Atmospheric
Chemistry
and Physics
Discussions

Open Access

ACPD

Interactive
comment

Interactive comment on "A Neural Network Aerosol Typing Algorithm Based on Lidar Data" *by* Doina Nicolae et al.

Anonymous Referee #1

Received and published: 9 August 2018

The ACPD paper entitled A Neural Network Aerosol Typing Algorithm Based on Lidar Data by Nicolae et al. plays certainly a difference in the atmospheric aerosol typing research.

The proposed aerosol typing algorithm provides realistic range of results, as compared to values reported in literature in last two decades. The classification is done for pure aerosol, different component mixtures and/or predominant aerosol type. The selection of aerosol classes is scientifically relevant and detailed.

The authors made the developed code available, which should be stressed as a section related to Software Availability, after/before Acknowledgements.

I am somewhat missing discussion of the lower and upper range-limit of the typed

Printer-friendly version

Discussion paper

layers. How deep does it go into the boundary layer? What is the highest aerosol layer observed? Especially, for the validation exercise this would be relevant to know.

In Conclusions, I would expect more on the outlook.

The paper is written in an easy to read, casual manner. I made many specific comments that should be addressed to clarify the methodology. Suggestions for cosmetic changes in mathematical formulas are outlined. The Figures and Tables are in general good. I proposed some improvement. The captions of Figures and Tables are not acceptable, as they do not guide the reader.

The language seem in general fine for non-native speaker. In some sentences I suggest corrections and pointed typos.

I listed below many comment, all of which are actually only minor revisions. Therefore, I recommend this article for publishing, provided minor revisions are undertaken.

————————————

Minor revisions & comments

Throughout the text, please consider following changes: 1) remove coma after i.e. & e.g. 2) change 'ration' to 'ratio' 3) change 'humidity' to 'relative humidity' 4) change 'backscatter' to 'backscattering' 5) check color ratio vs color index - these are mixed up and the latter is not defined 6) fonts are sometimes mixed up (on purpose?) in text, tables and equations, what looks strange 7) make a decision to use the names of aerosol types either with small-latter or with Capital letter, and keep it consistent.

p.1 l.3 lidar-based techniques

l.4 remove 'depicted'

l.4 change to 'One of such techniques in based'

l.6 change to 'multispectral lidar data'

l.8 change to 'parameters, calculated'

l.10 change to '(or not)'

l.11 change to 'identification of 14'

l.13 change to 'the most probable aerosol type'

l.14 change to 'The limitation of NATALI is'

l.15-16 change to 'Additional applications of NATALI, e.g. to test'

l.16 change to 'or insufficient cloud screening is feasible.'

l.17 change to capability of NATALI to retrive

p.2 l.1 change to' (Cavalo et al., 2013, and references therein)'

l.6 change to 'IPCC reporting'

l.7 change to 'sources should reduce'

l.15 change to 'aerosol type,'

l.16 change to 'However, the measurements averaged ... column cannot'

l.18-20 move references as: 'aerosol types by providing ... properties (Muller ...), as well as to understand'

l.24 change to '(e.g. David...), thus difficult'

l.25 change to 'helped to address'

l.28 remove 'collected'

l.31 spectral particle depolarization ?

p.3 l.1 change to ',such as Saharan' and later on 'large areas and long distances'

l.2 references - add also more recent demonstration of the network capability, e.g At-

mos. Chem. Phys., 17, 5931-5946, https://doi.org/10.5194/acp-17-5931-2017, 2017

l.10 change to 'Another issue in aerosol'

l.12 change to 'data of pure aerosol types are sparse.'

l.17-19 for consistency, move ref.Tesche after 'Marocco,' and ref.Gross15 after 'interaction Experiment,'

l.18 remove 'n' and add coma before ref.Gross11

l.20-21 clarify: ... of pure types (e.g. dust and their mixtures) <- so pure or mixtures?

l.24 change to 'aerosol properties to be'

l.25 an aim? and later remove ', for example,'

l.27 consider changing 'things and situations' to 'data'

l.30 change to 'The ANN contain tens'

p.4 l.4-5 change to 'In this paper, an in-house-developed ANN algorithm for aerosol typing is introduced.'

l.9 change to 'are used at first to compute'

l.10 start in new line: The ability...

l.12 coma before 'including'

l.13 change to 'of different aerosols was' later on 'capable of reproducing the observed aerosol properties, and'

l.15 change to 'needed doe to sparse observational data sets'

l.16 change to 'The aerosol model was constructed to'

l.27 remove 'in this article'

l.31 chemical composition <- give hint on which parameters

l.32 certain limits <- specify in brackets what limits you mean

l.32 remove 'as much as possible'

p.5.1.4 remove 'average'

l.14-15 change to 'In the aerosol model, particles ... ratios to simulate the aerosol anisotropy'

l. 17 microscopical properties <- give hint on which properties

l.18 change to 'in certain limits the critical component (in brackets give hint what you mean!)'

l.24 change to 'For each'

p.6 l.1 change to 'for the j components of the aerosol mixture, respectively.'

l.4 change to 'The size ... is mono-modal, log-normal and given by ..... where sigma_p, sigma_j sigma is ... and r is ... ' (note: in Eg.5 is r not r_j ?)

l.13 change to 'Use of the calculated microphysical properties with the T-matrix code (provide reference), the effective cross-section for the particle scattering C_sca and extinction C_ext, as well as ...'

l.14-15 and l.17-18 change to '... alpha is determined from Eq.6 and ... beta from Eq. 7, where F11 is...'

l.7-11 move to the end of l.20, then correct it as follows: The integration domain (R_min:R_max), for which the effective radius r?, r_p?, the extinction coefficient alpha and scattering coefficient sigma? sigma_p? are calculated, covers medium size particles with radius between ... The radius was not increased further due to computing time limitation and model design limitations (i.e. ...). However, the latter limitation is not considered critical for the range of the lidar wavelengths.

C5

l.10 correct and start in new line: 'The single scattering albedo omega is yield as the ratio of the ....

p.7 l.5 change to 'The algorithm is iterated for each'

l.6 it is confusing -> what domain you mean here?

l.11 in brackets change to '(above 90%)'

l.10 change to 'considered as too rarely present in the ambient atmosphere.'

l.12 change to 'Thus, within ... is considered linear for '

l.15 rather: 'This is partly due to the limitations of the model itself and partly due to the various uncertainties associated to the measurement ...'

l.16 explain what you mean by linearity here.

l.17 change to 'applied in pre-processing to correct' later on change to 'as close'

l.17-21 this is not very clear, do you mean that 'To reproduce lidar measurement, for each simulated parameter an associated relative error (deltax) is assumed and the possible variable range is then computed as [x_med-deltax:x_med+deltax]', were x_med is the mean? median? value within the selected layer boundaries? But then why this is applied to syntetic data and true measurement? So you add additional error to the true data? what you mean by 'certain step' in finesse? is the deltax relative or absolute error and finally which is assumed, both?

l.23 start in new line: Based on the values ...

l.25 rather it should be: 'extinction coefficient derived with Raman method, mainly due to noisy Raman lidar signals.'

l.24,27, be specific, in brackets give uncertainties of ext, bscat and depol that you mean!

l.27-28 change to 'in the case'

l.28 change to 'Lidars, were the'

l.29 change to 'cross-talk... is still'

l.31 change to 'inclusive to mimic ...parameters.'

p.8 l. 2 add come before 'as long as'

l.5-6 remove sentence 'There was.. mixtures' and then change to 'The mixtures composed of only two pure types were considered not sufficient.'

l.10-11 change to 'of possible mixtures ... three pure aerosols (35 mixtures), only ... (9 mixtures, see Table 4).'

l.12 you mean: the time-performance of the algorithm and the minimum number of output aerosol types considered significant in atmosphere.

l.13-14 correct and change to 'the extinction Angstrom exponent (k_ext) and the backscattering Angstrom exponent (k_bsca), also referred to as color ratio'

l.21 titla of 2.2: The ANN architecture ?

l.22-26 belongs to Introduction! -> start the section with: The ANNs can be calibrated or...

l.2 change to 'elements of identical parameters (e.g. ...) with an associated ...'

l.4 start as new paragraph: No significant ...

l.5-6 unclear sentence: 'The TanhAxon ... -1 and 1.' you mean: a bias-offset B ? each neuron x_i ? in the layer omega_i ?

l.9 remove 'all'

l.9 start in the new line with: Supervised ...

l.13 change to ' training approach... ANN: the input... and the output compared with the [specify: true? or initially assumed? or pure?] aerosol type, in order...'

l.15 specify: the minimum classification errors ?

l.17 clarify: the standard gradient was not tested at all?

l.18 change to 'It provides'

l.21 change to The ConjugateGradient ... has no parameters ... (such as learning ...) and'

l.22 change to 'algorithm'

l.24 change to 'function, were found inadequate for aerosol typing purpose.'

l.25 change to 'but its active'

l.26 start as new paragraph: The cross-validation ...

l.27 please make it more specific here what you mean by 'when error begins to increase'? is it at the beginning zero? or has some initial constant value? how is it defined? assumed?

l.29 be more specific, what you mean by the exact type?

l.30-31 change to In total 68 ANN structures ... in order to compromise between the minimum ...

l.31 specify what is meant here by saturation effects?

l. 32-33 rather change to 'Examples of 6 pure, 7 double-component mixtures and 2 triple-component mixtures obtained within the 68 explored ANN ...

p.10 l.2-5 state clearly if for the training, tasting and validation the same classification error <25% is used.

l.7-8 should this not read 'the Jordan-Elman with 6 or 8 hidden laers and the Generalized Feedforward with 10 hidden layers' ?

l.9 what you mean by response with higher confidence? the classification error?

C8

l.10 define the bast weights

l.11 the title of 2.3 The typing algorithm ?

l.12 should not be 'the uncertainty threshold'?

l.12-13 change to 'parameter, a bundle of imputs ... was generated.'

l.14 remove: 'then'

l.15 remove: 'and'

l.16 start at new line: For the NATALI aerosol typing algorithm, the input ...

l.17 optionally?

l.18 remove: all

l.21 good quality <- specify in brackets the uncertainty threshold

l.24 high uncertainty <- specify in brackets the uncertainty threshold range

l.25 change to 'available, the'

l.26 change to 'meaning again that the'

l.27 change to 'retrieved for 5 outputs (...), whereby if... from Dust nor Continental ..., and therefore is ...'

l.30-31 change to 'The three ANNs (Table 3) were developed ...schemes (Table 4) to increase the confidence of the aerosol typing. A voting procedure selects the most...

p.11 l.1 the title of 2.4 The NATALI Code ?

l.2 change to 'The ... developed in the Phyton programing environment is buitl on ...

l.6-7 change to 'calculates within each layer the mean intensive optical parameters (name them in brackets) with uncertainties (Fig.3).'

l.8 should be: The layer boundaries

l.8 please clarify what is used boundary layer detection: the gradient (1st derivative of signal) or as in l.9 the inflection point (2nd derivative of signal)

l.8 is the range&time resolution of 1064 signal the same as the resolution of the optical EARLINET profiles?

l.9 specify what parameters for SG-filter are used, are they applied the same for all profiles?

l.11-14 From own experience, layers of thickness < 300m can have higher SNR than layers > 300m, as SNR depends on aerosol optical depth, which even for very thin layers can also be very high! I suggest to revise the fragment as follows: The layers with thickness of more than 300 m are considered, whereby the intensive optical properties and their uncertainties are computed for the middle of each layer in the range of at least 200m thickness, to exclude the margins likely affected by the smoothing.

l.15-16 The first 2 sentences are repetitions, pls remove it, and start directly with: Several filters ... pass the following criteria ...

l.22 change to 'the input model generates an adjustable number N of values x with uncertainties deltax in the range of [x-deltax:x_deltax]'.

l.24 start in new line: The typing module ...

l.25 change to 'In the case that the depolarization ratio is'

l.28 remove 'some'

l.29-32 change to ', and returns only the ... type (Volcanic overlaps completely (...) ... polluted type and cannot be retrieved in low-resolution.'

l.32 remove as irrelevant sentence: Thus...

l.33 change to 'telegrams contain the identification of the data sets for which typing is

C10

performedand provides for each identified layer following parameters:'

p.12 l.3-7 start each parameter with 'the'

l.10 change to 'The NATALI code'

l.10-12 check the names of the python routines there is problem with the font.

l.12 should be: The three

l.13 rather module routines/codes than methods!

l.15-18 should be Firstly, ... Secondly, ... Thirdly, ...

l.22 change to 'by the developed aerosol model'

l.25change to 'As shown in'

l.26 move the '(i.e. ...) to the end of the sentence.

l.27-29 could you comment on what could be the reason for the given discrepancies?

p.13 l.3 OPAC hygroscopicity was NOT available?

l.3-4 change to 'However, the changes in OPAC are not expected to produce ...'

l.6-8 chnage to 'In the Figure 4 comparisons ... by Gross et al. (2013) are provided. =

l.8-13 Based on the Airborne... properties measured during campaigns in ... EU-CAARI), Gross et al. (2013) developed an aerosol classification scheme for six aerosol types ...'

l.15 should be: observations at 532 nm from ... (Fig.4 a and d), especially for smoke...

l.17 should be: therefore were not compared.

l.18-20 should be: Clusters are identified both in synthetic and observational data , which means ... one wavelength can be sufficient for the ANN training.

l.18-20 I am interested if you did actually try to train the ANNs with the single wavelength extinction, backscattering and depolarization?

l.23 either add after 'measurement campaigns' comment in brackets: (listed in caption of Fig.5)' or list the campaigns in the text and leave them out in the Fig.5 caption, what will be consistent with Fig.4

l.29 it is difficult to see overlapping of smoke and continental polluted

l.29-31 I would revise this fragment as follows: Smoke and continental pollution almost completely overlap (Fig.6a), which is consisted with measurements reported in literature as in Table 6. This makes the typing challenging. The importance of particle depolarization shown relatively recently (e.g. =Freudenthaler et al., 2019), can improve the aerosol typing (Fig.6b).

l.33 should be: calibration, recently addressed by e.g. McCullough ...

l.34 should be: the depolarization ratio quality criteria.

l.34-35 rephrase this sentence to positive: However, even without particle depolarization information, the low-resolution typing can identify the aerosol predominant types in a mixture.

p.14 l.9 change 'Also,' to '; continental ...'

l.10 change 'while' to '.Note that 33% ...'

l.12 add at the end of 1st sentence: (Fig.7b).

l.16 start in new line: A3H and A3L ...

l.16 most or best performing ?

l.18-19 last sentence -> pls clarify, I do not see this from provided Figs!

l.21 change 'for example,' to 'olny'

l.21 I think there should be dust continental instead of marine mineral for high-resolution typing

l.22 I think there should be continental smoke and mixed smoke for low-resolution typing

l.23 consider adding sentence: For costal polluted, the relative humidity increase results in an increase of typing performance.

l.25 I suggest to separate the CALIPSO and INOE data sets, therefore I propose to rewrite this as: Observational data from EARLINET Data Base (www.earlinet.org), related to the CALIPSO (...) overpasses over different EARLINET observational sites, were compared with ... model. The EARLINET-CALIPSO Data base (Pappalardo et al. 2010), covers the data of 2000-2018 ??? and includes...

l.29 should be: all of the

l.31-32 change to: '... was added, assuming values reported in literature as typical for the corresponding aerosol type. This way, 105 cases ... were obtained.'

p.15 l.1-2 change to: 'The cases for which all parameters were within 20% of relative error were selected (63 cases), whereby 57 corresponded to known aerosol types.'

l.3 add sentence: 'Additionally, profiles available at the EARLINET site in Bucharest/Magurele, established by the Romanian National Institute for Research and Development of Optoelectronics (INOE), were used to increase the validation measurements sample.'

l.4-5 change to: 'About 44.6% of measurements were conducted at nighttime (thus, include the Rama-derived extinction coefficient profiles).

l.5 871 profiles or layers?

l.6 remove: profiles

l.6 instead of maximum of 1h, give range of temporal averaging (e.g. 30min-1h)

l.6 significant aerosol load ?

l.7 change 'could be' to 'were' later on should be 'higher than 20%'

l.10-11 in fect this are 311 and 182 layers?

l.11 start last sentence is new line as: The time-series of lidar measurements (532 nm vol.depol.ratio and ... range corrected signals) were used to identify the aerosol layers.

l.13 start last sentence in new line : The identification...

l.14-16 change to: 'source was assumed at the region ... was closest to the ground, providing guidance ... The rainfall along trajectory was used a indicator for likely wet deposition.'

l.16 meteo parameters or fields?

l.17-18 should be: was based on ... reanalyses (Kalnay et al., 1996) to confirm the aerosol trajectories ...

l.21-23 be specific: the mixtures of three components, incorrect calibration or inappropriate estimation of aerosol type. On the other hand the sparse observational data lead to ...

l.25 make new sentence: Low values ... lidar ratio are seen for several cases of ...

l.26 should be: This is most likely ...

l.27 coma before 'and thus'

l.29 change 'could not' to 'failed to'

l.29 consider adding information: Actually, in general the agreement of classification of the simulated data and the real observations is very good, given all limitations discussed.

p.16 l.2-4 rather unclear, especially the prerequisite hypothesis

l.6 strong or crucial?

l.7 should be: the depolarization methodology is still maturing and only a few lidar stations ...

l.15 should be: for the input data gave realistic retrieval of aerosol ...

l.21 I reckon you should repeat very valuable information in the section: Software Availability next to the Acknowledgements!

l.22 easy to use or: The NATALI software is user-friendly; a user-guide is provided.

l.23 should be: the limitations of the algorithm, i.e. the results are ... on the quality of the input data

l.27 should be: of less that

l.28 rather: of the cases for medium and high

l.33 rather: for example, it can be applied to test ... data, identify ...

l.33 could you please comment whether you did actually tested this specific QC-procedure? is there any conference paper on that you could cite?

p.17 l.2-3 comment on expected results for data sets of more than 3beta, 2 alpha and 1sigma, e.g. adding polarization at 355 and/or 1064 nm; adding lidar-derived RH profiles? How you expect this improves the results?

—-

References

p.21 l.17 should be Janicka; p.25 l.10 should be Muller, D.; p.25 l.13 remove 2x .and

—-

Equations

Eg.9 more elegant formulation: LR=4pi/omega*F11(180)

Eg. 13 please do not use beta and omega in this equation, as they both have already associated physical quantities! the current beta-bias-offset can be denoted as B. What is the omega here? as there is x_iˆlin is this not LinearThanAxon?

—-

Figures

Fig.1&2&3 improve captions, specifically add info that schemes are used for the NATALI code.

Fig.3 I suggest to change the description to: Mean intensive optical parameters within the layer, Mean uncertainty of intensive optical parameters within the layer, Scrambled seta of mean intensive optical parameters +- uncertainty

Fig.4 indicate in upper sub-figs. what wavelengths they depict; for consistency in legend of lower sub-figs give aerosol types in sequence: dust, dust polluted, smoke, continental polluted, marine

Fig.5 indicate for which wavelength is the upper figure! for comparisons the same wavelengths of the observations and synthetic data must be used. Caption is lengthy, the description of the campaigns should go to the text.

Figs.6-9 for consistency, order the aerosol types as in AH definition in Table 4 and use THE SAME COLOR SCALE for each aerosol type.

Fig.6 upper figure should be CRvsLR not ARvsDR! In caption add info that this is NATALI classification

Fig.7&8 change Marine/CC to marine! Plot confidence-levels? vote-number? in grayscale (white being 0-20%, black 80-100%). Note that low-resolution typing for 7 pure

categories is excellent. Note that high and low resolution results are better of any info on RH is available. Give more explanation on how to understand figs in the captions. Again, add info that this is for NATALI code.

Fig.9 Add info that INOE site is also EARLINET site. Add over the left column info synthetic-data and over the right one observational-data

—-

Tables

Table 1 change big to large; change right column title to Particle characteristics, for marine -> should not be non-absorbing? continental polluted and smoke are identical but smoke can be aspherical.

Table 2 what is in the 3th column, whet is this critical component proportion? is has no units? you mean axis-ratio or aspect-ratio? Is this table defined by results reported in literature?

Table 3 check font of than-axons

Table 4 specify whether aerosol types are retrieved directly by the NATALI or measurements? Note there is no italic font! why categories in Arial-font are given in low-resolution if they are not really retrieved?

Table 5 AE and CR have the same values? is the AE, CR of 6 realistic? the Rayleigy being 4 should not be the limit? Which CI is correct? Is the lower CI not LR? Please check carefully this table for typos!

Table 6 consider adding reference for Continental polluted/industrial: Raman lidar-derived AE550/350 of 1.17-1.19 is reported in Remote Sensing 2017, 9(11), 1199; doi:10.3390/rs9111199.

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2018-492, 2018.