

REVIEW: “A neural network approach for the simultaneous retrieval of volcanic ash parameters and SO₂ using MODIS data”, by A. Piscini, M. Picchiani, M. Chini, S. Corradini, L. Merucci, F. Del Frate, and S. Stramondo.

General Comments:

This paper presents a back-propagation method on a feed-forward multi-layer perceptron neural network, using data from MODIS satellite measurements to constrain mass of volcanic gas and ash in eruption-derived ash clouds produced in May 2010 by Eyjafjallajökull volcano in Iceland. The ash clouds were measured and simulated over sea surface alone and over clouds overlying a sea surface, and the results between the two situations are compared and analyzed. The results are very good, but the method raises questions regarding over-fitting and the use of prior distributions. Both concerns are related to the complexity of the neural network.

Specific comments:

1) Over-fitting:

a) The back propagation algorithm in complex neural networks is notorious for overfitting (an output highly tuned to a particular set of inputs), producing overly-confident results. That is why procedures such as cross-validation (which are used in the method the authors present here) were developed. As I am concerned about this, I spent some time analyzing the authors work and previous publications.

What concerns me most is the choice of data sets to train the network and the choice of data to test it. The network is trained using brightness temperature differences, or BT_D, for ash mass, effective radius, and aerosol optical depth, and training the network to estimate sulfur dioxide using the differences between simulated and observed radiances in the band around 8.7 microns. The test data set, or validation set, is bracketed by the data sets used to calibrate the model in time. As a consequence, this is an in-data-set or (IS) model versus an out-of-data-set or (OOS) model. Models tuned and tested with IS data do not test the veracity of a back-propagation algorithm, even in this method where care has been taken to limit over-fitting. The results appear to be very good, but are only compelling to me because the same method was used by Picchiani et al, 2011. What I would have liked to see was the use of a data set outside the time frame used to train the model, rather than having the model trained on data sets that bracket the validation set in time, or tests of the model using independent data from other instruments. For example, CALIPSO data exist for May 8 2010 for cloud temperature and emissivity (Pavolonis et al., 2013). These results could be compared with the weighted estimates of cloud temperature and emissivity determined from their training algorithm using MODIS data alone. Such a comparison would be a better measure of the tendency towards overfitting with their algorithm.

I urge the authors to more completely test their model, as they did using the same approach for ash detection and mass retrieval in Picchiani et al., 2011, and to use other measurements of ash mass obtained from different equipment to test their method, as in Corradini et al., 2010. As reported here, their model is trained to minimize differences with BT_D measurements of the same clouds observed just before and just after the validation data set, and can be expected to produce a very good fit to the clouds within this sequence.

b) In the optimal brain surgeon approach to obtaining a good balance between the data and the degrees of freedom in their model, they use a Hessian matrix. Is this Hessian determined from the output relative to the inner hidden layer output, or determined from the output of their neural network relative to the input to the neural network? The latter may be more appropriate since the results are used to prune the number of inputs.

c) The scatter plots suggest that using only 3 channels, as Picchiani et al., 2011 did, would provide a better Occam factor than using all 28 channels. This would apply both to gas as well as to ash estimates. The Occam factor can be estimated from the curvature of the posterior distribution of weights output from the network, evaluated at the maximum likelihood. Defining this matrix as \mathbf{A} , and Taylor-expanding the log posterior probability around the value of maximum likelihood for the weights, the posterior can be locally approximated as a Gaussian and with covariance matrix \mathbf{A}^{-1} . This approximation provides an additional criteria for pruning network inputs, and may provide a better comparison of results obtained with 3 channels versus 28 channels.

2) The use of prior data:

This concern is secondary to over-fitting, but may still affect the outcomes if informative priors are used. Placing a prior constraint on one parameter, such as water vapor, may influence the determination of other parameters in complex ways that affect the learning. A single layer network avoids this since the prior distributions are simple and are directly used to calculate output, but in a multilevel network prior distributions are developed at each level. If biases are excluded and unconstrained, then the priors across multiple levels form *improper* (unable to normalize) distributions. On page 10, line 1 it is stated that a 2 layer network with sigmoidal activation functions are independent from a priori assumptions. Unless the prior distributions are constructed properly, this is not strictly true. Sensitivity to prior distributions was a point raised by Picchiani et al., 2011.

Technical corrections:

Page 4:

Line 9: delete 'in'

Line 10: replace 'has' with 'have'

Line 12: replace 'have been also' with 'were'

Line 12: insert 'a scenario for' between 'to' and 'the detection'

Page 6:

Line 17: replace 'to' with 'of'

Page 9:

Line 2: replace 'insist' with 'exist'

Line 12: replace 'the achieved' with 'can be achieved'

Line 24: delete 'such as'

Line 25: replace 'as a' with 'that an'

Page 10:

Line 2: replace 'type' with 'types'

Line 6: replace 'a MLP' with 'an MLP'

Line 6: replace 'neuron' with 'a neuron'

Line 7: replace 'input,' with 'input'

Line 13: delete 'in the form'

Line 14: replace 'Digital Numbers (DNs), e.g.' with 'data such as'

Line 14: replace 'temperatures,' with 'temperatures'

Line 15: replace 'through a' with 'where the'

Line 15: replace 'equals' with 'is equal'

Line 22: replace 'vanish' with 'eliminate'

Line 29: delete 'of'

Page 11:

Line 2: replace 'aiming' with 'aimed'
Line 3: replace 'are,' with 'are'
Line 5: delete 'which are'
Line 7: delete 'the'
Line 20: replace 'is capable to give' with 'gives'
Line 21: delete 'its'

Page 12:

Line 10: replace 'the all' with 'all'
Line 11: insert 'from the' after 'parameters'
Line 11: delete 'have been used'
Line 16: delete 'an error step'
Line 16: replace 'The' with 'This'

Page 13:

Line 1: replace 'has' with 'have'
Lines 11 and 23: replace 'plume' with 'cloud'
Line 14: replace 'parameters' with 'parameter'
Line 23: replace 'sea and clouds' with 'sea and meteorological clouds'
Line 24: replace 'were lying' with 'were located'

Page 14:

Line 21: delete 'also'
Line 21: replace 'for the two' with 'show'
Line 21: insert 'that' between 'images' and 'have'
Line 22: replace 'in' with 'into'

Page 15:

Line 6: replace 'on' with 'at the'
Line 15: replace 'allowing to train the network' with 'allowing the network to train'
Line 28: replace 'As regard to the' with 'With regard to'

Page 16:

Line 13: replace 'of ' with 'for' in both cases.
Line 14: replace 'retrieval' with 'retrieval,'
Line 14: replace 'considered' with 'considered,'

Page 17:

Line 20: replace 'plume and clouds' with 'the ash cloud and meteorological clouds'

Page 27:

Line 1: replace 'divided' with 'in Table 2 and 3 by separately'

Page 28:

Line 1: replace 'for divided' with 'for parameters in Tables 2 and 3 divided'
Line 1: replace 'plume on' with 'ash clouds over'
Line 2: replace 'plume on' with 'ash clouds over meteorological'

Page 34:

Line 4: replace 'laying on sea' with 'clouds over sea alone'
Line 5: replace 'on' with 'and ash clouds over'

Pages 34-47:

The axes on all figures are labeled with miniscule labels that are unreadable without significant magnification. Please increase these font sizes to be readable.

References Cited:

Corradini, S., L. Merucci, A.J. Prata, and A. Piscini, 2010, Volcanic ash and SO₂ in the 2008 Kasatochi eruption: Retrievals comparison from different IR satellite sensors, JGR, doi: 10.1029/2009/JD013634.

Pavolonis, M.J., A.K. Heidinger, and Justin Sieglaff, Automated retrievals of volcanic ash and dust cloud properties from upwelling infrared measurements, JGR: Atmospheres, doi: 10.1002/jgrd.50173

Picchiani, M., M. Chini, S. Corradini, L. Merucci, P. Sellitto, F. Del Frate, and S. Stramondo, 2011, Volcanic ash detection and retrievals using MODIS data by means of neural networks, Atmos. Meas. Tech., doi: 10.5194/amt-4-2619-2011.