Atmos. Meas. Tech. Discuss., 6, 1649–1681, 2013
www.atmos-meas-tech-discuss.net/6/1649/2013/
doi:10.5194/amtd-6-1649-2013

Atmospheric Measurement Techniques Discussions

Development of a neural network model for cloud fraction detection using NASA-Aura OMI VIS radiance measurements

G. Saponaro[1], P. Kolmonen[1], J. Karhunen[2], J. Tamminen[1], and G. de Leeuw[1,3]

[1]Finnish Meteorological Institute, Climate Change Unit, P.O. Box 503, 00101 Helsinki, Finland
[2]Aalto University School of Science, Department of Information and Computer Science, P.O. Box 15400, 00076 Aalto Espoo, Finland
[3]University of Helsinki, Department of Physics, P.O. Box 64, 00014 Finland

Received: 20 October 2012 – Accepted: 14 January 2013 – Published: 13 February 2013

Correspondence to: G. Saponaro (giulia.saponaro@fmi.fi)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The discrimination of cloudy pixels is required in almost any estimate of a parameter retrieved from a satellite image in the ultraviolet (UV), visual (VIS) or infra-red (IR) parts of the electromagnetic spectrum. Also, the distincion of clouds within satellite imagery and the distribution of their micro-physical properties is essential to the understanding of radiative transfer through the atmosphere.

This paper reports the development of neural network algorithms for cloud detection for the NASA-Aura Ozone Monitoring Instrument (OMI). We present and discuss the results obtained by training mathematical neural networks with simultaneous application to OMI and Aqua-MODerate Resolution Imaging Spectrometer (MODIS) data. The neural network delivers cloud fraction estimates in a fast and automated way. The developed neural network approach performs generally well in the training. Highly reflective surfaces, such as ice, snow, sun glint and desert, or atmospheric dust mislead the neural network to a wrong predicted cloud fraction.

1 Introduction

The retrieval of atmospheric constituents from satellite data requires the accurate identification of clouds due to their high reflectance which overwhelms the contribution of other constituents to the reflectance measured at the top-of-atmosphere (TOA) reflectance (Andreae and Rosenfeld, 2008; Koelemeijer and Stammes, 1999). For instance, for the retrieval of aerosol properties all cloud-contaminated pixels are discarded. Commonly cloud detection is performed using several tests, and different algorithms have been developed to extract cloud information (Ackerman et al., 1998; Kokhanovsky et al., 2011). The most consolidated methods are based on thresholding the measured radiance or reflectance at certain wavelengths using values which are empirically estimated, or set with information from, e.g. radiative transfer models. The application of these methods requires the extraction of cloud-related information

AMTD
6, 1649–1681, 2013

from large quantities of satellite imagery using manual/visual interpretation which is a time-consuming and complicated task and, in addition, results are observer dependent. Hence, highly efficient and reliable algorithms for the automatic processing of satellite imagery are required to discriminate between cloudy and clear pixels for climatological, physical or chemical applications.

In this paper we focus on OMI cloud screening which is challenging because of the coarse spatial resolution and the lack of thermal channels in the instrument. The current method for the retrieval of OMI cloud mask is based on two individual tests: the first test combines a radiance threshold and the UV aerosol index, while the second one takes into account the spatial homogeneity of the so-called small-pixel data (van der Oord, 2002). Pixels failing either of the two tests are classified as cloudy (Acarreta and de Haan, 2002).

In this work we propose and discuss a novel approach using neural networks (NN) for the direct determination of the pixel cloud fraction from VIS measurements provided by the Ozone Monitoring Instrument (OMI). The approach employs the OMI radiance measurements together with auxiliary cloud information from the Aqua-MODerate Resolution Imaging Spectrometer (MODIS) to determine the presence of clouds mask. OMI and MODIS fly in the A-train constellation on different platforms, respectively AURA and AQUA, with a time lag of about 7 min.

In recent years neural networks have been adopted for a wide range of applications from atmospheric sciences to electromagnetic modelling. The developed applications include, e.g. forward and inverse radiative transfer problems (Krasnopolsky, 2008), the prediction of atmospheric parameters (Grivas and Chaloulakou, 2006; Sellitto et al., 2012), the inversion of satellite observations and post processing of satellite data (Mas and Flores, 2008; Del Frate and Schiavon, 1998), ozone retrievals (Del Frate et al., 2002, 2005a; Iapaolo et al., 2007), cloud classification (Tian et al., 1999), land cover classification (Aitkenhead and Aalders, 2008), and feature extraction (Del Frate et al., 2005b).

AMTD
6, 1649–1681, 2013

Here we propose the application of a NN as an alternative approach to the cloud screening task for OMI. The objective is to give information about cloud fraction, or at least discriminate cloudy from clear pixels using only spectral information. A similar approach is described in Preusker et al. (2008), where the cloud screening problem is solved by applying a neural network scheme (for the Medium Resolution Imaging Spectrometer (MERIS)) trained with a database of simulated cloudy and cloud free spectra. Here we use real data obtained from MODIS, with a spatial resolution which is much smaller than that of OMI. Hence the OMI cloud fraction in an OMI pixel can be determined using MODIS data and used to train the NN for cloud fraction determination in real OMI data.

In the next sections, we describe the adopted design for the cloud screening algorithm. This starts with some preprocessing of the OMI measured VIS radiance. Then, two different learning algorithms are employed for training the neural networks, namely the back-propagation and Extreme Learning Machine ones. We next report the comparison of the two methods and their results with an analysis of the performances of the learning algorithms for separated land and ocean pixels. Results are compared to the MODIS cloud fraction data. Finally, we show an example of application of the algorithms, which at the same time represents a test of their performance using four random orbits. The use of only 4 orbits is not sufficient to provide a complete and accurately trained NN which provides good solution for all relevant situations. However, the purpose of the current paper is to present and evaluate the method and its potential for cloud screening.

The neural networks approach described here has been developed for the OMI/MODIS combination and it will work also for the TROPOMI/VIIRS one without extensive modifications. The TROPOMI radiance spectrum has the added benefit of the oxygen A-band when compared to the OMI spectrum. This is expected to greatly enhance the performance of the neural network method as this band contains cloud information.

AMTD
6, 1649–1681, 2013

2 Instruments

OMI is a nadir-viewing near-UV-Visible spectrometer on board NASA's Earth Observing System's (EOS) Aura satellite. The Ozone Monitoring Instrument measures radiances at 751 wavelengths in the visible parts of the electromagnetic spectrum covering the wavelength range from 349 nm to 504 nm. The nominal ground footprint is $13 \times 24\ km^2$ at nadir. Complete global coverage was achieved daily (Levelt et al., 2006) between 2002 and 2008, while after 2008 the global coverage is achieved in two days due to the Row Anomaly affecting the quality of OMI level 1b radiance data. Aura flies in formation about 7 min behind Aqua in the A-train satellite constellation, orbiting the earth in a polar sun-synchronous pattern. One instrument on board Aqua is MODIS which produces many cloud related products (e.g.: cloud fraction, cloud top pressure, cloud optical thickness) (Hubanks, 2012; King et al., 1992, 1998). Since Aura and Aqua are separated by only 7 minutes the MODIS products can be used together with OMI products with quite high confidence as has been shown in Stammes et al. (2008), Vasilkov (2008) and in Sneep et al. (2008).

3 Neural Networks

Neural networks algorithms aim at identifying the relationship between input and output parameters by learning either from real or simulated reference data, rather than directly from the application of a representative physical model.

Owing to the fact that cloud properties are highly variable and sometimes difficult to detect, neural networks with their adaptive learning nature offer an attractive and computationally efficient alternative for cloud screening. It has been proven that neural networks algorithms are able to approximate any continuous multivariate non-linear function, provided that the learning data set is statistically representative of the process to be modelled and an appropriate structure for the network has been selected (Hornik et al., 1989).

AMTD
6, 1649–1681, 2013

One important class of neural networks is considered here which is referred to as multilayer perceptron (MLP). Figure 1 shows the architectural graph of a multilayer perceptron with one input layer, one hidden layer and an output layer. The input signal comes in at the input layer of the network, flows through the network on a layer-by-layer basis, and emerges at the output nodes of the network as an output signal, the response of the network to the inputs. The input nodes represent the satellite information that is fed into the network.

The role of each node in the hidden layer is determined by the activities of the input nodes and the weights in the connection between input and hidden nodes. The behaviour of the output node depends on the activity of the hidden nodes and the weights between the hidden nodes and output ones. The weights between the input and the hidden nodes determine when each hidden node is active. In the MLP, the model of each node is based on a differentiable non-linear activation function. These functions enable the network to learn complex tasks by extracting features from the input signal as explained below. More details on artificial neural networks theory can be found in Ham and Kostanic (2001) and Haykin (2009). A multilayer perceptron network must be trained and several algorithms have been developed for this task, including the basic back-propagation, algorithm, batch version of it, and the Levenberg-Marquardt algorithm.

The two learning algorithms, back-propagation and extreme learning machine, that were applied in this work to train the MLP neural networks are described in the next sections.

3.1 Back-propagation

The error back-propagation algorithm is a popular learning algorithm used to train neural networks by modifying the weights during the training phase in order to model a particular learning task correctly for the training examples (Haykin, 2009). The training phase of a NN updates the weights iteratively using the negative gradient of a Mean Square Error (MSE) function. Basically, the error back-propagation algorithms perform

AMTD
6, 1649–1681, 2013

two passes through each layer of the network: the first pass starts with the application of the input vector to the input nodes of the network, and its effect is forwarded trough the layers. This is the forward pass during which all the weights of the network are fixed. Then, a set of output is produced as the response of the network to the input signal, and is subtracted from a desired (target) response to produce an error signal. The error signal is propagated backward through the network. This process represents the backward pass. During the backward pass the weights are adjusted to make the actual response of the network move closer to the desired one in a statistical sense. The model of each node is based on a non-linear activation function.

The particular activation function adopted in this work is the sigmoidal function

$$g(x) = \tanh(ax), \tag{1}$$

where $a > 0$ is a scaling parameter.

The network structure consist of an input layer in which each component of the training vectors is inputted into the network in its own node, one hidden layer of 25 nonlinear nodes, and a linear output layer. All the computations take place in the hidden layer and in the output layer where there is only one node, which is commonplace in classification applications. The back-propagation algorithm is a stochastic gradient algorithm for minimizing the mean-square error (MSE) between the target values and the respective true outputs of the MLP network. It converges slowly and often to some local minimum of the MSE error only instead of achieving global minimum. This should be taken into account when the solution is analyzed.

3.2 Extreme learning machine

The Extreme learning machine uses a special MLP network structure with one hidden layer where the weights between the input layer and the nodes of the hidden layer are chosen randomly beforehand, and similarly for the bias terms of the hidden layer nodes. The output layer is to be taken linear. The extreme learning machine method for neural networks consists of the following steps:

AMTD
6, 1649–1681, 2013

– Assume that we have a training set $\boldsymbol{x}_i, \boldsymbol{t}_i$, $i = 1, 2, ..., N$, where $\boldsymbol{x}_i$ is the i-th input vector of dimension n, and $\boldsymbol{t}_i$ is the corresponding target vector of dimension m.

– Choose the activation function $g(t)$ and number M of nodes in the hidden layer. In our case, the selected activation function is $g(x) = \tanh(wx + \beta)$.

– Assign randomly the hidden layer scaling parameter vectors $\boldsymbol{w}_j$ and biases $\boldsymbol{\beta}_j$, $j = 1, 2, ..., M$.

– Calculate the $N \times M$ hidden layer output matrix **H**. Its elements are

$$h_{ij} = g(w_j^{\mathbf{T}} x_i + \beta_j). \tag{2}$$

– Calculate the $M \times m$ weight matrix **B** of the output layer from:

$$\mathbf{B} = \mathbf{H}^{+}\mathbf{T}, \tag{3}$$

where $\mathbf{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, ..., \boldsymbol{t}_N]^{\mathbf{T}}$ is the $N \times m$ target matrix and $\mathbf{H}^{+}$ is the pseudo-inverse of the matrix **H**. The matrix **B** is

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{w}_{1,m} \\ \boldsymbol{w}_{2,m} \\ . \\ . \\ . \\ \boldsymbol{w}_{M,m} \end{bmatrix}. \tag{4}$$

This learning method provides an easy implementation, it reaches small training error, and it runs extremely fast as compared e.g. to the standard back-propagation algorithms (more detailed information in Huang et al., 2006). Since this type of algorithm does not require tuning and the hidden layer parameters can be fixed, the optimal solution can be resolved from a system of linear equations using the least-squares method

AMTD
6, 1649–1681, 2013

(pseudo-inverse) and avoiding problems related to gradient learning methods, such as local minima in the back-propagation. The drawback of ELM is an increased number of neurons in the hidden layer because the scaling parameters and biases are not learned from data but chosen randomly.

4 Application OMI VIS radiance measurements for cloud detection by means of neural networks

To investigate the potential and limits of the application of NN for cloud screening, a representative data set for the observed phenomena is required, including a statistically significant set of observational uncertainties. The training dataset needs to be as complete as possible and of sufficient quality. In the NN training phase of the described cloud detection method, the input consists of quantities derived from OMI VIS measurements and MODIS cloud fraction product which are described in Sect. 4.1.

4.1 The OMI training input

This section describes the OMI products included in the training dataset.

OMI measures radiances at a large number of wavelengths in the VIS band. Radiances are converted to reflectance to scale the input information between zero and unity for the neural networks. The conversion was done using equation

$$\rho(\lambda) = \frac{I(\lambda)}{\pi \cos(\theta_z) L(\lambda)}, \quad (5)$$

where ρ is the calculated reflectance, I is the OMI measured radiance at wavelengths between 349 nm and 504 nm, θ_z is the solar zenith angle, and L is solar irradiance at wavelength λ.

OMI provides, at one wavelength, a five times higher spatial sampling in the flight direction than normal which is called small-pixel data. This capability provides more

AMTD
6, 1649–1681, 2013

information about spatial inhomogeneity in a pixel caused by, e.g. clouds. The small pixel data is included in the level-1B data set (van der Oord, 2002). The small-pixel reflectance has been derived from the small pixel radiance using Eq. (5), and the values were used to calculate the variance of the reflectance in each OMI pixel which was added to the training data set.

Moreover, the solar zenith angle (SZA), providing information about measurement geometry, and OMI Surface Reflectance Climatology Data Product (OMLER) were also included to the training input data. OMLER is an OMI product describing the monthly climatology of the Earth's surface Lambertian Equivalent Reflectance (LER) value. LER is defined as the required reflectance of an isotropic surface necessary to match the observed top of the atmosphere (TOA) reflectance in a pure Rayleigh scattering atmosphere under cloud and aerosol free conditions. The product has a spatial resolution of 0.5 by 0.5 degrees and is build by using five years of OMI data, obtained between January 2005 and December 2009.

4.2 Singular value decomposition for the OMI reflectance

In our study, the OMI reflectance data consists of 751 measurements for each pixel. Therefore, a dimensionality reduction is necessary to help lessen the effect of large datasets on computing time. Dimensionality reduction is the transformation of high-dimensional data into a representation of reduced dimensionality without loosing valuable information. The singular value decomposition (SVD) is a method that converts a matrix to its diagonal form, and it is a very powerful tool in data reduction, power spectrum estimation as well in image restoration (Golub and Van Loan, 1996). In the present study the SVD procedure is applied to reduce the number of reflectances at 751 wavelengths to a set of 20 values. This was done as follows.

Consider an $N \times M$ matrix $\mathbf{X}$ where $N \geq M$. It is possible to represent this matrix in the r-dimensional subspace where $r \leq M$. Let $\mathbf{U} = \mathbf{X}\mathbf{X}^{\mathbf{T}}$ and $\mathbf{V} = \mathbf{X}^{\mathbf{T}}\mathbf{X}$ be non negative, symmetric matrices with the same eigenvalues $\lambda_1, \lambda_2, ..., \lambda_r$, where it is assumed that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ...\lambda_r$. The square roots of these eigenvalues are called the singular

AMTD
6, 1649–1681, 2013

values of **X**. Now, if we form matrices **Ψ** and **Φ** from the corresponding eigenvectors of **U** and **V**, then **X** can be diagonalized as:

$$\mathbf{X} = \mathbf{\Psi \Lambda \Phi}^{\mathrm{T}} \tag{6}$$

where **Λ** is the diagonal matrix of the eigenvalues, i.e. diag(**Λ**) = $[\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_r}]$. Basically, each singular value represents the information content of the matrix **X** projected into each subspace.

The reduction of the reflectance data set is done by using only the part of the diagonalized system where the eigenvalues are significant. We tested several numbers of non-zero eigenvalues such as 5, 10 and 20. They all provide a large reduction when compared to the original size of 751. Finally, 20 eigenvalues were chosen as adding more values did not enhance the results in the initial testing.

4.3 The MODIS cloud fraction training data

The MODIS cloud fraction is employed as auxiliary reference data for training the neural networks. To this end, OMI and MODIS pixels need to be matched (Stammes et al. (2008)). Since the OMI Level 1B data product provides geodetic latitude and longitude only for the center of each ground pixel, a separate product, OMI ground pixel corner (OMPIXCOR), was used to arrange cross-platform pixel mapping. The OMI latitude-longitude corner coordinates are used to construct boxes representing the pixel area. The MODIS geo-located data is then searched for measurements falling within each box: a MODIS pixel is considered to fall within a particular OMI box if the center lies inside the OMI pixel boundary. The matches allow the use of MODIS data to determine the cloud fraction within an OMI pixel, as the number of cloudy pixels divided by the total number of MODIS pixels falling within the considered OMI pixel. The re-gridded data is then included in the training input as the reference data.

AMTD
6, 1649–1681, 2013

4.4 Dataset composition for training and NN structure

For the purpose of training the NN algorithm, the training dataset was composed of the OMI SVD reduced reflectance values, the OMLER climatological data, the solar zenith angle, the small pixel variance. The corresponding MODIS geometrical cloud fraction is the reference data. The block diagram in Fig. 2 shows the different components of the training data set which represents the input to the neural network. The neural network processes this information and provides the predicted cloud fraction.

The neural network considered for this study was a multilayer perceptron (MLP) with a single hidden layer, as is illustrated in Fig. 1. The input layer consists of 24 nodes: 20 input nodes for the SVD reduced OMI reflectance data, one input node for the solar zenith angle, the small pixel variance, OMLER data, and one output node for the reference MODIS cloud fraction.

Since we compared the performances of two learning algorithms with different characteristics, back-propagation and Extreme Learning Machine, the structure of the NN changes in the two considered cases. The two algorithms differ in the number of nodes in the hidden layer. To determine the optimal number of nodes in the hidden layer, several tests were made. For the back-propagation algorithm, 25 hidden nodes guarantee a good performance either in terms of training accuracy or training time. The ELM algorithm, because of its nature, requires a much larger amount of hidden nodes. The optimal number was found to be 240.

5 Results

The back-propagation and Extreme Learning Machine learning algorithms were trained with the same set of input data. Four random OMI orbits (2004m1028t0755, 2005m0828 t1257, 2006m0912t0828, 2006m0113t1325), covering different areas and seasons, were selected to compose the training datasets. The performance of the learning algorithms in predicting cloud fraction was assessed in terms of the percentage of pixels

AMTD
6, 1649–1681, 2013

that the NNs estimate to be cloudy/clear and which are actually cloudy/clear according to the MODIS re-gridded cloud fraction.

To compare the cloud masking capabilities of two NNs, two cloud thresholds were considered, 60 % and 30 %, respectively. A threshold $t = 60\,\%$ implies that 60 % of an OMI pixel contains clouds. Values above the thresholds represent cloudy pixels while values below the thresholds represent cloud-free ones. The selected orbits were divided into a cloudy and a non-cloudy set based on the threshold values and the analysis was based on these separated sets. We analysed the results by selecting MODIS cloud fraction pixel with values above (below) the 60 % or 30 % thresholds and compare these reference data with the corresponding values predicted by the NNs.

Several options were explored in order to prepare the training dataset. Data was first analysed by observing each orbit individually, then by separating land- and water-covered pixels and discarding the ice-covered ones. Although the performances of the two training algorithms in Fig. 3 are comparable when each orbit is individually observed, the back-propagation algorithm sets apart because of the long processing time required for training the algorithm. The histograms shown in Fig. 3 represent the percentages of correctly detected cloudy and cloud-free pixels for the two thresholds for the four selected orbits presented along the x-axis. Both algorithms, back-propagation and ELM, detect cloud fraction accurately when the cloudiness threshold is set to 60 % rather than 30 %. The lowest performance is obtained for orbit 2005m0828t1257 and orbit 2006m0912t0828 for cloud-free pixels. Generally cloud-free pixels are detected with lower accuracy. This can be explained by considering the chosen orbits: most of the pixels are fully covered by clouds, thus not enough information is provided to the NN for the training.

Next, in order to test whether land or ocean pixels influence the NN behaviour, we analysed the same data but trained the learning algorithms with separated date-sets for land/water pixels. Results are presented in Fig. 4. The histograms show that the back-propagation algorithm performs with lower efficiency. The four plots show that the lowest number of correctly detected pixels corresponds to clear pixels. This

AMTD
6, 1649–1681, 2013

confirms the lack of information regarding cloud-free reflectance in each single orbit. Referring to results from the ELM learning algorithm for orbits 2005m0828t1257 and 2006m0912t0828, we unexpectedly found that NN perform worse over ocean than over land. The ocean, in fact, acts as an homogeneous dark surface in the background, hence contrast should be enhanced between cloudy and cloud-free measurements.

To supply the neural networks with the largest amount of information and scenarios, all four orbits were included in one dataset for the training phase. Providing a larger amount of data for the training, with a larger variety as regards seasons and locations, is expected to enhance the overall performance. The results are presented in Fig. 5. By comparing Figs. 4 and 5, we can observe that the performance of the NN are enhanced when detecting clear pixels for the threshold of 60% for land and ocean pixels. Results for the 30 % threshold do not present major improvement even when the training dataset is enlarged. As mentioned previously, the observed orbits are fully clouded meaning that the 30 % threshold is not representativefor the selected data. After training the ELM algorithm with a training dataset formed by the four observed orbits taking into consideration the ground pixel coverage type (land/water), we ran the resulting NNs on each single considered orbit. Pixels covered with ice or snow were discarded. This process leads to predicted cloud fraction values for each orbit.

The cloud fractions estimated from the ELM neural network are plotted with MODIS re-gridded data for each of the considered orbits in Fig. 6. The colour scale provides a measure for the number of points. Correlation coefficients are shown in the upper right corner of each figure. The images are organized in a matrix where the four rows represent the four analysed orbits and the two columns divide land and water pixels, respectively. A good correlation is observed for the the orbits in the first and last rows. For small cloud fraction over land the value retrieved by the NN and the MODIS geometrical cloud fraction appear non-linearly related as we find in Fig. 6 a bias of 10 % and 12 % between the datasets, with correlations of 0.81 and 0.89 for orbit 2004m1028t0755 and orbit 2006m0113t1325, respectively. This implies that land surface albedo leads

AMTD
6, 1649–1681, 2013

the NN to underestimate the cloud fraction retrieved by MODIS. Low correlations are found for orbits 2005m0828t1257 and 2006m0912t0828, over both land and ocean.

Figure 7 shows the results of testing of the NN model for cloud fraction prediction. Once the ELM neural network has been trained with the dataset including all orbits, the final weights are applied to each single orbit and the accuracy of the cloud fraction estimates is determined. The algorithm fails in distinguishing cloud-free pixels, especially for the lower threshold for both land and ocean measurements and the reason, as explained above, is due to the lack of cloud-free pixels in the training dataset. Orbits 2005m0828t1257 and 2006m0912t0828 present low accuracy values for predicted cloud fractions over land and there is an overestimation of cloudy pixels.

Overall, a poor correlation between predicted and MODIS cloud fraction is always present in orbits 2005m0828t1257 and 2006m0912t0828 which are presented in Figs. 8 and 9. Figures 8a and 9a show the MODIS geometrical cloud fraction and the color scale indicates the cloud fraction between 0 (cloud-free) and 100 (totally clouded). Figures 8b and 9b are the MODIS RGB granule images. Figures 8c and 9c show the difference between MODIS geometrical cloud fraction and the ELM estimates: the value 0 of the color scale represents a perfect agreement between the cloud fraction of the two datasets, while a value of 100 signifies a total mismatch. Some of the causes of failure are indicated by the MODIS granules inserted in Figs. 8b and 9b. By inspecting MODIS RGB granule images, the ocean areas resulting in non-satisfactory cloud retrieval accuracy are affected by dust and sun glint while over land highly reflective surfaces such as desert and ice (low contrast against clouds) result in low accuracy. Good performances of the neural network in cloud detection are presented in Figs. 10 and 11.

A preliminary validation of the ELM algorithm for cloud masking was made using observations which were not included in the training dataset. The dataset of four orbits was divided into two subsets. To this end, the training set was now composed of three orbits (2004m1028t0755, 2005m0828t1257, and 2006m0912t0828), while orbit 2006m0113t1325 was left for independent testing. The weights determined during the

AMTD
6, 1649–1681, 2013

training of the ELM were used for the independent orbit. The result of the independent validation is presented in Fig. 12. The resulting predicted cloud fraction is rather inaccurate and this might be due to two factors: first the very limited training dataset, and second two orbits of the training set presented problems in discriminating bright surfaces from cloud reflectances, as previously shown in Figs. 8 and 9. The three orbits of the training set were mainly covered by clouds, thus they provided insufficient data describing cloud-free pixels and moreover cloud-free areas were influenced by dust, sun-glint or bright surface. Hence, the trained NN was unable to produce totally cloud free estimates for the cloud fraction. However, the general features of the cloud fraction can be observed in Fig. 12.

6 Conclusions

In this work, a neural networks-based solution was described for the problem of cloud screening for OMI VIS satellite imagery. Neural networks are attractive for cloud screening because of their capability of high computational speed for large datasets. Moreover, they rely on auxiliary data only during the training and they are independent from the instrument platform which makes the approach portable to other combinations of instruments such as TROPOMI/VIIRS.

A comprehensive study was made to select the optimal learning algorithm for the cloud screening task, and two neural network learning algorithms, back-propagation and ELM, were compared. An SVD procedure was included in order to reduce the 751 VIS channels from OMI to 20 without loosing information of the original data.

The results show that the ELM-based solution achieved higher cloud screening accuracy than the back-propagation algorithm. The back-propagation based scheme was also extremely time consuming in the training phase when compared to the ELM training approach. Both algorithms performed poorly over cloud-free scenes because the training dataset lacks information representing clear sky. All the orbits were mainly

AMTD
6, 1649–1681, 2013

totally cloudy, and moreover the few clear-pixels were contaminated with dust, sun glint or covered by bright surfaces.

It was found that the spectral features alone can discriminate cloudy from clear pixels with a reasonable accuracy. Nevertheless, none of these feature-sets provide consistently good discrimination ability for all cloud/background classes which will be taken into consideration in future work. NN failed in predicting cloud fraction in the presence of ice, snow, dust and sun glint because of their spectral similarities which mislead the neural network to a wrong final prediction.

In the future we will study the possibility to compose a complete database of orbits for learning the neural network where the orbits having low correlation with the reference data are iteratively discarded, and an alternative approach for the dimensionality reduction based on information content.

This work serves as a preliminary study on cloud screening for the upcoming TROPOspheric Monitoring Instrument (Veefkind et al., 2012). The TROPOMI instrument is expected to provide daily high-quality global information of atmospheric parameters for climate and air quality applications. TROPOMI is planned to be launched in 2015 on board the ESA/GMES Sentinel 5 Precursor (S5P) satellite, extending the current data records from OMI on board NASA EOS-Aura and SCIAMACHY on board ESA Envisat. Precise cloud screening is a challenge at the TROPOMI spatial resolution of $7 \times 7\,\mathrm{km}^2$ and without thermal infrared measurements available. TROPOMI will fly in constellation with the Visible Infrared Imager Radiometer Suit (VIIRS) from the US NPOESS Preparatory Project (NPP) mission. VIIRS provides cloud imagery. Preliminary investigation of temporal co-registration between the VIIRS measurements of clouds and TROPOMI atmospheric measurements have been analysed by Genkova et al. (2012). To apply the developed NN cloud screening for TROPOMI only minor modifications have to be made. The VIIRS instrument information about clouds can be used for the TROPOMI pixel. The TROPOMI radiance spectrum consists not only of the VIS band but also of the Oxygen A-band. While the VIS band can be processed as described

AMTD
6, 1649–1681, 2013

here, the A-band may need separate treatment because of its high information content about clouds.

As the NN is developed to provide cloud fraction or cloud screening independently from the location and time of the measured radiance, the training dataset has to include different climatologies with different seasonal coverage. These demands are outside the scope of the method description and initial testing described here but they will be addressed in future testing and validation.

Acknowledgements. All RGB MODIS granules are publicly available and are provided by the NASA Level1 and Atmoshere Archive and Distribution System (LAADS) website: http://ladsweb.nascom.nasa.gov/data/search.html. The authors would like to acknowledge the KNMI-OMI team. Funding for this work was provided by the PP-TROPOMI project and Tekes (the Finnish funding Agency for Technology and Innovation and Ministry of Transport and Communications).

References

Acarreta, J. R. and de Haan, J. F.: Cloud pressure algorithm based on the O_2-O_2 absorption, OMI Algorithm Theoretical Basis Document (ATBD), vol. III, Clouds, Aerosols, and Surface UV Irradiance, edited by: Stammes, P., 17–29, R. Neth. Meteorol. Inst., De Bilt, 2002. 1651

Ackerman, S. A., Strabala, K. I., Menzel, W. P., Frey, R. A., Moeller, C. C., and Gumley, L. E.: Discriminating clear sky from clouds with MODIS, Earth-Sci. Rev., 89, 13–42, 1998. 1650

Aitkenhead, M. J. and Aalders, I. H.: Classification of Landsat Thematic Mapper imagery for land covering using neural networks, Int. J. Remote Sens., 29, 2075–2084, 2008. 1651

Andreae, M. O. and Rosenfeld, D.: Aerosol-cloud-precipitation interaction. Part 1. The nature and sources of cloud-active aerosols, J. Geophys. Res., 103, 32141–32157, 2008. 1650

Bartlett, P. L.: The sample complexity of pattern classification with neural networks: the size of weights is more important than the size of the network, IEEE Trans. Neural Netw., 44, 525–536, 1998.

Del Frate, F. and Schiavon, G.: A combined natural orthogonal function/neural network technique for the radiometric estimation of atmopheric profiles, Radio Sci., 33, 405–410, 1998. 1651

AMTD
6, 1649–1681, 2013

Del Frate, F., Ortezi A., Casadio, S., and Zehner, C.: Application of neural algorithms for a real-time estimation of ozone profiles from GOME measurements, IEEE Trans. Geosci. Remote Sens., 40, 2263–2270, 2002. 1651

Del Frate, F., Iapaolo, M., and Casadio, S.: Intercomparison between GOME ozone profiles retrieved by neural network inversion schemes and ILAS product, J. Atmosp. Ocean. Technol., 22, 9, 1433–1440, 2005a. 1651

Del Frate, F., Iapaolo, M., Casadio, S., Godin-Beekman, S., and Petitdidier, M.: Neural network for the dimensionality reduction of GOME measurement vector in the estimation of ozone profiles, J. Quant. Spectrosc. Radiat. Trans., 92, 275–291, 2005b. 1651

Genkova, I., Robaidek, J., Roebling, R., Sneep, M., and Veefkind, P.: Temporal co-registration for TROPOMI cloud clearing, Atmos. Meas. Tech., 5, 595–602, doi:10.5194/amt-5-595-2012, 2012. 1665

Grivas, G. and Chaloulakou, A.: Artificial neural networks for prediction of PM10 hourly concentrations in the Greater Area of Athens, Greece, Atmos. Environ., 40, 1216–1229, 2006. 1651

Golub, G. H. and Van Loan, C. F.: Matrix computations, JHU Press, 1996. 1658

Ham, F. and Kostanic, I:, Principles of neurocomputing for science and engineering, McGraw-Hill, 2001. 1654

Haykin, S.: Neural Networks and Learning Machines, 3rd Edn., Pearson Int. Ed., 2009. 1654

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural Netw., 2, 359–366, 1989. 1653

Huang, G. B., Zhu, Q. Y., and Siew, C. K.: Extreme learning machine: theory and applications, Neurocomp., 70, 489–501, 2006. 1656

Hubanks, P. A.: MODIS Atmosphere QA Plan for Collection 005 and 051, Version 3.1, 2012. 1653

Iapaolo, M., Godin-Beekman, S., Del Frate, F., Casadio, S., Petitdidier, M., McDermid, I. S, Leblanc, T., Swart, D., Meijer, Y., Hansen, G., and Stebel, K.: Ozone profile retrieved by neural network techniques: A global validation with lidar measurements, J. Quant. Spectrosc. Radiat. Transf., 107, 105–119, 2007. 1651

Joiner, J. and Bhartia, P. K: The determination of cloud pressures from rotational Raman scattering in satellite backscatter ultraviolet measurements, J. Geophys. Res., 100, 19–26, 1995.

AMTD
6, 1649–1681, 2013

Joiner, J. and Vassilkov, A. P.: Firts results fromt the OMI rotational Raman Scattering cloud pressure algorithm, IEEE Trans. Geosci. Remote Sens., 44, 1272–1282, doi:10.1109/TGRS.2005.861385, 2006.

Karayiannis, N. and Venetsanopoulos, A. N.: Artificial neural networks: learning algorithms, performance evaluation, and applications, The Springer International Series in Engineering and Computer Science, 209, 460 p., 1993.

King, M. D., Kaufman, Y. J., Menzel, W. P., and Tanre, D.: Remote sensing of cloud, aerosol, and water vapor properties from the Moderate Resolution Imaging Spectrometer (MODIS), IEEE Trans. Geosci. Remote Sens., 30, 1–27, 1992. 1653

King, M. D., Tsay, S. C., Platnick, S. E., Wang, M., and Liou, K. N.: Cloud retrieval algorithms for MODIS: optical thickness, effective particle radius, and thermodynamic phase, Tech. Rep. ATBD-MOD-05, NASA, Washington, D.C., 1998 1653

Koelemeijer, R. B. A., Stammes, P., and de Haan, J. D.: Effects of clouds on the ozone column retrieval from GOME UV measurements, J. Geophys. Res., 104, 8281–8294, 1999. 1650

Koelemeijer, R. B. A., Stammes, P., Hovenier, J. W., and de Haan, J. D.: A fast method for retrieval of cloud parameters using oxygen A-band measurements from the Global Ozone Monitoring Instrument, J. Geophys. Res., 106, 3475–3490, 2001.

Kokhanovsky, A. A., Platnic , S., and King, M. D.: Remote sensing of terrestrial clouds from space using backscattering and thermal emission techniques, The Remote Sensing of Tropospheric Composition from Space, Physics of earth and space environments, doi:10.1007/978-3-642-14791-3, Springler-Verlag, Berlin, 2011. 1650

Krasnopolsky, V. M.: Neural network applications to solve forward and inverse problems in atmospheric and oceanic satellite remote sensing Artificial Intelligence Methods in the Environmental Sciences, Germany, Springer-Verlag, 2008. 1651

Levelt, P. F., van der Oord, G. H. J., Dobber, M. R., Malkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The Ozone Monitoring Instrument, IEEE Trans. Geosci. Remote Sens., 44, 1093–1101, 2006. 1653

Mas, J. F. and Flores, J. J.: The application of artificial neural networks to the analysis of remotely sensed data, Int, J. Remote Sens., 29, 617–663, 2008. 1651

Preusker, R., Hunerbein, A., and Fischer, J.: ATBD Cloud Detection – Revision 0, ATBD Cloud Detection, 2008. 1652

Sellitto, P., Del Frate, F., Solimini, D., and Casadio, S.: Tropospheric ozone column retrieval from ESA-Envisat SCIAMACHY Nadir UV/VIS radiance measurements by means

AMTD
6, 1649–1681, 2013

of a neural network algorithm, IEEE Trans. Geosci. Remote Sens., 50, 998–1011, doi:10.1109/TGRS.2011.2163198, 2012. 1651

Sneep, M., de Haan, J. F., Stammes, P., Wang, P., Vanbauce, C., Joiner, J., Vasilkov, A., and Levelt, P.: Three-way comparison between OMI and PARASOL cloud pressure products, J. Geophys. Res., 113, D15S23, doi:10.1029/2007JD008694, 2008. 1653

Stammes, P., Sneep, M., de Haan, J. F., Veefkind, J. P., Wang, P., and Levelt, P. F.: Effective cloud fraction from Ozone Monitoring Instrument. Theoretical framework and validation, J. Geophys. Res., 113, D16S38, doi:10.1029/2007JD008820, 2008. 1653, 1659

Tian, B., Shaikh, M. A., Azimi-Sadjadi, M. R., Vonder Haar, T. H., and Reinke, D. L.: A study of cloud classification with neural networks using spectral and textural features, IEEE Trans. Geosci. Remote Sens., 1, 138–151, 1999. 1651

van der Oord, B.: OMI small pixel data, TN-OMIE-KNMI-397, 1, 2002. 1651, 1658

Vasilkov, A., Joiner, J., Spurr, R., Bhartia, P. K., Levelt, P., and Stephens, G.: Evaluation of the OMI cloud pressure derived from rotational Raman scattering by comparison with other satellite data and radiative transfer simulations, J. Geophys. Res., 13, D15S19, doi:10.1029/2007JD008689, 2008. 1653

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt P. F.: TROPOMI on the ESA Sentinel-5 Precursor; a GMES mission for Global Observation of the Atmospheric Composition Climate, Air Quality Applications, Remote Sens. Environ., 120, 70–83, 2012. 1665

AMTD
6, 1649–1681, 2013

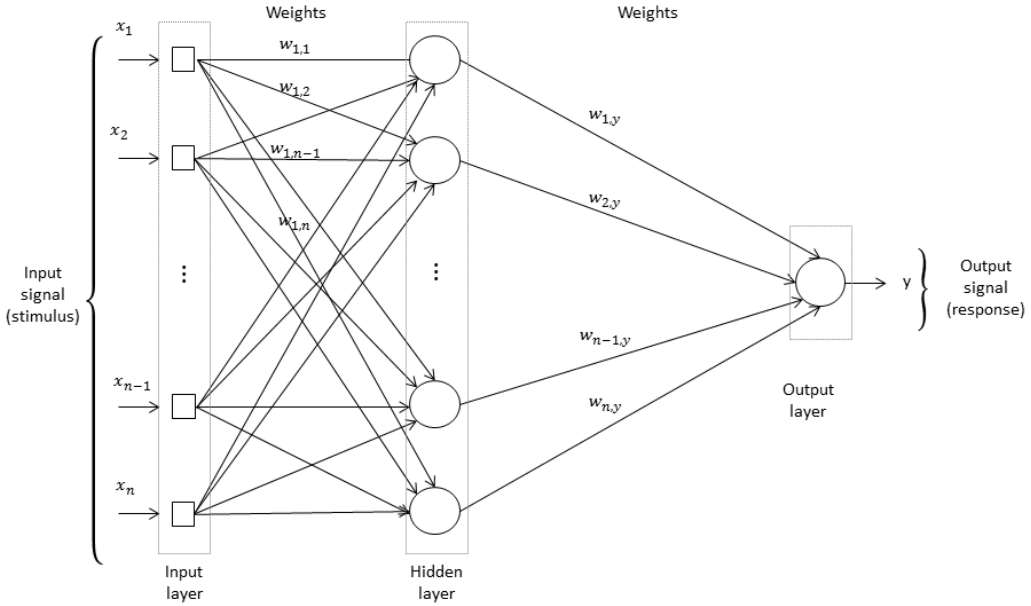

Fig. 1. Neural-network feedforward topology. x_n represents the n-th input unit, and y represents the output unit.

AMTD
6, 1649–1681, 2013

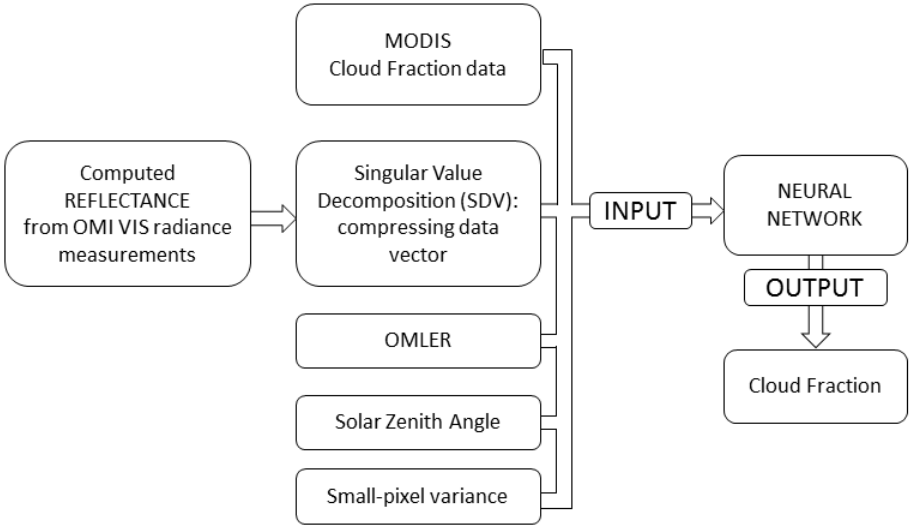

Fig. 2. Block diagram of the proposed approach for training the neural network. The training dataset is composed of the target data represented by the MODIS cloud fraction re-gridded onto the OMI orbit, the compressed OMI reflectance vector data, and additional data such as climatological data (OMLER), the solar zenith angle and the computed small-pixel variance. These data form the input vector which is fed to the neural network. The neural network response is a predicted cloud fraction for the given orbit.

AMTD
6, 1649–1681, 2013

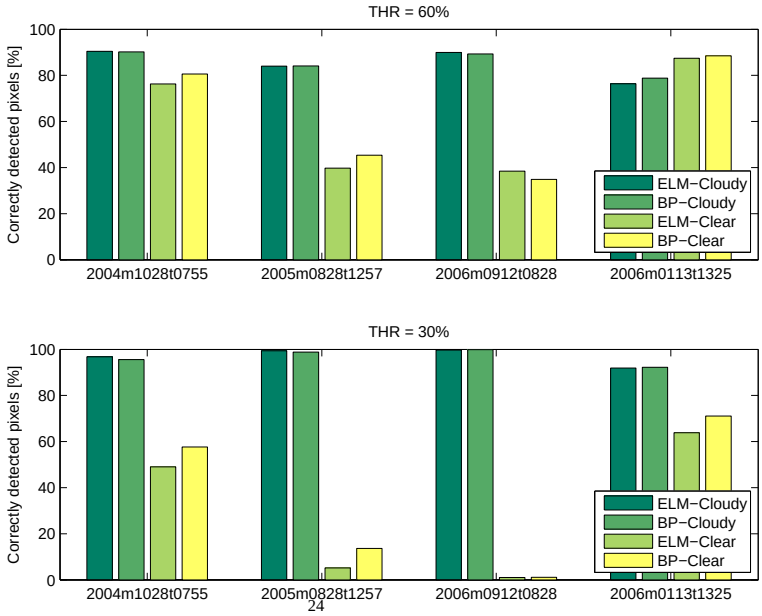

Fig. 3. Comparison of BackPropagation and Extreme Learning Machine accuracy in predicting cloud fraction for each single observed orbit.

AMTD
6, 1649–1681, 2013

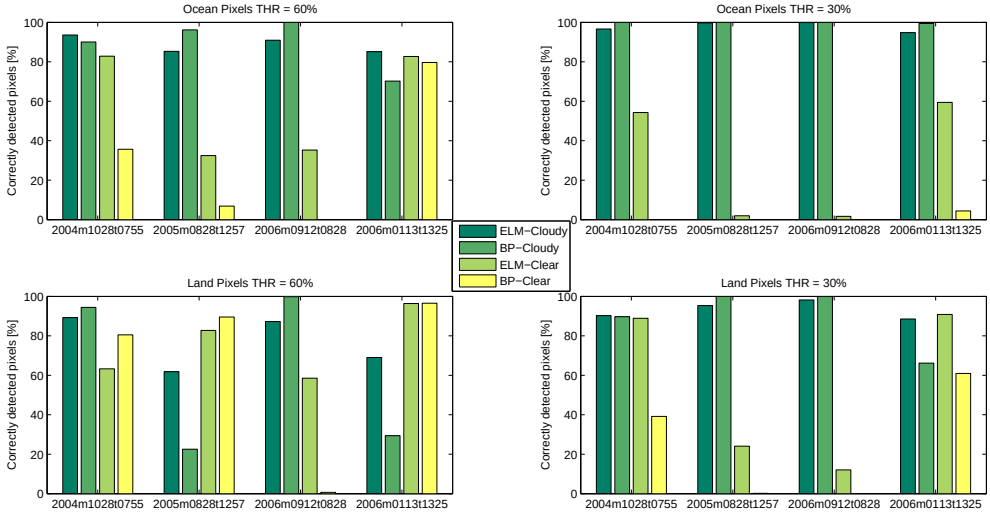

Fig. 4. Comparison of BackPropagation and Extreme Learning Machine performance in predicting cloud fraction for divided land and ocean pixel when ice/snow pixel are discarded.

AMTD
6, 1649–1681, 2013

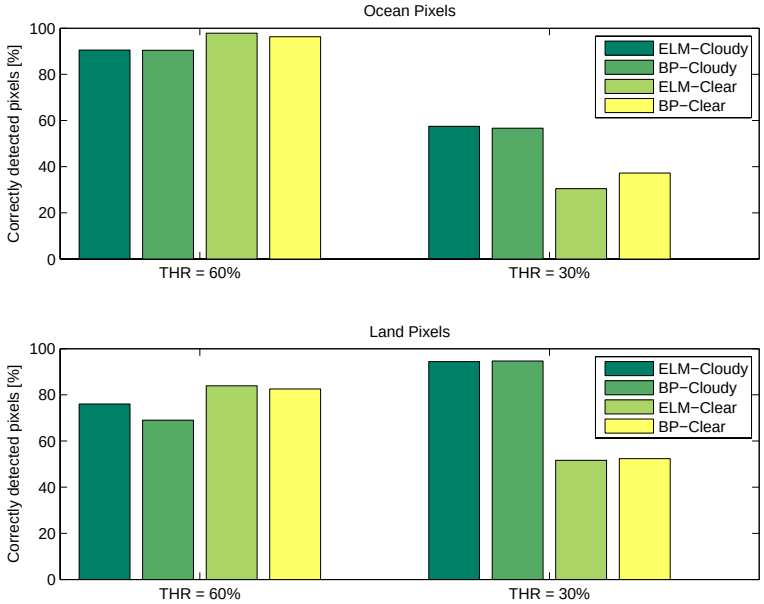

Fig. 5. Summary of results of cloud fraction estimation from ELM and BP algorithms when all four orbits are included in the training dataset.

AMTD
6, 1649–1681, 2013

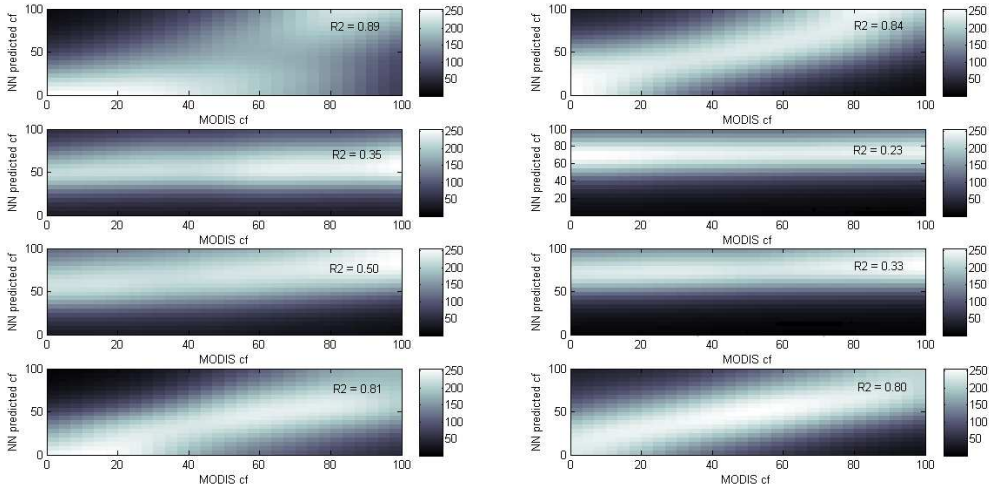

Fig. 6. Density plots between ELM NN predicted cloud fraction and reference MODIS cloud fraction. The images are organized in a matrix where the two columns divide land (left) from water (right) pixels, and the four rows represent the four analysed orbits presented in this order starting from the top: 2004m1028t0755, 2005m0828t1257, 2006m0912t0828, and 2006m0113t1325. The density plots show the correlation between the cloud fraction estimated by neural networks and the MODIS data. A good correlation is observed for the the orbits in the first and last row, respectively orbit 2004m1028t0755 and orbit 2006m0113t1325, and particularly over ocean pixels. low correlations are found for orbits 2005m0828t1257 and 2006m0912t0828, both on land and ocean.

AMTD
6, 1649–1681, 2013

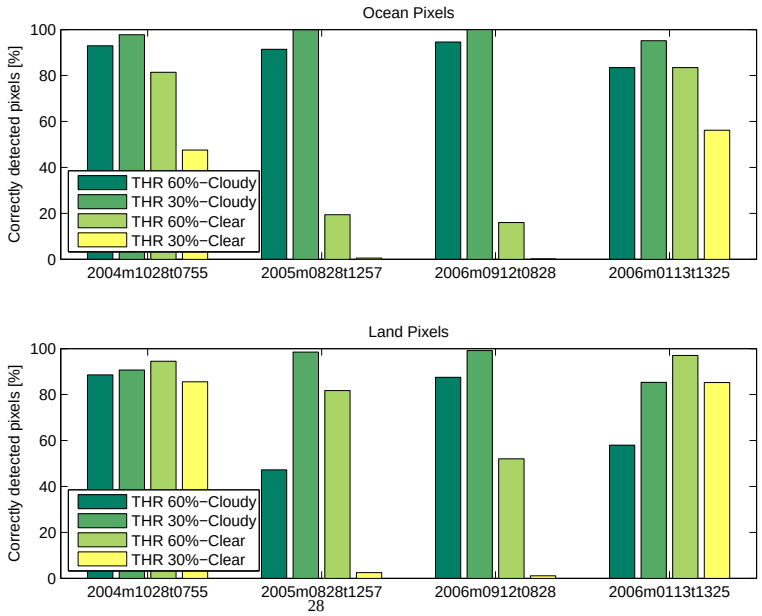

Fig. 7. Summary of accuracy of ELM NN in predicting cloud fraction once it has been trained with dataset composed of all observed orbits.

AMTD
6, 1649–1681, 2013

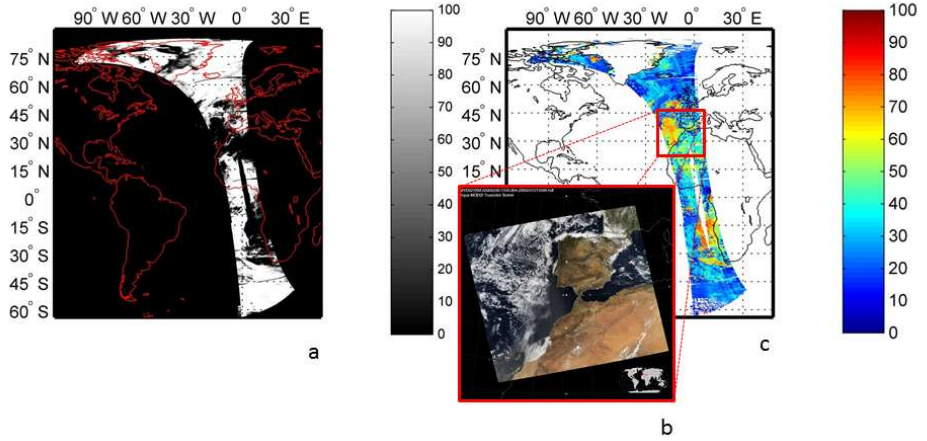

Fig. 8. Comparison of NN estimated cloud fraction and reference data for orbit 2005m0828t1257. **(a)** Computed MODIS geometrical cloud fraction re-located onto OMI grid. The color-code ranges from 0 (cloud free) to 100 (total cloud cover). **(b)**At those areas where the neural network fails in predicting cloud fraction, the corresponding MODIS RGB granule images are selected. In this case the MODIS image shows the presence of dust as the reason of failure. **(c)** Absolute difference between the MODIS geometrical cloud fraction and the NN predicted cloud fraction. The color-code ranges from 0 (perfect match) to 100 (complete mismatch).

AMTD
6, 1649–1681, 2013

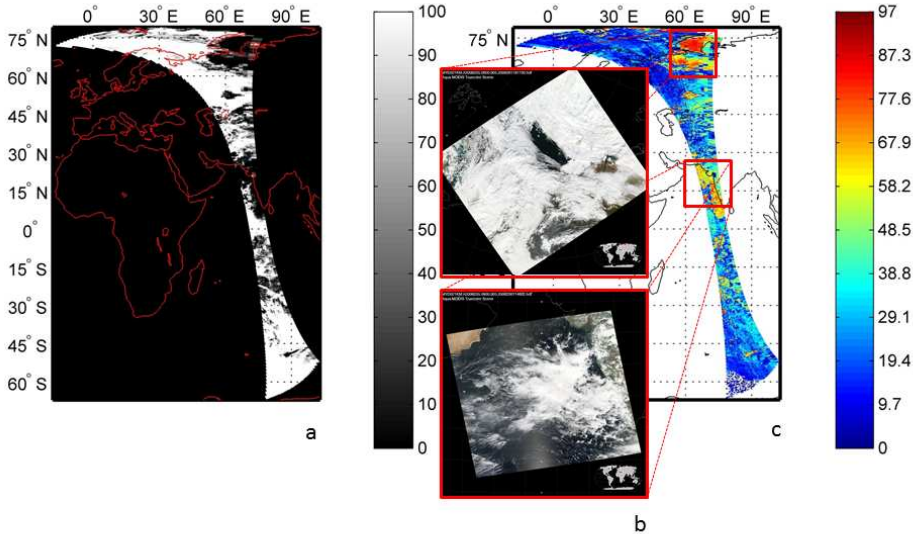

Fig. 9. Comparison of NN estimated cloud fraction and reference data for orbit 2006m0912t0828. **(a)** Computed MODIS geometrical cloud fraction re-located onto OMI grid. The color-code ranges from 0 (cloud free) to 100 (total cloud cover). **(b)** At those areas where the neural network fails in predicting cloud fraction, the corresponding MODIS RGB granule images are selected. In this case the MODIS image shows the presence of sun glint or ice as the reason of failure. **(c)** Absolute difference between the MODIS geometrical cloud fraction and the NN predicted cloud fraction. The color-code ranges from 0 (perfect match) to 100 (complete mismatch).

AMTD
6, 1649–1681, 2013

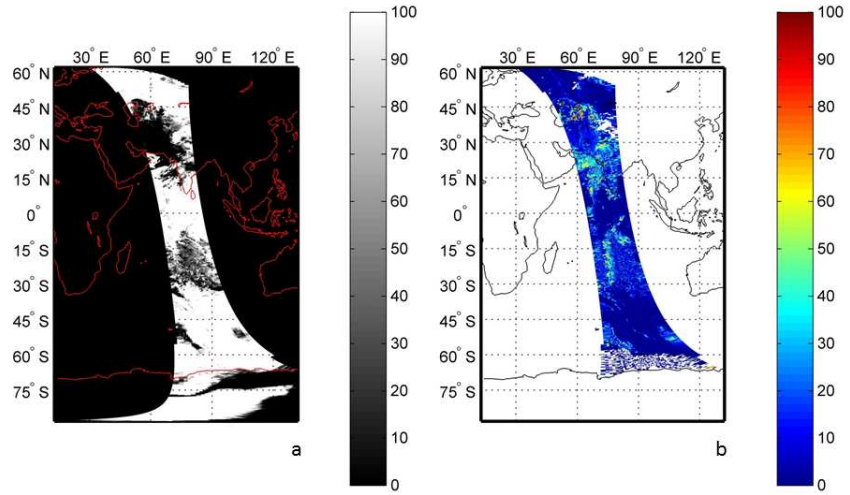

Fig. 10. Comparison of NN estimated cloud fraction and reference data for orbit 2006m0113t1325. **(a)** Computed MODIS geometrical cloud fraction re-located onto OMI grid. The color-code ranges from 0 (cloud free) to 100 (total cloud cover). **(b)** Absolute difference between MODIS geometrical cloud fraction and NN predicted cloud fraction. The color-code ranges from 0 (perfect match) to 100 (complete mismatch).

AMTD
6, 1649–1681, 2013

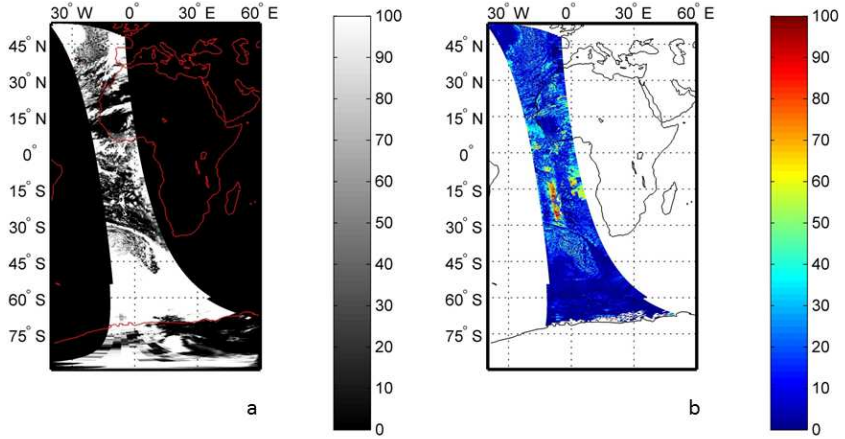

Fig. 11. Comparison of NN estimated cloud fraction and reference data for orbit 2004m1028t0755. **(a)** Computed MODIS geometrical cloud fraction re-located onto OMI grid.The color-code ranges from 0 (cloud free) to 100 (total cloud cover). **(b)** Absolute difference between MODIS geometrical cloud fraction and NN predicted cloud fraction. The color-code ranges from 0 (perfect match) to 100 (complete mismatch).

AMTD
6, 1649–1681, 2013

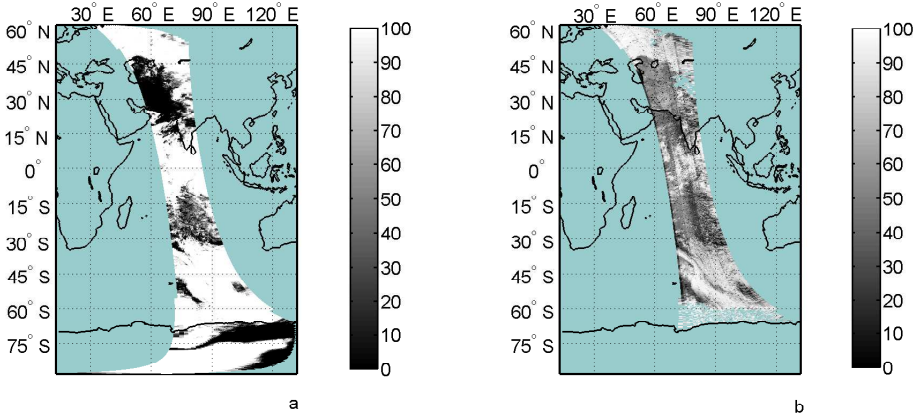

Fig. 12. Independent validation performed for orbit 2006m0113t1325. In both images the color-code ranges from 0 (cloud free) to 100 (total cloud cover). **(a)** Computed MODIS geometrical cloud fraction re-located onto OMI grid. **(b)** ELM predicted cloud fraction performed independently from training phase.

AMTD
6, 1649–1681, 2013