Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-64-RC1, 2017

Atmospheric Measurement Techniques Discussions

Interactive comment on "Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks" *by* Johan Strandgren et al.

Anonymous Referee #1

Received and published: 24 April 2017

The authors present a new retrieval method retrieving three cirrus-related climate variables from SEVIRI. The science appears solid and is mostly well-described. I recommend publication with minor revisions. Below I make a few general comments, followed by detailed comments line by line, which are mostly requests for clarification.

General comments

The authors characterise the performance of CiPS in several ways. Did you look at its performance depending on underlying surface type? Snow surfaces are famously difficult, but other surface aspects may be relevant as well.

The elephant in the room in many retrieval products, in particular those based on ma-

chine learning, is the uncertainty. Although the the paper provides characteristics on overall performance, is there any way to get an uncertainty estimate for a specific retrieval?

Is there, or is there planned to be, a publicly available data product based on CiPS, so that people can download the data and explore it on their own? I think there should be people interested in using it.

Could the approach be extended to other imagers than SEVIRI, as long as those have an overlap with CALIOP to be trained with? Or are the properties of SEVIRI (footprint size? scan speed? channels?) essential for CiPS to work?

Specific comments

Abstract

- Page 1, lines 1–2: replace "one of the largest uncertainties" by "one of the largest sources of uncertainty", and replace "they" by "their physical properties"
- Page 1, line 8: after 71, add %

1. Introduction

- Page 3, line 29: add "piece of" before "information"
- Page 3, line 32: please explain acronym LES (I guess this is large eddy simulation in the context of a cloud resolving model), which is missing in the text and in Appendix A. As this acronym appears to be used only once in the paper, I suggest just writing it out and avoiding the acronym altogether.

2.1. SEVIRI

Please expand this paragraph with:

- longitude above which SEVIRI is located (finally described on page 12, line 22)
- total range of field of view of the disk SEVIRI can observe

2.2 CALIOP

- Page 4, line 18: This usage of the word “frequency” is potentially confusing, maybe write that it measures 20.16 times per second or every 49.6 ms (when I see the word frequency I think of electromagnetic frequency).
- Page 4, line 24: please explain acronym IOT
- Page 4, line 26: This line has some typesetting issues: CPL should be explained at first use, and the formatting of the citation is incorrect.

2.3.2 Learning through backpropagation

- Page 6, line 16: replace “a” by “an”.

2.4 Validation metrics

- Page 8, equation 5: this MPE metric is risky because over- and underestimates can cancel out each other. I realise that is why the authors also look at MAPE but I think this risk should be explicitly pointed out.

3 CiPS

- Page 9, line 4: remove “though”
- Page 9, line 5: remove “the”

3.1 Multiple artificial neural networks

- Page 9, line 10: Remove “decimal” before “number”. I don’t think the authors actually mean a decimal number as defined by IEEE 754-2008, presumably it’s a regular binary in their software implementation.
- Page 9, line 23-24: Does CALIOP (reliably) identify when it is saturated?
- Page 10, line 2: The authors refer to “photon counts” but I don’t expect SEVIRI actually counts photons. The digital count level is probably rather a conversion from a voltage. I assume the authors use brightness temperatures already calibrated elsewhere, so I suggest to cite the relevant paper or technical report if available.

3.2.1 Brightness temperatures from SEVIRI

- Page 10, line 8: It would be useful to remind the reader to what surface area 19×19 pixels^2 corresponds (57×57 km^2?)

3.3 Output data: cirrus properties from CALIOP

- Page 12, line 3: Which spatial resolution do the authors use, finally?
- Page 12, line 10: Please add a bit more information about thin

`Opacity_Flag` product. How is this determined and how reliable is it? I understand that multiple profiles are combined. How is this done for the opacity flag?

3.4.1 Data collocation

- Page 12, line 25: "For this time period", referring to Sect. 3.3, but actually the time period is described in Sect. 3.4.
- Page 12, line 31: I'm confused. Higher on the same page the authors discuss how there are spatial resolutions at 5 km, 20 km, and 80 km. But now they seem to consider only 5 km. Then what is the relevance of the other spatial resolutions?
- Page 13, line 2: I believe the re-analysis also contains forecast variables at every hour, why not use those instead of interpolating the 6-hour time steps? Depending on what local time those correspond to a linear interpolation for surface temperature could introduce significant errors.

3.4.2 Training data

- Page 13, line 5: replace "millions" by "million"
- Page 13, line 11: do the authors use transparent as a synonym for "CALIOP signal did not get saturated"?
- Page 13, lines 15-19: the authors a huge training dataset, many orders of magnitude larger than in many other machine learning cases. Instead of duplicating certain cases, have the authors considered thinning the part of the state space where there are many cases, perhaps in a way similar to Chevallier (2016) https://nwpsaf.eu/downloads/profiles/profiles_91L.pdf ? That may provide a less biased dataset and (much) faster ANN training. Note that you are actually doing

this on page 15, line 10; first duplicating some points by a factor 4 and then using only 25% of the points is essentially thinning, depending on how the sub-selection is performed.

3.4.3 Validation data

- Page 13, line 28: what are the consequences of applying this balancing (or alternatively, as I propose above, thinning) to the training data but not to the internal validation data? This would mean that the statistical properties of the internal validation data differ from the ones for the training data. Can this introduce biased results?

3.5 Training

- Page 14, line 3: move "described in Sect. 2.3.2" to after "mini-batch gradient descent", because both backpropagation and mini-batch gradient descent are described there.

3.5.1 Training meta-parameters

- Page 14, Figure 2, legend: Replace "Tranparent" by "Transparent"
- Page 14, line 12: How is this random search performed? This is an optimisation problem and there are different ways of finding local or global minima.
- Page 14, lines 14-15: How is "best performing" defined? You have multiple metrics but it's not clear how those have been used exactly.

3.5.2 MLP structure optimisation

- Page 15, line 10: see my comment at page 13, lines 15–19
- Page 15, lines 18-19: Do you mean the differences between structures are very small, and/or the differences among the two trained for each structure?
- Page 31, Figure 3: It is hard to tell the differences between the performances. Could the authors add a figure showing the actual improvement (in %-point) between the network 3-64 and 1-16 and/or between the finally selected network and 1-16?

4.1 Application

- Page 17, line 16: replace "12.30" by "12:30"
- Page 32, figures 4(d)–(f): The colourmap chosen by the authors may not be optimal. As explained by Borland and Taylor (2007), the rainbow colour map and other colourmaps that are not perceptually uniform may be deceptive in their visual interpretation. The authors may wish to study the data using a perceptually uniform colourmap. Secondly, I do not understand why the colourmap in Figure 4(d) is different (opposite?) to the ones in 4(e) and 4(f). In this case, white has been used to indicate areas without cirrus clouds, so the colourmap should ideally not contain a colour similar to white (perhaps possibly at the low end of the IWP and IOT scales)

  Borland and Taylor (2007), Rainbow Color Map (Still) Considered Harmful, in: IEEE Computer Graphics and Applications (Volume: 27, Issue: 2, March-April 2007), doi:10.1109/MCG.2007.323435
- Page 19, line 4: Replace "along side" by "alongside"

4.2.2 Cirrus properties

- Page 35, Figure 8: you are validating against CALIOP so the CALIOP measurement should be on the x-axis instead of on the y-axis. The same comment applies to Figures 10 and 11.
- Page 35, Figure 8: Why does the CALIOP scale go down to 4.0 km if the dataset excludes data with CTH < 4.5 km (poles) or 9.5 km (tropics)?
- Page 35, Figure 8, right panel: most of the lower part of the panel is empty. I think you can restrict the y-axis to -20% or so, and abandon the symmetry on both sides of the y=0-line. The same applies to Figures 10 and 11.
- Page 20, line 4: You might want to again point out here that the 4.5 km CTH in the dataset are all near the poles, so the problem for these pixels is actually more difficult than for others.
- Page 20, line 17 / Page 35, Figure 9: could you add a panel to Figure 9 showing the density of points that make up the statistics shown in Figure 9? You write in the text that those cases where there is a bias are relatively rare. Such a 2D histogram could show how rare.
- Page 36, Figure 10: comment at Figure 35, Figure 8 applies
- Page 36, Figure 10: caption should describe what the shaded area in the middle panel indicates. Currently this is only stated in the main text.
- Page 20, line 28: How is this (lack of correlation when $\mathrm{IOT}_{\mathrm{CALIOP}} < 0.04$) apparent from Figure 10?
- Page 21, lines 19–21: $\mathrm{IOT}_{\mathrm{CALIOP}}$ is your reference for the training and the validation. Why would a bias between $\mathrm{IOT}_{\mathrm{CALIOP}}$ w.r.t. truth contribute to your error relative to $\mathrm{IOT}_{\mathrm{CALIOP}}$?

- Page 36, Figure 11: comment at Figure 35, Figure 8 applies
- Page 22, line 4: I believe this result is not shown, so the authors may wish to indicate this for clarify (i.e. "(not shown)").

5 The cirrus life cycle with CiPS

- Page 22, line 12: I believe you mean Pyrenees, not Alps, or I'm confused.
- Page 22, line 15–20: Is this method an established technique or something that the authors developed? If the former, can you add a reference to a source containing more details? I realise it is not the main focus of the study but it would seem something the interested reader may wish to learn more details about.
- Page 22, line 33: replace "does also present" by "also presents".
- Page 23, line 5: Your cloud has its maximum area exactly at the time from which you tracked it forward or backward in time. Is this just a coincidence, or could it mean that your tracking is not entirely reliable? I'm a bit worried that the analysis in this paragraph may say more about your tracking method than about the cloud evolution, in particular for the surface area (not saying this is the case, but in theory it could be and therefore should be addressed or ruled out as an alternative explanation).
- Page 23, line 20: remove "though"

6 Conclusions

- Page 24, lines 9–10: there is a double negative here ("cannot ... neither ... nor"). I suggest to replace "neither ... nor" by "either ... or".

- Page 24, lines 11–14: the authors might want to briefly repeat the main points of why/how CiPS improves upon COCS.

Needs some lines on recommended future work / next steps.

Acknowledgements

Please expand/explain the acronyms in the acknowledgements, where they were not explained before (DLR, DAAD).

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-64, 2017.