the Sentinel-2 images used in this study cannot be adequately represented by the Hollstein dataset, since spectra from deep water areas or spectra with high noise levels and sunglint are scarce. Interesting results were produced when applying on the spectra of the images the feature scaling parameters that corresponded to the images instead of those of the training set, where acceptable results were produced for 34 images, which showed that using the feature scaling parameters of the test set could be a factor that could alter the predictions of an ANN towards a positive direction.

The overall accuracy of the ANN of the first configuration trained on the spectra dataset extracted from Sentinel-2 images with several levels of noise and sunglint was over 0.92 on the test set. In addition, the TSS score was over 0.89. The predictions of this ANN were evaluated by visual observation and compared with the results produced by three state-of-the-art algorithms: Fmask, MAJA, and Sen2Cor. Its predictions were considered to be very favourable compared to the above mentioned state-of-the-art algorithms. It produced robust results since none of the 79 cases presented outlier classification output and it proved to be unaffected by water areas with high noise levels and sunglint. In addition, its masks better represented the natural shape of the clouds. Although a small omission error was overall observed, the results were acceptable in all cases. A quantitative evaluation was also performed which was in accordance with the visual and showed that the ANN produced the highest accuracy/TSS scores and presented the strongest correlation with the MAJA masks followed by Fmask and Sen2Cor. As a general conclusion, this ANN

showed the best performance not only concerning the quality of the masks, but also the time efficiency.

The weights of the first hidden layer for the four configurations were observed since they represent the importance of the bands for the ANN and a simple importance measure was defined. It was observed that band 11 (1.6 μm) was given high weights in all configurations. In addition, when observing the weights of the ANNs trained on the dataset with high noise levels and sunglint extracted from the Sentinel-2 images, it was demonstrated that the cirrus band which is less affected by sunglint and two bands typically used for atmospheric correction (443 and 940 nm) are the ones responsible for the mitigation of high noise levels and sunglint in the training set. Regarding the rest of the bands, the ranking of importance greatly varied, a behaviour which can be explained by the fact that a variety of VISNIR and SWIR bands carry useful information for cloud detection applications.

There is no perfect method to mask all clouds and retain water pixels, but this study proved that ANNs are a simple, fast, and effective cloud masking algorithm that can avoid the influence of deep water areas with high noise levels and sunglint. The developed ANN successfully detects clouds on Sentinel-2 images which present serious directional reflectance effects. It also showed that the database created by Hollstein et al. (2016) needs to be expanded with more spectra from deep water areas since successful ANN results are closely connected with training on a dataset that adequately represents the wide variability of cloud and water spectra. Finally, it was shown that there are cases where making predictions using the feature scaling parameters of the test set instead of those of the training set can improve the ANN results when the test set is not adequately represented by the training set. The possibility of generalizing this finding in other applications will be further investigated in our future work.

Data availability statement

The manual database created for the purpose of this study will be published in the public repository: 'Figshare' (<https://doi.org/10.6084/m9.figshare.8075396>) under CC0 licence after the publication of the article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was conducted in the framework of the SEO-DWARF project (<https://cordis.europa.eu/project/id/691071>) supported by H2020 Marie Skłodowska-Curie Actions (<http://dx.doi.org/10.13039/100010665>).

ORCID

Vassilia Karathanassi  <http://orcid.org/0000-0002-8834-4734>

References

- Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin et al. 2016. "Tensorflow: A System for Large-scale Machine Learning." *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*: 265–283. <https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf>
- Abu-Ein, A. A. K. H. 2014. "A Novel Methodology for Digital Removal of Periodic Noise Using 2D Fast Fourier Transforms." *Contemporary Engineering Sciences* 7 (3): 103–116. doi:10.12988/ces.2014.31065.
- Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley. 1998. "Discriminating Clear Sky from Clouds with MODIS." *Journal of Geophysical Research: Atmospheres* 103 (D24): 32141–32157. doi:10.1029/1998JD200032.
- Ackerman, S. A., R. E. Holz, R. Frey, E. W. Eloranta, B. C. Maddux, and M. McGill. 2008. "Cloud Detection with MODIS. Part II: Validation." *Journal of Atmospheric and Oceanic Technology* 25 (7): 1073–1086. doi:10.1175/2007JTECHA1053.1.
- Agarap, A. F. 2018. "Deep Learning Using Rectified Linear Units (Relu)." 1(March): 2–8. <http://arxiv.org/abs/1803.08375>
- Ahmad, A., and S. Quegan. 2014. "Multitemporal Cloud Detection and Masking Using MODIS Data." *Applied Mathematical Sciences* 8 (7): 345–353. doi:10.12988/ams.2014.311619.
- Allouche, O., A. Tsoar, and R. Kadmon. 2006. "Assessing the Accuracy of Species Distribution Models: Prevalence, Kappa and the True Skill Statistic (TSS)." *Journal of Applied Ecology* 43: 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x.
- Bai, T., D. Li, K. Sun, Y. Chen, and W. Li. 2016. "Cloud Detection for High-Resolution Satellite Imagery Using Machine Learning and Multi-Feature Fusion." *Remote Sensing* 8 (9): 715. doi:10.3390/rs8090715.
- Banks, A. C., and F. Mélin. 2015. "An Assessment of Cloud Masking Schemes for Satellite Ocean Colour Data of Marine Optical Extremes." *International Journal of Remote Sensing* 36 (3): 797–821. doi:10.1080/01431161.2014.1001085.
- Baseski, E., and C. Cenaras. 2015. "Texture and Color Based Cloud Detection." *2015 7th International Conference on Recent Advances in Space Technologies (RAST)*, 311–315. IEEE. doi:10.1109/RAST.2015.7208361.
- Bishop, C. M. 1995. "Training with Noise Is Equivalent to Tikhonov Regularization." *Neural Computation* 7: 108–116. doi:10.1162/neco.1995.7.1.108.
- Candra, D. S., S. Phinn, and P. Scarth. 2016. "Cloud and Cloud Shadow Masking Using Multi-Temporal Cloud Masking Algorithm in Tropical Environmental." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B2* July: 95–100. doi:10.5194/isprsarchives-XLI-B2-95-2016.
- CESBIO. 2018. "On-demand SENTINEL2 L2A processing with MAJA on PEPS" <https://labo.obs-mip.fr/multitemp/on-demand-sentinel2-l2a-processing-with-maja-on-peps/>
- Charantonis, A. A., S. Thiria, B. Berthelot, H. Brogniez, and M. Roux. 2009. "Océan, Atmosphère, Climat et Télédétection Section Méthodes physiques en télédétection." *MASTER2 Recherche*. Ecole Polytechnique.
- Chen, S., and T. Zhang. 2015. "An Improved Cloud Masking Algorithm for MODIS Ocean Colour Data Processing." *Remote Sensing Letters* 6 (3). Taylor & Francis: 218–227. doi:10.1080/2150704X.2015.1026955.
- Chollet, F. (2015) "Keras." <https://github.com/fchollet/keras>
- Cox, C., and W. Munk. 1954. "Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter." *Journal of the Optical Society of America* 44: 838. doi:10.1364/josa.44.000838.
- Cramer, H. 1946. "Mathematical Methods of Statistics." Princeton University Press. ISBN: 0-691-08004-6
- ESA. 2019. "MultiSpectral Instrument (MSI) Overview." <https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>
- Fisher, A. 2014. "Cloud and Cloud-Shadow Detection in SPOT5 HRG Imagery with Automated Morphological Feature Extraction." *Remote Sensing* 6 (1): 776–800. doi:10.3390/rs6010776.

- Foga, S., P. L. Scaramuzza, S. Guo, Z. Zhu, R. D. Dilley, T. Beckmann, G. L. Schmidt, J. L. Dwyer, M. J. Hughes, and B. Laue. 2017. "Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products." *Remote Sensing of Environment* 194 (June). Elsevier Inc: 379–390. doi: [10.1016/j.rse.2017.03.026](https://doi.org/10.1016/j.rse.2017.03.026).
- Gallegos, S. C., J. D. Hawkins, and C. F. Cheng. 2008. "A New Automated Method of Cloud Masking for Advanced Very High Resolution Radiometer Full-Resolution Data over the Ocean." *Journal of Geophysical Research: Oceans* 98 (C5): 8505–8516. doi:[10.1029/93jc00100](https://doi.org/10.1029/93jc00100).
- Garaba, S. P., J. Schulz, M. R. Wernand, and O. Zielinski. 2012. "Sunglint Detection for Unmanned and Automated Platforms." *Sensors (Switzerland)* 12 (9): 12545–12561. doi:[10.3390/s120912545](https://doi.org/10.3390/s120912545).
- Gould, Jr., R. A. Arnone, and M. Sydor. 2001. "Absorption, Scattering, and Remote-Sensing Reflectance Relationships in Coastal Waters: Testing a New Inversion Algorithm." *Journal of Coastal Research*. <https://www.jstor.org/stable/4300183>
- Hagolle, O., M. Huc, C. Desjardins, S. Auer, and R. Richter. 2017. "MAJA ATBD Algorithm Theoretical Basis Document." *tech. rep., CNES+ CESBIO and DLR*. https://www.theia-land.fr/sites/default/files/imce/produits/atbd_maja_071217.pdf
- Hagolle, O., M. Huc, D. Villa Pascual, and G. Dedieu. 2010. "A Multi-Temporal Method for Cloud Detection, Applied to FORMOSAT-2, VEN μ S, LANDSAT and SENTINEL-2 Images." *Remote Sensing of Environment* 114 (8). Elsevier Inc.: 1747–1755. doi: [10.1016/j.rse.2010.03.002](https://doi.org/10.1016/j.rse.2010.03.002).
- Hocking, J., P. N. Francis, and R. Saunders. 2011. "Cloud Detection in Meteosat Second Generation Imagery at the Met Office." *Meteorological Applications* 18 (3): 307–323. doi:[10.1002/met.239](https://doi.org/10.1002/met.239).
- Hollstein, A., K. Segl, L. Guanter, M. Brell, and M. Enesco. 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 (8): 666. doi:[10.3390/rs8080666](https://doi.org/10.3390/rs8080666).
- Hu, C. 2011. "An Empirical Approach to Derive MODIS Ocean Color Patterns under Severe Sun Glint." *Geophysical Research Letters* 38: 1. doi:[10.1029/2010GL045422](https://doi.org/10.1029/2010GL045422).
- Hughes, M. J., and D. J. Hayes. 2014. "Automated Detection of Cloud and Cloud Shadow in Single-Date Landsat Imagery Using Neural Networks and Spatial Post-Processing." *Remote Sensing* 6 (6): 4907–4926. doi:[10.3390/rs6064907](https://doi.org/10.3390/rs6064907).
- Iannone, R. Q., F. Niro, P. Goryl, S. Dransfeld, B. Hoersch, K. Stelzer, G. Kirches, et al. 2017. "Proba-V Cloud Detection Round Robin: Validation Results and Recommendations." *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 1–8. IEEE. doi:[10.1109/Multi-Temp.2017.8035219](https://doi.org/10.1109/Multi-Temp.2017.8035219).
- Imai, T., and R. Yoshida. 2016. "Algorithm Theoretical Basis for Himawari-8 Cloud Mask Product." *Meteorological Satellite Center Technical Note* <https://www.data.jma.go.jp/mscweb/technotes/msctechrep61-1.pdf>
- Irish, R. R. 2000. "Landsat 7 Automatic Cloud Cover Assessment." *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI* 4049: 348–356. International Society for Optics and Photonics. doi:[10.1117/12.410358](https://doi.org/10.1117/12.410358).
- Ishida, H., Y. Oishi, K. Morita, K. Moriwaki, and T. Y. Nakajima. 2018. "Development of a Support Vector Machine Based Cloud Detection Method for MODIS with the Adjustability to Various Conditions." *Remote Sensing of Environment* 205. Elsevier: 390–407. doi:[10.1016/j.rse.2017.11.003](https://doi.org/10.1016/j.rse.2017.11.003).
- Karvonen, J. 2014. "Cloud Masking of MODIS Imagery Based on Multitemporal Image Analysis." *International Journal of Remote Sensing* 35 (23). Taylor & Francis: 8008–8024. doi:[10.1080/01431161.2014.978955](https://doi.org/10.1080/01431161.2014.978955).
- Kingma, D. P., and J. L. Ba. 2014. "Adam: A Method for Stochastic Optimization," December: 1–15. <http://arxiv.org/abs/1412.6980>
- Le Goff, M., J.-Y. Tournieret, H. Wendt, M. Ortner, and M. Spigai. 2017. "Deep Learning for Cloud Detection." *8th International Conference of Pattern Recognition Systems (ICPRS 2017)*, 10 (6). Institution of Engineering and Technology. doi:[10.1049/cp.2017.0139](https://doi.org/10.1049/cp.2017.0139).
- Lee, Z., Y.-H. Ahn, C. Mobley, and R. Arnone. 2010. "Removal of Surface-Reflected Light for the Measurement of Remote-Sensing Reflectance from an above-Surface Platform." *Optics Express* 18: 26313. doi:[10.1364/oe.18.026313](https://doi.org/10.1364/oe.18.026313).

- Li, Z., H. Shen, Y. Wei, Q. Cheng, and Q. Yuan. 2018. "Cloud Detection by Fusing Multi-Scale Convolutional Features." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4 (3): 149–152. doi:10.5194/isprs-annals-IV-3-149-2018.
- Lin, C. H., B. B. Y. Lin, K. B. Y. Lee, and Y. C. Chen. 2015. "Radiometric Normalization and Cloud Detection of Optical Satellite Images Using Invariant Pixels." *ISPRS Journal of Photogrammetry and Remote Sensing* 106 (August). International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS): 107–117. doi:10.1016/j.isprsjprs.2015.05.003.
- Martins, J. V., D. Tanré, L. Remer, Y. Kaufman, S. Mattoo, and R. Levy. 2002. "MODIS Cloud Screening for Remote Sensing of Aerosols over Oceans Using Spatial Variability." *Geophysical Research Letters* 29. doi:10.1029/2001GL013252.
- Mateo-García, G., L. Gomez-Chova, and G. Camps-Valls. 2017. "Convolutional Neural Networks for Multispectral Image Cloud Masking." *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2017–July:2255–2258. IEEE. doi:10.1109/IGARSS.2017.8127438.
- Mateo-García, G., L. Gómez-Chova, J. Amorós-López, J. Muñoz-Marí, and G. Camps-Valls. 2018. "Multitemporal Cloud Masking in the Google Earth Engine." *Remote Sensing* 10 (7): 1079. doi:10.3390/rs10071079.
- Mobley, C. D. 1999. "Estimation of the Remote-Sensing Reflectance from above-Surface Measurements." *Applied Optics* 38: 7442. doi:10.1364/ao.38.007442.
- Nakajima, T. Y., H. Ishida, T. M. Nagao, M. Hori, H. Letu, R. Higuchi, N. Tamaru, N. Imoto, and A. Yamazaki. 2019. "Theoretical Basis of the Algorithms and Early Phase Results of the GCOM-C (Shikisai) SGLI Cloud Products." *Progress in Earth and Planetary Science* 6 (1). Progress in Earth and Planetary Science: 1–25. doi:10.1186/s40645-019-0295-9.
- Nicolas, J. M., P. Y. Deschamps, H. Loisel, and C. Moulin. 2005. "Algorithm Theoretical Basis Document, POLDER-2/Ocean Color/Atmospheric Corrections." https://smc.cnes.fr/fr/POLDER/Algo/JMN_ATBD_polder_classe2_V1.1.pdf
- Nwankpa, C. E., W. Ijomah, A. Gachagan, and S. Marshall. 2018. "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning," 1–20. <http://arxiv.org/abs/1811.03378>
- Park, D. H., C. M. Ho, and Y. Chang. 2018. "Achieving Strong Regularization for Deep Neural Networks." https://openreview.net/pdf?id=Bys_NzbC-
- Platnick, S., M. D. King, S. A. Ackerman, W. P. Menzel, B. A. Baum, J. C. Riedi, and R. A. Frey. 2003. "The MODIS Cloud Products: Algorithms and Examples from Terra." *IEEE Transactions on Geoscience and Remote Sensing* 41 (2): 459–473. doi:10.1109/TGRS.2002.808301.
- Ricciardelli, E., F. Romano, and V. Cuomo. 2008. "Physical and Statistical Approaches for Cloud Identification Using Meteosat Second Generation-Spinning Enhanced Visible and Infrared Imager Data." *Remote Sensing of Environment* 112 (6): 2741–2760. doi:10.1016/j.rse.2008.01.015.
- Richter, R., J. Louis, and U. Müller-Wilm. 2012. "Sentinel-2 Msi-level 2a Products Algorithm Theoretical Basis Document." *European Space Agency,(Special Publication) ESA SP* 49: 1–72.
- Rosenblatt, F. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review* 65 (6): 386–408. doi:10.1037/h0042519.
- Roslan, N., M. N. M. Reba, M. Askari, and M. K. A. Halim. 2014. "Linear and Non-Linear Enhancement for Sun Glint Reduction in Advanced Very High Resolution Radiometer (AVHRR) Image." *IOP Conference Series: Earth and Environmental Science* 18 (1). doi: 10.1088/1755-1315/18/1/012041.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Representations by Back Propagating Errors." *Nature* 323: 533–536. doi:10.1038/323533a0.
- Schröder, M., R. Bennartz, L. Schüller, R. Preusker, P. Albert, and J. Fischer. 2002. "Generating Cloudmasks in Spatial High-Resolution Observations of Clouds Using Texture and Radiance Information." *International Journal of Remote Sensing* 23: 4247–4261. doi:10.1080/01431160110114547.
- Shin, D., J. K. Pollard, and J.-P. Muller. 1996. "Cloud Detection from Thermal Infrared Images Using a Segmentation Technique." *International Journal of Remote Sensing* 17 (14): 2845–2856. doi:10.1080/01431169608949110.
- Sholar, J. M. 2017. "Lightweight Deconvolutional Neural Networks for Efficient Cloud Identification in Satellite Images." <http://cs231n.stanford.edu/reports/2017/pdfs/561.pdf>

- Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. 2014. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *The Journal of Machine Learning Research* 15 (1): 1929–1958. <http://jmlr.org/papers/v15/srivastava14a.html>
- STEP. 2016a. "Sentinel-2 Cloud Mask with Fmask." <https://forum.step.esa.int/t/sentinel-2-cloud-mask-with-fmask/4152>
- STEP. 2016b. "Reducing Sen2Cor Processing Time" <https://forum.step.esa.int/t/reducing-sen2cor-processing-time/1675>
- Strandgren, J., L. Bugliaro, F. Sehnke, and L. Schröder. 2017. "Cirrus Cloud Retrieval with MSG/SEVIRI Using Artificial Neural Networks." *Atmospheric Measurement Techniques* 10 (9): 3547–3573. doi:10.5194/amt-10-3547-2017.
- Taravat, A., S. Peronaci, M. Sist, F. Del Frate, and N. Oppelt. 2015. "The Combination of Band Ratioing Techniques and Neural Networks Algorithms for MSG SEVIRI and Landsat ETM+ Cloud Masking." 2015 *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2315–2318. IEEE. doi:10.1109/IGARSS.2015.7326271.
- Tian, B., M. A. Shaikh, M. R. Azimi-Sadjadi, T. H. Vonder Haar, and D. L. Reinke. 1999. "A Study of Cloud Classification with Neural Networks Using Spectral and Textural Features." *IEEE Transactions on Neural Networks* 10 (1): 138–151. doi:10.1109/72.737500.
- Wang, M., and W. Shi. 2006. "Cloud Masking for Ocean Color Data Processing in the Coastal Regions." *IEEE Transactions on Geoscience and Remote Sensing* 44 (11): 3196–3105. doi:10.1109/TGRS.2006.876293.
- Weng, L. G., W. B. Kong, and M. Xia. 2016. "Computing Cloud Cover Fraction in Satellite Images Using Deep Extreme Learning Machine." *International Journal of Simulation: Systems, Science and Technology* 17 (48): 37.1–37.9. doi:10.5013/IJSSST.a.17.48.37.
- Xie, F., M. Shi, Z. Shi, J. Yin, and D. Zhao. 2017. "Multilevel Cloud Detection in Remote Sensing Images Based on Deep Learning." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (8): 3631–3640. doi:10.1109/JSTARS.2017.2686488.
- Yhann, S. R., and J. J. Simpson. 1995. "Application of Neural Networks to AVHRR Cloud Segmentation." *IEEE Transactions on Geoscience and Remote Sensing* 33 (3): 590–604. doi:10.1109/36.387575.
- Yuan, Y., and X. Hu. 2015. "Bag-of-Words and Object-Based Classification for Cloud Extraction from Satellite Imagery." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (8): 4197–4205. doi:10.1109/JSTARS.2015.2431676.
- Zhai, H., H. Zhang, L. Zhang, and P. Li. 2018. "Cloud/Shadow Detection Based on Spectral Indices for Multi/Hyperspectral Optical Remote Sensing Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 144 (August). Elsevier: 235–253. doi:10.1016/j.isprsjprs.2018.07.006.
- Zhang, H., and M. Wang. 2010. "Evaluation of Sun Glint Models Using MODIS Measurements." *Journal of Quantitative Spectroscopy and Radiative Transfer* 111: 492–506. doi:10.1016/j.jqsrt.2009.10.001.
- Zhang, Z., A. Iwasaki, G. Xu, and J. Song. 2018. "Small Satellite Cloud Detection Based on Deep Learning and Image Compression." *Preprints* (February): 1–12. doi:10.20944/preprints201802.0103.v1.
- Zhu, Z., and C. E. Woodcock. 2012. "Object-Based Cloud and Cloud Shadow Detection in Landsat Imagery." *Remote Sensing of Environment* 118 (1–4). Elsevier Inc: 83–94. doi:10.1016/j.rse.2011.10.028.
- Zhu, Z., S. Wang, and C. E. Woodcock. 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 (March). Elsevier Inc.: 269–277. doi:10.1016/j.rse.2014.12.014.