Atmos. Meas. Tech. Discuss., 6, C229–C239, 2013
www.atmos-meas-tech-discuss.net/6/C229/2013/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Development of a neural network model for cloud fraction detection using NASA-Aura OMI VIS radiance measurements" *by* G. Saponaro et al.

Anonymous Referee #2

Received and published: 10 March 2013

GENERAL COMMENTS

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. Besides this, and besides some other presentation issues that can be solved with relatively little effort (e.g. through a careful reading by a native English speaker), there are other serious problems in this work that prevent me from recommending its publication as it is now.

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.

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

AMTD
6, C229–C239, 2013

Interactive Comment

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?

DETAILED COMMENTS

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

2. P1650, L5. is -> are.

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).

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”.

5. P1650, L19. Are you sure that Andreae and Rosenfeld (2008) is an appropriate reference for this statement?

6. P1650, L19-21, sentence “for instance . . . pixels are discarded”. Could you provide a citation for this?

7. P1650, L24, sentence “The most consolidated methods . . . radiative transfer models”. Could you cite some examples of these methods as well?

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 (O_2-A band, O_2-O_2, 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.

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 O_2-O_2 band) with limited interferences from other factors except surface albedo.

10. P1651, L17. Remove “mask” at the end of the sentence. Furthermore, I do not

AMTD

6, C229–C239, 2013

Interactive Comment

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.

11. P1651, L23. Why do you place Sellitto et al. (2012) in “prediction of atmospheric parameters” rather than “ozone retrievals”?

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

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.

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

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.

16. P1653, L17. “Neural networks algorithms” -> “Neural network algorithms”.

17. P1654, L2. It could be good to cite Werbos (1974) after “MLP”.

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.

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

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.

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.

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.

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

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.

25. P1657, Eq. 5. Pi should be at the numerator, not at the denominator.

26. P1657, L14. Why did the Authors not use the viewing zenith angle and the relative azimuth angle as inputs?

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

28. P1659, L13. I would suggest removing "auxiliary".

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).

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".

31. P1659, L22. Remove "," between "pixel" and "as".

32. P1660, L3. "For the purpose ... training" -> "The training dataset for the NN

consists of . . .”.

33. P1660, L18-20. What is a “good performance in term of training accuracy or training time?”.

34. P1660, L25. How many training data have been produced in total?

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

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

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

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.

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

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?

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.”

42. P1666, L23. Bartlett (1998) is not referenced in the text.

AMTD

6, C229–C239, 2013

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

43. P1667, L1. Ortezi -> Ortenzi

44. P1667, L10. Roebling -> Roebeling

45. P1667, L30. Joiner and Bhartia (1995) is not referenced in the text.

46. P1668, L1. Vassilkov -> Vasilkov; Firts -> First.

47. P1668, L1, L5. Joiner and Vasilkov (2006), and Karayiannis and Venetsanopoulos (1993) are not referenced in the text.

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

49. P1668, L15. de Haan, J. D. -> de Haan, J. F.

50. P1668, L17. Platnic -> Platnick

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

REFERENCES

Acarreta, J. R., de Haan, J. F., and Stammes, P. (2004), “Cloud pressure retrieval using the O2-O2 absorption band at 477 nm”, J. Geophys. Res., 109, D05204. doi: 10.1029/2003JD003915

Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006), "Extreme learning machine: Theory and applications", Neurocomputing, 70, 489-501. doi: 10.1016/j.neucom.2005.12.126

Loyola, D. G., Thomas. W., Livschitz, Y., Ruppert, T., Albert, P., and Hollmann, R. (2007), “Cloud properties derived from GOME/ERS-2 backscatter data for trace gas retrieval”, IEEE Trans. Geosci. Remote Sens., 45 (9), 2747-2758. doi: 10.1109/TGRS.2007.901043

Loyola, D. G., Thomas, W., Spurr, R. and Mayer, B. (2010), “Global patterns in daytime cloud properties derived from GOME backscatter UV-VIS measurements”, Int. J. Rem. Sens., 31 (16), 4295-4318. doi: 10.1080/01431160903246741

AMTD

6, C229–C239, 2013

Interactive Comment

Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986), “Learning representations by back-propagating errors”, Nature, 323, 533-536. doi: 10.1038/323533a0

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

Werbos, P. J. (1974), “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences”, PhD thesis, Harvard University.

Interactive comment on Atmos. Meas. Tech. Discuss., 6, 1649, 2013.