Saponaro et al., ‘Development of a neural network model for cloud fraction detection using NASA-Aura OMI VIS radiance measurements’, amtd-6-1649-2013

Response to referee comments

First, we would like to thank the editor and the four anonymous referees for their critical but fruitful comments which much improved the quality of our manuscript ‘Development of a neural network model for cloud fraction detection using NASA-Aura OMI VIS radiance measurements’.

The aim of our manuscript was to develop a fast and automated approach for detecting cloud fractions in OMI measurements, hence investigate the application of neural networks for this purpose. We have trained two neural networks with a reference dataset consisting of Aqua-MODIS cloud fractions. Once the neural networks were successfully trained, MODIS data are no longer needed and the resulting coefficients can be applied directly to any OMI orbit. After the comments from the reviewers, the authors included in their work a validation of the two developed algorithms on an independent dataset. The results clearly show the NNs capabilities to generalize the cloud fraction estimation. In addition to extra work with new orbits, the manuscript has been significantly re-written and a reduced set of new figures is provided.

In the following, we respond to the comments from the four referees. We have included the comments in *italics* and the responses are in normal text.

Anonymous Referee Nr.1

MAJOR AND GENERAL ISSUES:

1) The authors base their discussion and their conclusions on the results over the training dataset only, except for one single validation with an independent orbit (Fig.12). To me, it makes no sense to discuss the outputs of the training datasets, because it is only with an independent validation set that you can test the performances of the algorithm. So, in my opinion, Fig.s 3-11 are not really interesting. I suggest to select only a few Fig.s to report the performances on the training dataset (maybe Fig. 6, 8 and 9), if you really need them to discuss some aspects of your algorithm (the causes of large deviations with respect to MODIS reference data, e.g., as resulting from Fig.s 8-9). And, most important, the authors must provide more validation and discuss the performances of the algorithm based on the validation. The lack of a real validation is the main flaw of this manuscript and I recommend rejection if this aspect is not properly addressed in the revised manuscript.
The aim of this manuscript is to present neural networks as an alternative approach for cloud detection focusing on the methodology part. For this reason the validation/testing of the described method was at first thought to be outside the scope of the manuscript. However, as we explain in the revised MS the dataset has been changed. It has been expanded to 11 orbits divided in 3 subsets: a training set (6 orbits), a testing set (3 orbits) and a validation set (2 orbits). Only results from validation on the two independent orbits are provided and discussed in the MS. In this way the generalization capabilities of the NNs are assessed.

2) Even if we consider the training dataset only, the results shown in, e.g., Fig.s 3, 4, 7 are not really encouraging. Values of 0-20% for the "Correctly detected pixels" are found for clear sky pixels on some orbits. It is worth nothing that you say that the selected orbits for the training are almost completely cloudy: you just need to choose other orbits to train the NN. So, I recommend reconsidering the training dataset to include more representative orbits and/or to reconsider the design process of the NNs to obtain better performances on the training dataset.
The dataset used in the present work has been modified as explained in the previous comment. In the validation, both NNs present inaccuracies to estimate cloud-free pixels over land bright

surfaces as shown in Fig.6.d,e and Fig. 7.d, e in the MS. This is a consequence of NNs inability to discriminate high values in reflectance backscattered from clouds and from surfaces.

3) The authors need to give more details on the singular value decomposition of the input OMI spectra. Why did you choose 20 eigenvalues? I recommend to add a figure showing the eigenvalues as a function of the eigenvectors to show that cutting off at 20 is an optimal choice.
These choices have been changed in the revised MS. The authors decided to train each NN with a separate model for land and for ocean pixels due to their different characteristics. In order to avoid over-fitting the NNs, an optimization of the number of eigenvalues necessary to represent the spectrum and of the hidden neurons was necessary.
The approach suggested by Reviewer#2 in comment #1 was implemented. Several NNs were trained with BP and ELM and their performances over a testing subset were monitored by MSE (for BP) and RMSE (for ELM). The combination of eigenvalues and hidden neurons providing the best performance (over the testing data) is the optimized BP/ELM NN.
The ELM structure is designed so that the number of hidden neurons is determined by the number of inputs multiplied by a multiplication factor. This factor and the number of eigenvalues are the parameters to be optimized. We tested several NNs with eigenvalues ranging from 1:10 and a multiplication factor from 1 to 15 for land and ocean separately. The best performances over ocean were achieved with 2 eigenvalues and 75 hidden neurons given by the multiplication of the training inputs (2 eigenvalues for SVD + small pixel reflectance + OMLER data + solar zenith angle) and the optimized multiplication factor (15). Over land, the best combination consisted of 3 eigenvalues and 24 hidden nodes. Fig.1 shows which combinations of hidden nodes and eigenvalues provide the smallest RMSE in the optimization of the ELM algorithm modeled for ocean and land data.
The same procedure was used to optimize the BP NN. In this case the parameters considered were the number of eigenvalues and of hidden nodes. Figure 2 presents the results for land and ocean.

4) I suggest to modify the title to "A neural network model for cloud fraction detection using NASA-Aura OMI VIS radiance measurements".
The title of the revised MS is: “A neural network algorithm for cloud fraction estimation using NASA-Aura OMI VIS radiance measurements".

5) The definition of MODIS data as "reference inputs" of the algorithm is very confusing. Indeed, the MODIS data are "reference outputs". Please correct throughout the text and modify Fig. 2 accordingly.
In the revised MS we now refer to MODIS data as reference data. Figure 2 has not been modified: in the training phase of the NN, the MODIS cloud fraction is included in the training dataset (hence it is not used as output reference). It is only in the validation, that MODIS cloud fraction is used as reference data for comparing it with the NN predicted cloud fraction.

6) The manuscript is not well organized and very hard to read and understand. The written English needs to be improved. In the following, I suggest some very obvious corrections, but I recommend to ask an English speaker to copy-edit your manuscript. Please make a particular effort to clarify the introduction and the conclusions, as I found them very confusing.
Ok. The Authors will take care of reviewing the quality of the text giving particular attention to the introduction and results sections.

<u>MINOR COMMENTS:</u>

Page 1650

1) L3: "visual"->"visible"
Accepted.

2) L5-7: the sentence "Also...atmosphere" is not clear. You want to say "detection" or

"discrimination" instead of "distinction"? "Essential for" instead of "Essential to"?
The sentence has been deleted.

3) L8: "from NASA-Aura Ozone (OMI Monitoring Instrument (OMI) observations"
The sentence has been rephrased.

4) L9: "mathematical" seems not pertinent here, please remove
The sentence has been rephrased.

5) L9-10: "simultaneous application to OMI and Aqua-MODIS data" is very confusing. You wanted to say that OMI and MODIS are inputs/(reference)outputs of the NN? Please rephrase
The sentence has been rephrased.

6) L12: remove "approach"
The sentence has been rephrased.

7) L12: "Highly reflective..."->"However, highly reflective..."
Accepted.

8) L18-19: "...(TOA) reflectance..."->"...(TOA)...", as "reflectance" is said twice
Accepted.

9) L21: "...are discarded.", you need to add a reference here
The following reference has been added:
Martins, J. V., Tanré, D., Remer, L., Kaufman, Y., Mattoo, S., Levy, R.: MODIS Cloud screening for remote sensing of aerosols over oceans using spatial variability, Geophys. Res. Lett., 29(12), doi:10.1029/2001GL013252, 2002.

10) L21: "Commonly"-> "Usually"
Accepted.

11) L21: "...is performed using several tests.", you talk about the methods described at L23-etc? Please clarify
The sentence has been rephrased.

12) L25: "...with information from..."->"...with additional information coming from..."
Ok.

Page 1651

13) L1: "...from large quantities..."->"...from a large quantity..."
The sentence has been deleted.

14) L2: "...task and, in addition..."->"...task. In addition..."
Accepted.

15) L3: "efficient" doesn't seem the proper word here. You mean "fast"?
Accepted.

16) L6: "...on OMI cloud screening..."->"...on the cloud screening of OMI observations..."
The sentence has been rephrased.

17) L7: you talk about thermal channels: why? You wanted to say that TIR channels are useful for the cloud detection? Please explain and possibly reference to existing *literature.*

The TIR band allows measuring the thermal radiation emitted from clouds, thus increasing the information available to the algorithm to estimate cloud fraction. The following reference has been added:

Rees, W. G., Physical principles of remote sensing, 3rd ed., ISBN 978- 1-107-00473-3, Cambridge University Press, 2012.

18) L9: what do you mean with the word "combines"? Please explain

The Authors meant "to employ".

19) L13: "In this work we propose...", eliminate "In this work" and don't start a new paragraph

Ok.

20) L14: "VIS" (visible) is not defined. It has been defined in the abstract but it needs to be defines also in the main text

In the revised MS the acronym has been defined accordingly.

21) L15: you use "OMI" at L6 but you define the acronym only here

In the revised MS the acronym has been defined accordingly.

22) L18-19: "AQUA"->"Aqua", "AURA"->"Aura"

Accepted.

23) L23: the paper Sellitto et al., 2012 doesn't talk about the prediction of atmospheric parameters. As the authors can easily see from just the title, it talks about ozone retrievals, so please put it at L25, with the other ozone retrieval references

Thanks. The reference to Sellitto et al. (2012) has been moved in the text.

24) L25-26: the references Del Frate et al., 2002,2005a, and Iapaolo et al., 2007 are not the most recent papers on ozone retrievals with NNs. Indeed, here are two references of NNs algorithms to invert OMI radiances that are more relevant for your paper and need to be referenced:

Sellitto, P., Bojkov, B. R., Liu, X., Chance, K., and Del Frate, F.: Tropospheric ozone column retrieval at northern mid-latitudes from the Ozone Monitoring Instrument by means of a neural network algorithm, Atmos. Meas. Tech., 4, 2375-2388, doi:10.5194/amt-4-2375-2011, 2011

Di Noia, A., Sellitto, P., Del Frate, F., and de Laat, J.: Global tropospheric ozone column retrievals from OMI data by means of neural networks, Atmos. Meas. Tech. Discuss., 5, 7675-7727, doi:10.5194/amtd-5-7675-2012, 2012

The references Del Frate et al., 2002, 2005a, and Iapaolo et al., 2007 have been substituted with the two references suggested by the Referee.

Page 1652

25) L2: eliminate "task"

The sentence has been rephrased.

26) L7: Please specify that MODIS data are taken as "truth reference" or "reference output"

Accepted.

27) L8: if the MODIS pixel is smaller, its spatial resolution is higher than OMI, not smaller. Please correct

Accepted.

28) L8-10: the sentence "Hence the OMI...real OMI data." is not clear. Please rephrase
The sentence has been removed.

29) L11: "...the adopted design for the cloud..."->"...the design of the cloud..."
The sentence has been rephrased.

30) L13: "employed"->"used"
Accepted.

31) L14: eliminate "ones"
Accepted.

32) L16: eliminate "separated"
Accepted.

33) L19-20: the authors say "The use of only 4 orbits is not sufficient...": so why you don't use more orbits to train the NNs?
In the revised MS 6 orbits are included in the training dataset as explained in the general response.

34) L20: what do you mean with "good" solution? Maybe you should mention the concept of "generalization"
The sentence has been removed.

35) L24: "...also for the TROPOMI/VIIRS one without..." ->"...also with TROPOMI/VIIRS without..."
The sentence has been removed.

36) L24: please define the acronyms "TROPOMI" and "VIIRS"
The sentence has been removed.

37) L24: your OMI/MODIS NN will not work with TROPOMI/VIIRS. You would need to re-train the NN with different data and maybe also change some other aspects of the design. From my point of view, the only thing in common for the two algorithms is that they both are NNs...
The sentence has been removed.

38) L25: "...the added benefit of the oxygen a-band..."->"...the added benefit of the presence of the oxygen a-band..."
The sentence has been removed.

Page 1653

39) In the description of OMI it seems that OMI only takes measurements in the VIS. Even if only VIS measurements have been used in this context, the authors need to state that OMI has also two UV channels (UV-1: 270 to 314 nm, UV-2: 306 to 380 nm UV-1: 270 to 314 nm, UV-2: 306 to 380 nm)
Accepted.

40) L3: why not using the acronym "OMI"?
Accepted.

41) L5: "...in the visible parts of the electromagnetic spectrum..."->"...in the VIS..."
Accepted.

42) L5-6: "...is 13x24 km2 at nadir."->"...is 13x24 km2 at nadir, in the normal global operation mode." (There is also a "zoom mode")
Accepted.

43) L8: "Row Anomaly"->"row anomaly"
Accepted.

44) L8: put a reference for the row anomaly
Accepted.

45) L10: "earth"->"Earth"
The sentence has been rephrased.

46) L10-11: "...is MODIS which..."->"...is MODIS, which..."
The sentence has been rephrased.

47) L13: "...7 minutes the MODIS..."->"...7 minutes, the MODIS..."
The sentence has been rephrased.

48) L18: maybe "variables" is better than "parameters" here
Accepted.

49) L21: I'm not sure that "to detect" is pertinent here, maybe "to measure" is better
Accepted.

Page 1654

50) L1-2 "...is considered here which is referred to as multilayer perceptron..."->"...is the multilayer perceptron..."
Accepted.

51) Fig. 1 (caption): "...feedforward topology..."->"...feedforward structure..." (the figure is general, while the topology depends on the number of inputs/hidden/outputs neurons, for your specific problem)
Thanks. Accepted.

52) Fig. 1 (caption): x_n actually represents the last neuron, not the generic one (I would use another index, e.g. "i" to refer to the generic index), in this figure. Has this figure been produced by the authors? Otherwise please cite the reference of the source.
Accepted. The figure has been created by the first author.

53) L4: "...comes in at the..."->"...is collected by the..."; you can also eliminate "of the network" (it is obvious)
Ok, thanks. Accepted.

54) L7: "into"->"to"
Accepted.

55) L8-11: the sentence is very confusing, please rephrase. In general, the authors never talk about the "neuron" (they use the word "unit") or "activation function", which is very confusing
The Authors have included a definition for 'node' and 'activation function' in the section Neural Networks.

56) L12-13: with "the model of each node" you mean "the activation function of the neurons"?
Yes, thanks.

57) L15: "...neural networks theory..."->"...neural networks..."
Accepted.

58) L16: "...multilayer perceptron network..."->"...multilayer perceptron..."
Ok.

59) L18: "...back-propagation, algorithm..."->"...back-propagation algorithm..."
Ok.

60) L26-28: if you say "The training phase of a NN...", the sentence seems more general than a specific discussion on the back-propagation. The authors wanted to be more general here?
No, the sentence strictly refers to a back propagation neural network. The sentence has been modified accordingly.

Page 1655

61) L9: this sentence is very general: in this subsection weren't you specifically talking about the back-propagation training algorithm? (and here you finally mention the "activation function", but without defining it)
The definition of an activation function is introduced in the section Neural Networks. The sentence at L.9 has been substituted with a more detailed one explaining why sigmoid activation function is useful for BP.

62) L10: "adopted"->"used"
Ok.

63) L12: define symbols of Eq. 1
Ok.

64) L13-17: now you're talking about your algorithm? Is subsection 3.1 a general description of the back-propagation learning algorithm? I personally think that this subsection must be completely reconsidered
Authors agree with the Reviewer. Section 3.1 and section 3.2 are meant to be generally introducing the definition of the BP and ELM.

65) L23: "Extreme"->"extreme"
Ok.

66) L25: I missed where the authors introduced the "bias term" in the general description of the NNs
Its definition is introduced in the section Neural Networks.

Page 1656

67) L3: you use again "activation function", which is not defined
Its definition is introduced in section Neural Networks.

68) L5: again, the authors need to define the biases before (when they describe the general concepts of NNs)
Their definition is introduced in section Neural Networks.

69) L14: "provides"->"requires"
Ok.

70) L14-16: you say that the extreme learning machines reach smaller errors and are faster, so why using the back-propagation?
The sentence has been removed.

71) L18: "resolved from"->"found with"
Ok.

Page 1657

72) L2-4: the sentence "The drawback...randomly" is not clear. What do you mean with "scaling parameters"? Please clarify
The sentence has been removed in the revised MS.
After repeating the optimization of the NN parameters, an optimized NN structure has been selected. The number of hidden neurons of ELM has been reduced but it is still higher than the number of hidden neurons required by BP. The connection between the random choice of these scaling parameters and the number of hidden neurons could be explained as follows. Because the weights from the input to the hidden layer are not based on the NN training but randomly chosen, a NN trained using the ELM is not able to learn which input variables are more important for determining the output. This makes ELM much more sensitive to inputs poor in information contents than a NN trained, for example, with BP. This occurs because a high number of hidden neurons reduces the probability that some inputs are given near-zero weights by the random initialization (which would result in the loss of useful information by preventing those inputs to be used in the NN to estimate the output).

73) L9: "observation uncertainties" seems not proper here, since you don't use uncertainties as inputs of the NN
We agree with the Referee. The sentence has been rephrased.

74) L17: "unity"->"one"
Ok.

75) L17: calculating the reflectances does not assure to fully exploit the dynamical range of the input neurons (0-1). The authors should report the maximum and minimum value for the spectral inputs or rescale the inputs between 0 and 1.
Reflectances are normalized so that values are between 0 and 1.

76) L22-23: do you mean that the small-pixel data is provided at one single wavelength?
Yes. In the VIS band the small pixel data are provided only at 388 nm in the VIS. This information has been added in the MS.

Page 1658

77) L17-18: the authors don't mention at all the beneficial effect of a dimensionality reduction for the generalization capability of a NN (limitation of the overfitting effect and less local minima in the mapping function). This is important and must be discussed
In the revised version of the MS is explained how a correct reduction of the input variables can positively affect the generalization performance of the NN.

Page 1659

78) L11: what do the authors mean with "initial testing"?

We meant that additional eigenvalues would not result in higher MSE. Anyhow, the sentence has been removed from the revised MS because the results have changed.

79) Sub-section 4.3: please specify that the MODIS cloud fraction data are reference "outputs"

This section describes how MODIS data are used in the training phase of NNs. During the training of NNs, the MODIS cloud fraction data is included in the training dataset and it is not used for assessing the correlation between the estimated NN cloud fraction and MODIS cloud fraction (this is the case of validation where MODIS cloud fraction represents the reference output).

80) L14: you may want to say "...to be matched with the methodology described by Stammes et al. (2008)"

This sentence has been rephrased in the revised version of the MS.

Page 1660

81) L8-9: isn't the sentence "The neural...Fig.1" a repetition?

Ok. The sentence has been modified so that repetitions are avoided.

82) L14: why "Extreme Learning Machine" has initials in uppercase?

Modified accordingly.

83) L17-18: you say "For the back-propagation algorithm, 25 hidden nodes guarantee a good performance either in terms of training accuracy or training time", so you tested bigger NNs, with more hidden neurons? So the training time is not an issue? Trade-offs of hidden neurons is usually searched based on the accuracy and the generalization capability, by analyzing the results over an independent dataset (usually referred to as "test dataset", not to be confused with the "validation dataset"). I recommend to discuss the choice of the hidden layer dimensionality in terms of these quantities and not the training time.

The paragraph has been completely changed and a more specific explanation about NNs optimization is given. Please refer to 'Major and general Issues - Comment #3' of Anonymous Reviewer #1 for a more extensive explanation.

84) L22: why "Extreme Learning Machine" has initials in uppercase?

Corrected.

Page 1661

85) L1-10: there are lots of repetitions. Please be more concise

Text has been revised.

86) L11: "Data was..."->"Data were..." ("data is plural")

Ok.

87) L14: the authors talk about the results of Fig. 3 but they don't describe the figure in the text. Please first introduce the figure and then comment it

Figure 3 has been deleted and substituted with a different bar chart which is described in the Results chapter.

88) Fig. 3: as said in my major revision 2), the algorithm does not work (e.g., 0% clear sky detected pixels for the third orbit).

Results are now different due to a change in the NN structures.

89) L15: "observed"->"considered"

Ok.

90) L22-24: the authors say "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." To me this statement sounds strange. Cloud free and cloudy pixels are complementary (if a pixel is cloudy it is not cloud free, and viceversa), so if you have information on cloudy pixels you have also information on cloud free pixels. Please explain or correct your statement
The statement has been removed from the MS because the results have changed.

91) L26: "...analysed the same data but trained the learning algorithms with separated..."->"...analyzed the outputs of similar NN trained with separated..."
Ok.

92) L28: "efficiency" seems not pertinent here
Changed to “accuracy”.

P1662

93) L2-5: the authors say that worse performances over ocean than land are unexpected, and I agree. In any case, please attempt to investigate or at least to propose possible reasons for this unexpected behavior
The recent results are accurate over ocean while overestimated cloud fractions are detected over land bright surfaces.

94) L6-7: I think that the sentence "To supply...phase" is a repetition
The sentence has been deleted.

95) As said in the major revisions 1-2), do you really need to show all your tests if only this latter NN, with 60% threshold, works?
Ok, these results have been removed.

96) L13-14: "fully clouded", you mean that there not even 1 pixel with CF<100%? Maybe histograms with MODIS reference cloud fractions would help in this context
The sentence has been removed because the results have changed.

97) L14: put a space between "representative" and "for"
Ok, thanks.

98) L16: "ground pixel coverage type" seems not the best way to define your "Land/Water" flag
Ok.

99) L18: "predicted"->"estimated"
Ok.

100) L20-21: "The color scale...points", is it the number of points or the density?
It represents the density. It has been changed in the text of the revised MS.

101) Fig. 6: please annotate the units next to the colorbars, here and in every figure
Ok.

102) Fig. 6 (caption): again, "predicted"->"estimated"; "A good correlation...and ocean", this must be said in the main text and not in the figure’s caption
Ok.

103) L26: the authors talk about biases, but where are they reported? Not in the figure, as they claim here
The statement is cancelled due to changed results.

P1663

104) L18: delete "signifies"
Ok.

105) L23-24: you actually don't show "good performances" in Fig.s 10-11, but "MODIS geometrical cloud fraction...etc...for two orbit characterized by better performances than those shown in Fig.s 8-9..." or something similar (please explain the figures in the main text)
Due to changes in the MS, Figs.8-9-10-11 have been removed and replaced with two new figures.

106) Fig. 11: there are regions with about 100% mismatch, so this is not characterized by "good performances" over the whole orbit. Please clarify and attempt to explain for these marked mismatches
Figure 11 has been removed. In the latest results shown in the MS, over bright surfaces both NNs have mismatches between MODIS cloud fraction and the predicted cloud fraction. These mismatches are larger for BP than for ELM.

107) Fig. 9: please change the extreme values of the colorbar (c) to 0-100, as done for Fig.s 8, 10 and 11.
Fig.9 has been removed but colorbars are included in every new plot.

108) Fig. 8: "...MODIS image shows the presence of dust as the reason of failure...", actually MODIS shows the presence of dust and you suppose that this is the reason of mismatch (that's not shown by the MODIS image, it's your supposition): please rephrase accordingly
Corrected.

P1664

109) L2-3: "The resulting predicted cloud fraction is rather inaccurate", so the only validation you provide shows that your algorithm doesn't work/doesn't generalize from the training?
In the revised MS we have extensively modified the NNs (from increasing the dataset to an exhaustive optimization of NNs parameters) which now provide much better generalization performances as shown by Fig.6-7. The results have been validated with an independent data set which was not used in training or testing.

110) L9-10: "However...Fig. 12", the sentence is not clear, please explain better
The sentences have been removed.

111) L14-17: in my opinion, the statement "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" is not true, because a NN for TROPOMI/VIIRS will need re-design and re-training. Please modify or eliminate the sentence
Authors agree with the Referee. the sentence has been removed.

112) L18: this is not a "comprehensive" study. I would say, on the contrary, that this is a "preliminary" study

We agree that this is not a comprehensive study but a rather a study to evaluate the concept.

113) L23-25: you say more than once that the back-propagation is extremely time consuming with respect to ELM: please quantify
The sentence has been removed.

114) L25: you say that all the orbits are almost totally cloudy, again, so why not using other orbits?
In the revised MS the training dataset has been expanded to 6 orbits.

P1665

115) L3-5: you claim that "the spectral features alone can discriminate cloudy from clear pixels", but this is not true, because you show that cloud free pixels are often not correctly detected
The results have been improved and the sentence has been modified.

116) L10: "learning"->"training"
Ok.

117) L9-12: your description of future work is somewhat mysterious... If you wish to talk about it, please give more details
Sentence has been removed.

118) L25-26: again, this is not really portable to TROPOMI: you'd need to re-design and to re-train the NN. In addition, this sentence is a repetition: please delete
We agree and the sentence has been removed.

P1666

119) L1: please give some more details on how do you want to use the oxygen A-band in your future work with TROPOMI.
The text referring to the TROPOMI application has been modified.

120) L5-7: "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", I strongly disagree on this statement. In my opinion, if you don't provide a validation of your NN and you don't demonstrate its generalization to data not present in the training dataset, your work is not at publication level.
We agree, The NNs generalization performances have been tested on 2 independent orbits. Results are presented in the revised MS.

Anonymous Referee Nr.2

MAJOR AND GENERAL ISSUES:

The Authors have developed a neural network scheme to estimate the cloud fraction of OMI pixels using OMI VIS reflectance spectra. In my opinion, the title of the paper is slightly ambiguous, because "cloud fraction detection" is something like an hybrid between "cloud fraction estimation" (that is, assigning a continuous cloud fraction value to each pixel) and "cloud detection" (that would make me think that the output of the NN is a binary decision between a cloudy and a non cloudy pixel). This subtle ambiguity continues in the Introduction, because the Authors alternatively speak about "focusing on cloud screening" (P1651, L6); "a novel approach using neural networks for the direct determination of the pixel cloud fraction" (P1651, L13-15) and "the application

of a NN as an alternative approach to the cloud screening task" (P1652, L1-2). To my understanding, the NN proposed by the Authors gives a continuous cloud fraction as its output. In view of this, I would change "detection" to "estimation" in the title, and I would suggest the Authors to make the aforementioned parts of the Introduction more uniform.
Thanks for the suggestion. We have changed the title and tried to remove the ambiguities from the text.

1. In order to design their NN for cloud fraction estimation, the Authors compare two NNs trained with two different algorithms, namely standard backpropagation and ELM. Then, they judge the NN trained with the ELM to be better, because it trains faster and achieves smaller errors on the training set. As explained in any good textbook about neural networks and – in general – nonlinear regression, selecting the best regression model based on the performances on the training set is a serious methodological error, because it might lead to the selection of overfitted NN models, i.e. models that are able to reproduce the training data very well but give erratic results on unseen data. Let's clear out any misunderstanding: a NN that learns well but is uncapable of generalizing whatsoever is a NN that does not work! Such a NN would behave like a static memory and not as a regressor (it is able to "recall" almost exactly the correct response for every pattern already seen in the training set, but it is not capable of producing a reasonable output when a new input pattern is presented). It is not difficult to imagine that an overfitted NN is completely useless as a retrieval algorithm. In view of this, in order to decide which algorithm performs best and what is the most suitable number of hidden neurons for their NNs, the Authors should re-perform their comparisons as follows: (1) Split the dataset in a training and a validation subset. (2) Train several NN's with the backpropagation algorithm, eventually using some form of cross validation, and monitor the MSE over the validation subset as a function of the number of hidden neurons. The number of hidden neurons that leads to the best performance (on the validation subset) will make the best backpropagation NN, called NN_opt(BP). (3) Train several NN's with the ELM algorithm and monitor their RMSE on the validation subset as a function of the number of hidden neurons. The number of hidden neurons that gives the smallest RMSE on the validation subset will make the best ELM NN, called NN_opt(ELM). (4) The best network NN_opt will be the best between NN_opt(ELM) and NN_opt(BP) over the validation subset. The Authors should also consider performing an additional comparison between NN_opt(BP) and NN_opt(ELM) over a third independent set. This would be even more rigorous, but maybe it can be skipped if they do not possess enough data.
Thanks for the very important comment. As discussed above in the response to Referee #1, comment # 3, we have indeed followed this suggestion and added more data (orbits) and split the data set in a training, a test and a validation data set.

2. The Authors want to propose a novel method to estimate the cloud fraction from satellite data, but fail to produce any convincing demonstration that their algorithm can be actually used in an operational scenario. In fact, the only attempt they make to apply their NN to data that were not used in the training set gives quite poor results. They simply attribute this to an insufficiency of training data, but I would not be so sure about that, as other causes may explain this fact:
(I) Their model selection might have favoured an overfitted NN (see previous comment);
(II) The quantities used as inputs for the NN might not contain all the relevant information to detect cloud fraction, or the information might be "masked" by other irrelevant inputs (please note that the other methods that are used to estimate cloud fractions from UV/VIS/NIR hyperspectral observations usually focus on very specific wavelength intervals, rather than fitting a complete spectrum, and combine – or simultaneously retrieve – information on the cloud top pressure: such information is not used by the Authors, and this may well be one of the reasons why their NN is not working

outside the training set); (III) the co-location noise (caused by the differences in the field of view between MODIS and OMI, and by the motion of the cloud fields between the overpasses of the two instruments) and the different instrument sensitivities (to my awareness, MODIS is sensitive to a geometrical cloud fraction, whereas OMI is not) may be destroying the relationship between the OMI spectra and the cloud fractions used in the NN training dataset. In other words, the Authors should ask themselves how confident they are in the fact that a given OMI spectrum Y is actually produced by a cloud fraction X as measured by MODIS. The Authors seem to claim that this is not a real problem, because other papers have shown that OMI and MODIS cloud fractions "can be used together" (P1653, L13). However, the key question the Authors should address is "to do what" can OMI and MODIS cloud fractions be used together. For example, Stammes et al. (2008) - cited by the Authors in order to justify the use of MODIS cloud fraction as "truth" for the NN training - explicitly state that "the OMI effective cloud fraction c_eff is not a geometrical cloud fraction as retrieved by MODIS" and that "there is no direct method to compare the OMI c_eff with an existing MODIS product" (page 9 of the paper I am referring to). Therefore, they define an "effective cloud fraction" for MODIS, based on the cloud optical thickness, and compare it with co-located OMI cloud fractions retrieved with the O2-O2 absorption method. Even though they find a remarkable correlation coefficient, the scatter plot shown in Figure 9 of Stammes et al. (2008) shows a considerable spread between OMI and MODIS effective cloud fraction. Are the Authors sure that this fact does not affect the quality of their dataset?
Obviously there will be differences between MODIS and OMI effective cloud fraction due to the 7 minutes time lag. However, the MODIS pixel is expected to provide much more reliable cloud fraction information than the large OMI pixel and we do not see an alternative to do better.
The explanation of the poor results may be an overfitted NN. After the optimization of the NN parameters, both NNs provided good generalization performances.

<u>MINOR COMMENTS</u>

1. Title. Besides changing "detection" to "estimation", I would suggest the Authors to change "model" to "algorithm".
Ok.

2. P1650, L5. is -> are.
The sentence has been removed in the revised MS.

3. P1650, L6. I would say that "this paper reports on the development of a neural network algorithm to estimate cloud fractions from ...". In fact, the output of the NN is a cloud fraction and not a cloud detection (which would require a binary output).
The sentence has been rephrased in the revised MS.

4. P1650, L8-10. I do not like the sentence "We present . . . (MODIS) data" (by the way, "mathematical neural network" sounds quite trivial). I would say something like "The proposed NN is trained using OMI reflectance spectra, solar zenith angle and OMI climatological surface reflectance as input data and Aqua-MODIS as target data".
The sentence has been rephrased in the revised MS.

5. P1650, L19. Are you sure that Andreae and Rosenfeld (2008) is an appropriate reference for this statement?
We agree. The reference has been deleted.

6. P1650, L19-21, sentence "for instance . . . pixels are discarded". Could you provide a citation for this?
Included reference to:

Martins, J. V., Tanré, D., Remer, L., Kaufman, Y., Mattoo, S., Levy, R.: MODIS Cloud screening for remote sensing of aerosols over oceans using spatial variability, Geophys. Res. Lett., 29(12), doi:10.1029/2001GL013252, 2002.

7. P1650, L24, sentence "The most consolidated methods . . . radiative transfer models".
Could you cite some examples of these methods as well?
The following references have been added:
Wu, D. L., Jiang, J. H., and Davis, C. P.: EOS MLS cloud ice measurements and cloudy-sky radiative transfer model, IEEE Trans. Geosci. and Remote Sens., 44(5), 1156–1165, 10.1109/TGRS.2006.869994, 2006.
Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part I: Algorithm description.,
J. Appl. Meteor., 44, 39–54, 2005.
Loyola R., D. G.: Applications of neural network methods to the processing of earth observation satellite data, Neur. Netw., 19, 168-177, doi:10.1016/j.neunet.2006.01.010, 2006.

8. P1650, L25. I do not understand the meaning of the sentence "The application . . . observer dependent". If the Authors refer to the cloud fraction estimation, then I would say that the state-of-the-art methods to retrieve cloud fractions (O2-A band, O2-O2, Rotational Raman Scattering) are based on physical considerations, and are not observer dependent. If they refer to the thresholding of the estimated cloud fractions, then I do not see any difference between their NN method and others with respect to that. To my understanding, their NN only provides an estimate for the cloud fraction. It does not provide anything like an automatic threshold for cloud masking. In fact, later in the paper it is explained that, in order to perform cloud masking, the Authors threshold their estimated cloud fractions empirically as well, trying 30% and 60% as threshold values, and showing that only pixels with cloud fractions larger than 60% are detected with reasonable accuracy.
We agree. The sentence has been cancelled.

9. P1650, L25. Some discussion of the physical approaches to cloud fraction estimation from UV/VIS/NIR instruments might be worth at this point. For instance, the Authors might want to discuss the differences between their approach and the algorithms proposed by Joiner and Bhartia (1995), Koelemeijer et al. (2001), and Acarreta et al. (2004). Such a discussion could also help clarify why the Authors chose to include all the OMI VIS channel in the input vector rather than concentrating only on a specific spectral interval that is known to be sensitive to clouds (e.g. the O2-O2 band) with limited interferences from other factors except surface albedo.
In the present work, the Authors do not include all 751 OMI VIS wavelengths but they compress the relevant information content into a smaller number of features by applying a single values decomposition method. In the revised MS, the references Joiner and Bhartia (1995) and Koelemeijer et al. (2001) have been removed.

10. P1651, L17. Remove "mask" at the end of the sentence. Furthermore, I do not agree with the statement that the proposed NN uses auxiliary cloud information from MODIS to determine the presence of clouds. MODIS data are just used as reference values to construct the training dataset, but the NN operation is entirely based on OMI data as inputs.
The sentence has been rephrased in the revised MS.

11. P1651, L23. Why do you place Sellitto et al. (2012) in "prediction of atmospheric parameters" rather than "ozone retrievals"?
The reference has been moved in the text as suggested.

12. P1651, L25. Del Frate and Schiavon (1998) is not about satellite observations. It is about measurements from a ground-based microwave radiometer.
Text has been revised.

13. P1651, L27. Since this paper is about cloud fraction estimation, it might be appropriate to cite previous applications of NN's to similar tasks, like Loyola et al. (2007, 2010). The Authors might also want to add a discussion of the differences between their approach and that described by Loyola et al. (2007), where the cloud fraction is estimated outside the NN and the NN is then used to retrieve cloud albedo and cloud top height.
The suggested paper (Loyola et al., 2006) is now added to the references. Although, Loyola et al. (2006) use NN trained with radiative transfer simulated data to estimate cloud properties rather than to directly retrieve cloud fraction.

14. P1652, L1-10. I would suggest merging this paragraph with that at P1651, L13-19, eventually removing repetitions.
Text has been revised.

15. P1652, L17. If the four orbits the Authors refer to are those included in the training set, then these are not "four random orbits", and their example of application does not represent a performance test at all. In principle, they could have even chosen more hidden neurons, so as to achieve near-zero error on the training set (this is especially true for the ELM NN, where the training error approaches zero as the number of hidden neurons approaches the number of training data), but this would not tell anything about the real performance of the algorithm when applied to new cases.
The dataset has been modified as follows: 11 new orbits were randomly selected and divided in three sub-sets: 6 orbits for training set, 3 for testing, and 2 for independent validation. Also, the optimization of the neural networks (both backpropagation and extreme learning machine) is optimized by minimizing the RMSE/MSE (as suggested by Reviewer#2).

16. P1653, L17. "Neural networks algorithms" -> "Neural network algorithms".
Ok.

17. P1654, L2. It could be good to cite Werbos (1974) after "MLP".
Accepted. The following reference has been added:
Werbos, P. J.: Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, PhD thesis, Harvard University, 1974.

18. P1654, L18. "batch version of it" -> "its batch version". Maybe it would be worth to clarify the meaning of "batch" in this context, so as to make it clear for the occasional reader who is not an expert in NNs.
The sentence has been removed.

19. P1654, L24. Please cite Rumelhart et al. (1986) after "The error back-propagation algorithm".
Ok.

20. P1654, L27. As far as the standard backpropagation algorithm is concerned, the minimization of the cost function occurs pattern by pattern. Therefore, the cost function is not a mean square error function, but is the square norm of the error on the pattern itself. In the case of batch backpropagation it is correct to say that the cost function is a mean square error function.
We agree. We now refer to the square norm of the error in the revised MS.

21. P1655, L14. "inputted into" -> "input to", or "applied to the network as an input".. I

would remove “in its own node”, because it can be misleading and it adds nothing.
The sentence has been rephrased.

22. P1655, L15. Please remove “which is commonplace in classification applications” because (I) it is too generic; (II) what is shown here is not a classification application, as the output of the NN is not discrete.
The sentence has been removed.

23. P1655, L23. Please cite Huang et al. (2006) after “The Extreme Learning Machine”.
Ok.

24. P1657, L4. What do the Authors mean by “scaling parameters”? Maybe the weights from the input to the hidden layer? What is the connection between the random choice of these parameters and the need for more hidden neurons? To my understanding this occurs because having many hidden neurons (and hence many weights from the input to the hidden layer) reduces the probability that some inputs are assigned only near-zero weights by the random initialization (which would prevent those inputs to have an impact on the NN output). Am I correct? If this is the case, then I foresee two other drawbacks that are strictly related to that mentioned by the Authors. (I) Since the weights from the input to the hidden layer are not learned but chosen at random, a NN trained using the ELM is not able to learn which input variables are more important in order to determine the output (e.g. in the case of this algorithm, I could expect some reflectance singular vectors to contain more information about cloud fraction than others, etc.). Therefore, I would expect a NN trained with ELM to be much more sensitive to irrelevant inputs than a NN trained with the standard backpropagation method, or with any other algorithm that allows the input-to-hidden layer weights to be learned. This would ask for a very careful sensitivity analysis of the cloud fraction with respect to all the candidate inputs (to be done, e.g., with a RTM), so as to make sure that only quantities that are really relevant for the retrieval are included in the input vector. (II) If it is true that very large NN’s are necessary to achieve good performances with the ELM, then it means that this learning algorithm gives the designer no possibility to control the generalization capability of the NN by keeping its Vapnik-Chervonenkis (VC) dimension (Haykin, 1999) as small as possible.’
The Authors meant the weights from the input to the hidden layer**.** The comments given by the Reviewer are very interesting and could justify the NN behavior. Anyhoq, in the revised MS the NN has optimized and this lead the NN to have a better generalization ability.

25. P1657, Eq. 5. Pi should be at the numerator, not at the denominator.
The equation has been corrected.

26. P1657, L14. Why did the Authors not use the viewing zenith angle and the relative azimuth angle as inputs?
We indeed admit the oversight. Viewing zenith angles should have been included in the input dataset. We chose to include the solar zenith angle in the training dataset because we assumed it to have a larger variation in the data hence bringing more information to the dataset. The variation of the without the solar zenith angles did not have large impact on the final results.

27. P1658, L18. I would suggest changing “necessary to help . . . computing time” to “desirable in order to reduce the computation time”.
Ok.

28. P1659, L13. I would suggest removing “auxiliary”.
Ok.

29. P1659, L14. I would change the sentence “To this end . . . need to be matched”

to “The spatial matching between OMI and MODIS pixels was performed as follows”. I then propose to move the citation of Stammes et al. (2008) to a subsequent point in the paragraph (see next comment).
The text has been revised.

30. P1659, L21. After “OMI pixel boundary”, the Authors could say “as done in Stammes et al. (2008)”. Please change “matches” to “matchings” or “co-locations”.
The text has been revised.

31. P1659, L22. Remove “,” between “pixel” and “as”.
Ok.

32. P1660, L3. “For the purpose . . . training” -> “The training dataset for the NN consists of . . .”.
Ok.

33. P1660, L18-20. What is a “good performance in term of training accuracy or training time?”.
As suggested by the referee in general comment #1, the NN parameters (number of eigenvalues for SVD and hidden neurons) have been optimized by monitoring the MSE(for BP) /RMSE (for ELM) over the testing subset as a function of the number of hidden neurons and eigenvalues.
The combination giving the best performance with the test dataset are the parameters applied to NN for independent validation. This procedure for the optimization of NN parameters has been applied for land and ocean datasets.

34. P1660, L25. How many training data have been produced in total?
The dataset used in the revised paper consists of 11 random orbits divided in the following subsets: 6 for training, 3 for testing, and 2 for independent validation. The same dataset was applied to BP and ELM. Both NN have been trained with the same datasets divided into ocean and land data. So altogether, if the Reviewer is asking how many pixels ere used in the training the answer is 263461 pixels over ocean and 128941 pixels over land have been used in the training.

35. P1661, L10. “compare” -> “compared”.
Ok.

36. P1661, L19. "detect cloud fraction" -> "detect clouds"
Ok.

37. P1661, L26. “learning algorithms” -> “neural networks”.
Ok.

38. P1662, L9. It is said that increasing the amount of training data "is expected to enhance the overall performance". I would suggest the Authors to check immediately whether this is true. Downloading more OMI orbits and matching them with other MODIS data should not take too long. This would be very helpful in understanding whether the observed underperformance of the NN is merely due to a lack of training data or there are more fundamental issues, like those I pointed out above.
In the revised MS, results are included from optimizing the NN and also the dataset has been expanded.

39. P1662, L18. Remove the sentence “This process . . . each orbit” (it’s trivial, it directly descends from the previous sentence).
The sentence has been deleted.

40. P1663, L9. The Authors say that "the general features of the cloud fraction can

be observed in Fig. 12". However, the spatial patterns that I observe in the two panels of Fig. 12 are quite different. Can the Authors provide a scatter plot and some error statistics such as RMS error and correlation coefficient between retrieved and reference cloud fractions?
Figure 12 has been removed in the revised MS and proper validation was performed on two independent orbits. In the revised MS Fig. 4 and Fig. 5 show the validation results, i.e. a scatter plot and correlation of the results. Both NNs present very good correlation over ocean while over land ELM shows a correlation of about 0.70 and BP presents correlation of about 0.50.

41. P1666, L16, L21. There is something wrong with the citations. The journal for Ackerman et al. (1998) should be "J. Geophys. Res., 103 (D24), 32141-32157". The journal for Andreae and Rosenfeld (2008) should be "Earth Sci. Rev., 89, 13-42."
The references have been corrected. Thanks.

42. P1666, L23. Bartlett (1998) is not referenced in the text.
The reference has been removed in the revised MS.

43. P1667, L1. Ortezi -> Ortenzi
The reference has been removed in the revised MS.

44. P1667, L10. Roebling -> Roebeling
Ok.

45. P1667, L30. Joiner and Bhartia (1995) is not referenced in the text.
The reference has been removed.

46. P1668, L1. Vassilkov -> Vasilkov; Firts -> First.
The reference Joiner and Vasilkov (2006) has been removed.

47. P1668, L1, L5. Joiner and Vasilkov (2006), and Karayiannis and Venetsanopoulos (1993) are not referenced in the text.
The reference to Joiner and Vasilkov (2006) has been removed and Venetsanopoulos (1993) has been referenced in the revised MS.

48. P1668, L13. The authors are Koelemeijer, R. B. A. and Stammes, P. (de Haan, J. F. is not an author of the paper).
Ok, corrected.

49. P1668, L15. de Haan, J. D. -> de Haan, J. F.
The reference Koelemeijer et al. (2001) has been removed in the revised MS.

50. P1668, L17. Platnic -> Platnick
Ok.

51. P1670, L12. van der Oord -> van den Oord
Ok.

Anonymous Referee Nr.4

<u>GENERAL COMMENTS:</u>

The paper describes the development of neural network algorithms for cloud detection using NASA-Aura OMI measurements. The results obtained by training neural networks with OMI and MODIS data are presented and discussed. The results of OMI level 1b radiance measurements are compared with MODIS cloud fraction.

Exact cloud retrieval is an important task to determine the composition of lower atmosphere and surface characteristics. The described cloud detection system which is based on neural networks to develop an automated cloud clearing algorithm is well adapted to the fact that cloud properties are highly variable and difficult to detect. An optimal learning algorithm for the cloud screening task is selected. The described novel methodology for the direct retrieval of the cloudy or cloud-free pixels from VIS measurements by OMI contributes to the preparation of new satellite missions. The possibilities for the TROPOMI instrument to be launched in 2015 are discussed. Advantages and disadvantages of the method are discussed. The paper addresses relevant scientific questions within the scope of AMT. The paper presents novel concepts, ideas and tools. The scientific methods and assumptions are valid and clearly outlined so that substantial conclusions are reached. The description of experiments and calculations are sufficiently complete and precise to allow their reproduction by fellow scientists. The quality and information of the figures are very well. The related work is well cited as well as the number and quality of references appropriate i.e. the authors give proper credit to related work and clearly indicate their own new/original contribution. The title clearly reflects the contents of the paper and the abstract provides a concise and complete summary. The overall presentation is well structured and clear. The language is fluent and precise. The mathematical formulae, symbols, abbreviations, and units are correctly defined and used. Specific Comments No Technical corrections no.

The authors thank the referee for this positive statement on the MS.

Anonymous Referee Nr.5

<u>GENERAL COMMENTS:</u>

1*) The main problem of the paper is that both the training and validation of the algorithm seem to be a simple exercise and not an exhaustive study aimed at obtaining a robust cloud detection for OMI. The authors show some results for a limited dataset and as the abstract admits "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.". Since the use of neural networks for cloud detection has been already presented in previous works, a thorough development of the models and analysis of the results are required.*

For the revised MS a new data set has been used. 11 orbits have been randomly selected and divided in three datasets: 6 orbits for training the NN, 3 for testing, and 2 for independent validation. Also, the optimization of the neural networks (both backpropagation and extreme learning machine) has been optimized by minimizing the MSE/RMSE (as suggested by Reviewer 2). Validation on an independent subset has been performed and results are presented in the revised manuscript.

2) The authors apply SVD to reduce the number of reflectances at 751 wavelengths to a set of 20 values. They say that they tested several numbers of non-zero eigenvalues such as 5, 10 and 20, but they do not explain how they measure the different performances nor provide any classification accuracy results. The eigenvalues of the SVD decomposition explaining the variance of the transformed data should be shown, and the performance analysis used to select 20 dimensions has to be described and shown.

As suggested by the Reviewer #2 in general comment #1, the NN parameters (number of eigenvalues for SVD and hidden neurons) have been optimized by monitoring the MSE (for BP) and the RMSE (for ELM) over the testing subset as a function of the number of hidden neurons and eigenvalues. The combination giving the best performance with the test subset are the parameters applied to the NN for independent validation. This procedure for optimization of the NN parameters has been applied for land and ocean datasets. For more details, please refer to Reviewer #1, comment #3.

3) By using SVD for dimensionality reduction, one is preserving the maximum variance in the projected dataset, however some spectral channels with low variance (e.g. in some absorption bands) can be relevant for cloud detection. Other dimensionality reduction approaches such as partial least squares (PLS) should be explored and compared with SVD.
We thank the Referee for this suggestion. However, in the current work we have only used SVD. An exhaustive analysis of other data reduction methods is out of the scope of the current work, but in future work, the Authors will consider substituting the SVD with a single wavelength approach.

4) Again, for the selection of the model structure, the authors say that to determine the optimal number of nodes in the hidden layer, several tests were made. However, no information is given about the procedure followed to arrive to the 25 hidden nodes for the back-propagation (BP) algorithm, and the 240 nodes for the extreme learning machine (ELM) algorithm.
This comment is similar to that of Referee #1, comment #3. We refer to the extensive response to this comment.

5) The author say that "Data was first analyzed by observing each orbit individually, then by separating land- and water-covered pixels and discarding the ice-covered ones." Than means that they are intentionally excluding difficult ice/snow cases from the training, which explains the poor results obtained in highly reflective surfaces, such as ice or snow. In order to train a neural network performing well in most critical cloud detection cases, an effort must be done to include these cases in the training set.
We do not intentionally excluded difficult cases from the training dataset. The current training set includes six orbits but these had to be limited to +/- 70 degree latitudes because of light limitations. They do include deserts. More likely more training might be required to increase performances over bright surfaces.

6) Also, results are analyzed independently for land and water surfaces, but a better option is to develop different models for land and ocean due to their different characteristics.
In the revised MS, BP and ELM neural networks have been developed separately by minimizing a MSE or RMSE for land and ocean datasets. Once the models have been optimized, they have been applied to the validation subset.

7) The authors do not explain how many samples use for training the models nor how many samples per class they have in the training and validation set. In fact, they do not clearly describe any validation set nor the procedure they follow to validate the trained models (cross-validation, v-fold, ...?). All this information has to be clearly stated in the results section.
The dataset consists of 11 random orbits divided in the following subsets: 6 for training, 3 for testing, and 2 for independent validation. The same dataset was applied to BP and ELM. Both NN have been trained with the same datasets divided into ocean and land data. So altogether, if the Referee is asking how many pixels ere used in the training the answer is 263461 pixels over ocean and 228941 pixels over land have been used in the training.

8) Also, in the results section, the authors say that "The algorithm fails in distinguishing cloud-free pixels ... due to the lack of cloud-free pixels in the training dataset.". That is probably due to an unbalanced number of cloudy and cloud-free samples in the training set, which produces biased models. However, this is not an explanation for the poor results, it simply reveals a drawback in the generation of the training set that must be

solved before training the final models.
We assume the poor results in generalization are a consequence of poorly optimized NNs. We performed the optimization of NN parameters to avoid overfitting the NNs.

9) In the opinion of this reviewer, the training set has been generated almost randomly and with very few orbits. Therefore, comprehensive training and validation sets must be generated and all the results have to be repeated on a higher number of orbits to show the validity of the approach.
Please refer to previous replies regarding the new dataset.

10) Another aspect that the authors present as the main novelty of the paper is the comparison between back-propagation and extreme learning neural networks. However, the comparison is very limited since the authors only present the results of both methods. For example, the first layer of the extreme learning machines can be interpreted as a feature extraction. The analysis of the meaning of the SVD features and the first layer of ELM could be an interesting study of this paper.
The Authors appreciate the encouragement to investigate this interested topic. For the moment, we are now concentrating on the application of neural networks for cloud detection.

11) Also, the authors say that "results show that the ELM-based solution achieved higher cloud screening accuracy than the back-propagation algorithm" but in the opinion of this reviewer this conclusion is not justified by the presented results.
Correlation coefficients are reported in the results section. In Fig.4 and Fig.5 we present the density scatter plots and corresponding R^2 for ELM and BP NNs performances over land and ocean for the two validating orbits. While over ocean BP and ELM have similar results, the ELM performs better over land.

1*2) Finally, regarding the computation required by each neural network, the authors repeat during all the paper that the back-propagation based scheme is extremely time consuming in the training phase when compared to the ELM training approach. However, the training phase is performed only once ant the critical computing time is the testing time, which determines the time consumed for generating the cloud mask in a given orbit during the satellite operation.*
We agree. Text has been modified in the revised MS.

<u>MINOR COMMENTS:</u>

p1651, l17: "presence of clouds mask" -> "presence of clouds".
Text has been revised.

p1652, l7: "Here we use real data obtained from MODIS, with a spatial resolution which is much smaller than that of OMI." The sentence is misleading since the spatial resolution of MODIS is actually higher.
Corrected.

p1657, l16: Radiances are actually converted to TOA reflectance and hence some values can be higher than one.
Reflectance values are normalized so that they are values in the interval of [0,1] to assure to fully exploit the dynamical range of the input neurons (0-1).

p1668, l30: Preusker et al. (2008) reference cites a pdf file of a project report. More updated references from these authors presenting cloud screening algorithms based on neural networks can be found in international conferences.

Alternative references from these authors regarding neural network applications for cloud detection were not found. We added to the reference Preusker et al. (2008) the corresponding link to the website. If the Referee has better suggestion for this reference, the Authors will sincerely appreciate the contribution.

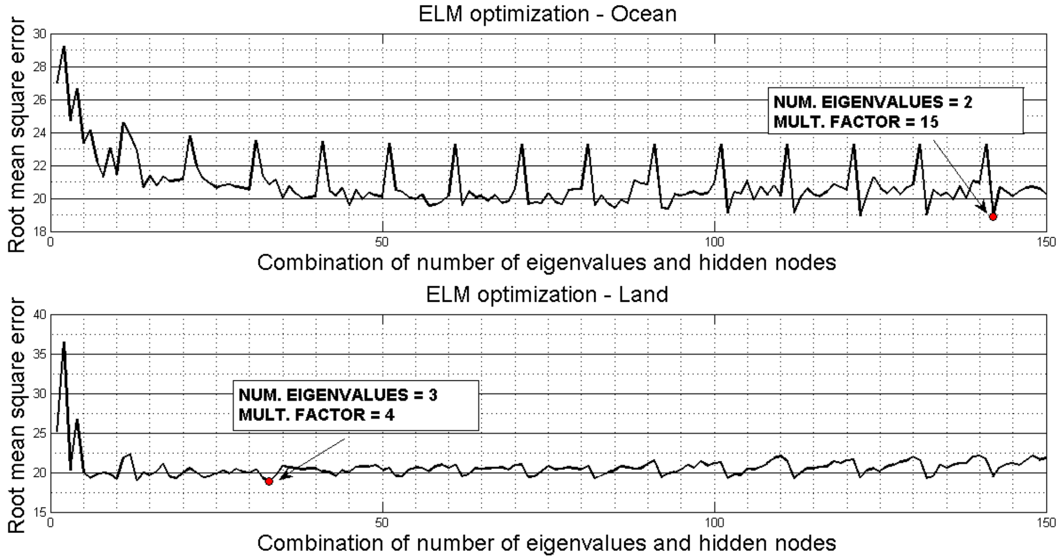

Figure 1: (TOP) Optimization of ELM NN for land data. (BOTTOM) Optimization of ELM NN for ocean data.

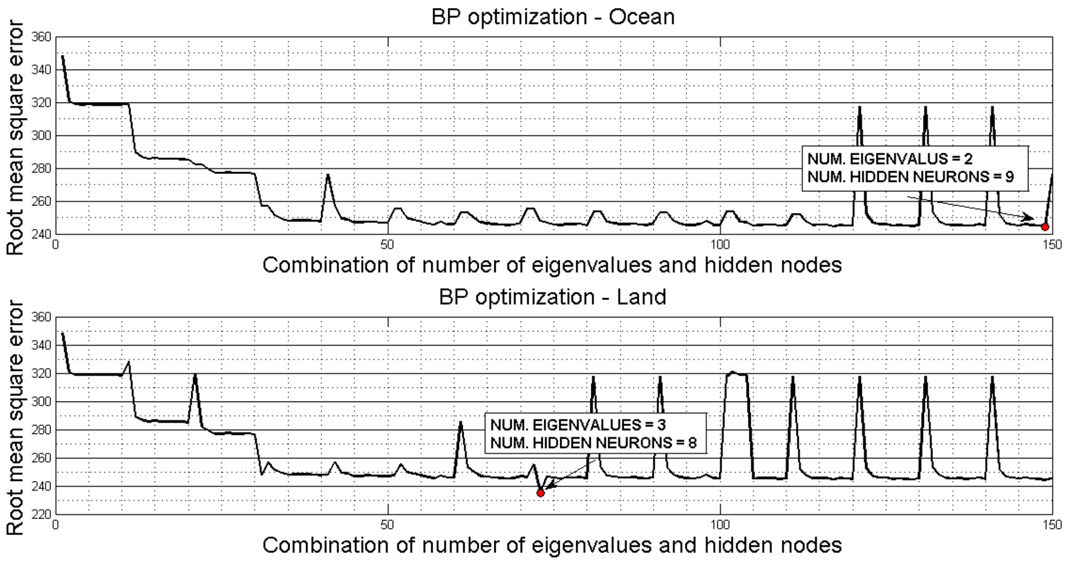

Figure 2 (TOP) Optimization of BP NN over ocean (BOTTOM) Optimization of BP NN for land data.