8th Mar 2024

Reviewer's comments

General comments

The preprint entitled *"`ampycloud`: an algorithm to characterize cloud layers above aerodromes using ceilometer measurements"* has been reviewed for possible publication in Atmos. Meas. Tech. The authors developed a new algorithm and associated Python package claiming it is suitable for the fully automated reporting of cloud information relevant for Meteorological Aerodrome Report (METAR) at Swiss civil aerodromes. The open source strategy with the Python package and its online documentation are positive, but the potential user group is likely limited to other weather services.

The manuscript is clearly organised, although individual paragraphs and figures belong in other sections or in the appendix. The diagnostic diagrams are really nice, although they do require a certain amount of background knowledge. The writing style is well understandable, even if some issues could have been explained better with more appropriate wording. This would also avoid unnecessary repetition of the same content both in the main sections and in the figure captions. Spelling and grammar are appropriate but need improvement and British English should be used consequently. Occasionally, the authors become entangled in contradictions and misunderstandings are inevitable for less knowledgeable readers. In a few places there is no corresponding reference, e.g. to the CL31 manual from *Vaisala* (2009).

Essentially, it is a matter of assigning the cloud hits originating from *n* ceilometers, which are derived from the actually measured backscatter signal, to cloud clusters, summarising these as far as possible into cloud layers and finally weighing up which of these are relevant for reporting, considering all the requirements and rules of the *ICAO* (2018). The authors use interesting approaches applying agglomerative clustering method and Gaussian mixture model, but it seems like they take a sledgehammer to crack a nut. In contrast, some key requirements of the *ICAO* (2018) are not met like reporting "NSC" (nil significant cloud) or vertical visibility (VV) with altitude value (see specific comments).

The quality of any similar cloud algorithm depends strongly on the performance of the ceilometer type used. Various ceilometers from the same or different manufacturers derive from their attenuated backscatter profile different cloud and VV hits at a given time and each device has its own strengths and weaknesses (e.g. *Illingworth et al.*, 2019). `ampycloud` is designed and tuned for CL31 and needs to be adapted or even extended if you replace the ceilometer. In the specific comments, I try to explain the limitations of the CL31 with regard to its use in aviation based on personal experience.

From the results presented, I cannot recognise any superiority of `ampycloud` over existing similar cloud algorithms, except that the implementation here is significantly more complex. Furthermore, its assessment based on observations is not convincing. Contrary to what the authors call an improvement of the algorithm, it would be an actual innovation with potential to be published, if a) the integration time has been made dependent on the horizontal wind speed or b) the measured backscatter signal has been used to derive cloud hits and VV hits by yourself to get rid of the manufacturer black box algorithms. In my opinion, this manuscript is not suitable for publication in its current form.

Specific comments

My recommendation for the title is:

> *"`ampycloud`: an algorithm for determining the amount and base altitude of cloud layers over aerodromes using ceilometer data"*

On the one hand, there are numerous macro- and microphysical cloud properties (e.g. *Gao et al.*, 2014; *Luebke et al.*, 2022), but `ampycloud` is only able to determine cloud amount and cloud base altitude based on the data of *n* ceilometers. On the other hand, the quantity actually measured by ceilometers is the backscatter signal (attenuated backscatter), while the cloud or VV hits are derived by application of manufacturer-specific algorithms due to the lack of a quantitative definition of a cloud (section 2.1.1 of *WMO*, 2017).

`ampycloud` is a strongly scientific rather than application-orientated implementation. There is no need to determine cloud amount and cloud base altitude with best accuracy, considering that for the METAR/SPECI the former is classified into FEW, SCT, BKN, OVC and the latter is rounded down to 100 ft increments below 10,000 ft altitude. The current version of `ampycloud` is only suitable to a limited extent for the complete, operational determination and reporting of cloud and VV information in METAR/SPECI or MET REPORT/SPECIAL syntax, i.e. major revisions are required:

- NSC is currently not considered but required by section 4.5.4 of *ICAO* (2018). Your Fig. B7 is just such a case where you have to report "NSC" in the METAR because there are clouds (above MSA) but they have no operational significance.
- Height/altitude values for VV can currently neither be determined nor reported although it is a requirement according to sections 4.5.4 and 4.6.5.1 of *ICAO* (2018). The CL31 provides VV hits and I wonder why the authors argue that another algorithm should handle it. Your data in Fig. B12 could also mean a so-called VV case, which leads to the possible result "VV001" in the METAR.
- Your line of argument for using a shorter integration time of Δt = 15 min is flawed. Such a short time period does not necessarily require a weighting of the latest cloud hits in order to better react to rapid changes in the cloud field, but e.g. for 30 min it makes perfect sense. METARs are representative for 8 km (even 12 km for convective clouds) radius around the airport reference point and the reporting time is 30 min. *Nadolski* (1998) pointed out that *"… a 30-minute time period provided an optimally representative and responsive observation similar to that depicted by an observer"*. Do you have observational evidence that your choice is better? Depending on the wind direction, the ceilometer detects clouds in the downstream half-space, i.e. only those clouds that have passed over the device. The crucial problem with a shorter Δt is the loss of information, which leads to less agreement with the observer who is looking at the entire half-space.
- I wonder why the authors explain `ampycloud` using Fig. 1, where obviously an integration time of Δt = 20 min (1200 s) was used, but they only speak of Δt = 15 min (900 s) both in the main text and in the figure captions of the appendix. This is an inconsistency and leads to misunderstandings. Why did you switch from 20 minutes to 15?
- Your slicing step applies the Manhattan metric as a distance function and average linkage as a fusion method. Rescaling the time axis is a clever hack to reduce the 2-D problem to 1-D, but why are you still using a 2-D clustering method that is computationally less efficient and slower? The original (historical) implementation of the `ampycloud` algorithm is no reason to deny further developments. From your initial slicing step, which is similar to the binning and clustering steps of the ASOS (*Nadolski*, 1998) you have generated two new problems: 1. To combine two cloud clusters that were incorrectly assigned to different slices but actually belong to each other you use an alternative distance function (Euclidian metric) and fusion method (single linkage). In the latter outliers often give rise to chain effects ("single-link effect") and tiny clusters that consist of only a few elements. 2. The problem that two independent

cloud clusters were incorrectly assigned to the same slice you try to solve using Gaussian mixture models. Wouldn't it be more plausible to use Gaussian mixture models to combine cloud clusters to form final cloud layers? Models with lower BIC are generally preferred but a lower BIC does not necessarily indicate one model is better than another. Have you applied other distance functions and fusion methods for the slicing and grouping steps to be sure what is the ideal choice? Is the Gaussian distribution really suitable for cloud hits in the layering step? I doubt these questions have been adequately answered.

- The lack of comparison with the results of at least one similar cloud algorithm already in operation. The authors claim in section 4 that this *"… is outside the scope of this article."*, but they need to show the added value of `ampycloud` to justify the higher computational effort and complexity.
- In section 3.4 the authors mention to use *"a reference dataset of 2128 cases extracted over a 5 year period (2018-2022) from LSGG METARs"* for the statistical evaluation and assessment of `ampycloud`. Even if SPECIs are not considered, a single year comprises 17520 METARs (twice an hour for 365 days). On the other hand, 2128 cases correspond to about 44 days. In relation to the 5-year period, it is only slightly more than 2.4 %. From a statistical point of view, this is definitely insufficient, your results are therefore not significant and the pre-selection of cases harbours the risk of concealing serious shortcomings in your algorithm.
- There is neither a controversial discussion nor at least a citation to the strengths and weaknesses of the observer, who is not the "ground truth" in all situations. This must be considered when discussing the results of `ampycloud` with regard to a possible under- or overestimation.
- The authors only present results for ideal input data, but this does not always apply to fully automatic operation. Please provide more detailed insights into the exception handling. How does your algorithm react if …
    - there is a "complete failure", e.g. due to power or network failure, and no input data is (temporarily) available?
    - there is a failure of individual or multiple ceilometers (e.g. due to maintenance)?
    - a single ceilometer measures but transmits an error code?
- Some important information about the CL31 is missing, e.g. the output interval of 15 s (4 hits per minute), which can only be inferred from the *max. hits per layer* in the diagnostic diagrams with some difficulty.

The Vaisala CL31 algorithm derives from the actual measured backscatter signal either up to three cloud hits in different heights or a value for vertical optical range (VOR) above ground level at any given time. Similar to determining cloud base altitude VV hits should be used to derive an altitude value for VV. By definition, VOR is the height above the ground up to which the extinction coefficient must be integrated in order to reach a certain threshold value, usually around 3 (*Werner et al.*, 2005). VOR is always higher than all cloud hits detected at a given time, because a well-performing ceilometer can always "see" a little into optically thick clouds and even through optically thin clouds.

However, Vaisala obviously uses a significantly smaller threshold value in order to have an alternative definition for the derivation of cloud hits. They call it a VV hit, but it's basically a cloud hit. Figures B1 and B2 show very nicely that a) cloud hits and VV hits merge almost seamlessly into one another and b) `ampycloud` would be completely wrong if the VV hits were not considered cloud hits. The crucial problem is that these two quantities are completely different things from a physical/meteorological point of view.

In the following there are specific comments to certain lines or figures of the preprint:

ll 24	MET REPORT/SPECIAL is not representative of the runway, but of the threshold. Each runway has 2 thresholds. Unfortunately, the MET REPORT/SPECIAL syntax is e.g. "... CLD RWY 05 OVC 800FT ...", which can lead to confusion.

ll 84	There are also parameters that have only computational or tuning relevance, such as your OVC/FEW thresholds or α_s in Tab. 2, which are arbitrarily chosen.
Fig. 1	Figure 1 should be presented in section 2.4 if the abbreviations and parameters have already been introduced. The figure caption of Fig. 1 is very long. I prefer to describe the method and general information in the main text with reference to this figure. All figures in Appendix B and Fig. 1 should then be self-explanatory.
ll 120	As you mention in relation to "nnn" in the caption of Fig. 1, a ceilometer does not provide vertical values relative to mean sea level (altitude = elevation + height), but it gives the height above ground level. The elevation corresponding to a specific location must be explicitly added to the settings.
lls 127 – 128 and 134 – 135	*"Data are pooled together before the analysis, independently of the spatial distribution of the ceilometers."* and *"The specification of h_{cid} for every hit is necessary to compute a correct estimation of …"* are contradictory statements.
ll 140	What is meant with "signal range"? Is it the maximum range of detection?
Fig. 2	Figure 2 should not be shown in the main sections, but at most in the appendix, as it is essentially just a repetition of the results from section 5 of *Wauben* (2002). The added value of $n > 1$ ceilometers for distinguishing BKN and OVC is rather limited, as both mean a ceiling in the end.
ll 160	Meanwhile Vaisala provides an "airplane filter" and maybe it can even be utilized by CL31 with a firmware upgrade.
lls 262 - 263	The triangles and rectangles in the blue slice in Fig. 1 reveal that $\alpha_s \cdot (h_{max} – h_{min}) \approx 0.2 \cdot (2{,}100 \text{ ft} – 950 \text{ ft}) = 230$ ft. The distance between the two sub-layers is definitely larger than 230 ft. So why are they combined during the slicing step?
lls 350 – 351	Or the differences are due to the observer's limited ability to distinguish between slightly more or less than 50 % cloud cover.
ll 357	Lower cloud bases occur almost twelve times (24.6 %) as often as higher ones (2.1 %). Therefore, it is not a tendency but systematic. This is potentially problematic when applying the SPECI criterion in section 2.3.2 f) of *ICAO* (2018).
ll 360	Or your Δt is too short and you miss some clouds that passed over CL31 earlier.
Fig. B4	Perhaps it is the same cloud layer that has passed over at least 2 ceilometers with a time delay. Then the two cloud clusters (green and red) should be combined into a single cloud layer. This would be more consistent with the 2nd cloud group reported by the observer.
Fig. B5	The cloud hits around 9,100 ft and 9,500 ft belong definitely to a single cloud layer, as it is a similar situation as for the red slice in Fig. B6. All hits together would likely result in SCT.
Fig. B6	The cloud hit distribution and your slicing step clearly speak in favour of 3 cloud layers similar to the observer-based METAR meaning that the grouping and layering steps fail.
Fig. B10	Perhaps your Δt is too short and you therefore cannot reproduce the clouds at around 3,500 feet.

Technical corrections

lls 1, 15, 45, …	Either "sky coverage" or "cloud amount" but use a consistent wording.
ll 28	I would prefer "international" rather than "civil" airports, as regional airports as a different category are not limited to goods and trade.
ll 29	"in addition" instead of "an additional"

ll 59	“cloud layers” without 2nd “s”
ll 61	“its performance” singular is enough
lls 66 – 71	This paragraph should be moved to section 2 or even section 3.
lls 98 – 99	Move these two sentences and Fig. 1 to section 2.5 and shortly refer to it here avoiding repetition of the same content.
ll 111	Use “cloud and VV hits” instead of “ceilometer hits”.
Fig. 1	- Δt = 20 min should be the same as for figures in Appendix B (15 min) - omit sup-/subscripts and explanation of OVC/FEW thresholds as these information can also be found in Tab. 2. - which airport is shown? (MSA: 8,000 ft refers to LSZH but n_{ceilos}: 4 to LSGG) - max. hits per layer: 160 is incorrect $n_{\text{ceilos}} \cdot$ hits/min $\cdot\ \Delta t = 4 \cdot 4 \cdot 20 = 320$ (doesn't fit) but $= 2 \cdot 4 \cdot 20 = 160$ (fits)
ll 117	Use “cloud hit” instead of “ceilometer hit”.
ll 120	“derived cloud base height above ground level” or “derived cloud base altitude above mean sea level” but no mixture and not including “measured”
ll 124	“backscatter profiles”
Tab. 1	Maybe add the CL31 output interval of 15 s, the elevation, MSA, etc.
ll 148	- use “time interval”, “time increment” or “integration time” but not “length” - “can be bundled” instead of “can be pooled”
ll 151	“… time difference relative to the latest.”
lls 157 – 159	“Assuming *max. hits per layer* is 160, then we use a threshold value of 99.375 % for OVC and 1.875 % for FEW, accepting 1 hole for the former and requiring 4 hits for the latter.” Your wording is far too cumbersome. Why do you apply an implicit dependency on *max. hits per layer* instead of defining fixed percentages?
lls 178 – 179	“This distance threshold is Δh_l = 250 ft between 0 – 10,000 ft (1,000 ft above 10,000 ft).” Your wording is rather cryptic.
lls 186 – 187	such repetitions should be avoided
Tab. 2	- “behaviour” in the figure caption - table too long and covers page number - add the meaning of parameters in a separate column and omit Python variable
ll 206	misleading footnote (could be power 3)
lls 222 – 223	“By (our) definition, slices *i* and *j overlap* if *one* of the following two conditions *is true/met*:”
ll 227	use “For example, the two top slices in Fig. 1 (green and red) overlap, …”
ll 241	use “the latter value” instead of “this value”
ll 250	A reference to Tab. 2 is missing for L_{frac} and L_{it}
ll 271	“overestimation” instead of “overestimate”
ll 276	use precise wording to avoid "In other words, ..."
ll 277	use “assumes” and “number” instead of “will deem” and “amount”, respectively
ll 300	Avoid text extending beyond the edge of the page.
ll 304	the correct reference is “(ampycloud, 2024a)
ll 334	use “5-year period”
ll 348	use “The comparison of Fig. 3 and Fig. 4 shows, that false alarms …”
ll 355	“height/altitude of the cloud layer with the highest degree of coverage” instead of “densest cloud layer”
ll 357	“than” instead of “that”

ll 359	consistently use either "cloud base height" or "cloud base altitude"
Fig. 5	maybe use “hft” (hecto feet) instead of “100 ft”
ll 416	“the accompanying material” instead of “associated supplementary material”
lls 426 – 428	This information does not belong in the conclusions.
ll 487	“figure caption”
Figs. B5, B11	*max. hits per layer* must be 240

References

Gao, W., C.-H. Sui, and Z. Hu (2014): A study of macrophysical and microphysical properties of warm clouds over the Northern Hemisphere using CloudSat/CALIPSO data, *J. Geophys. Res. Atmos*.,119, 3268–3280, doi:10.1002/2013JD020948.

Illingworth, A. J., Cimini, D., Haefele, A., Haeffelin, M., Hervo, M., Kotthaus, S., Löhnert, U., Martinet, P., Mattis, I., O’Connor, E. J., and Potthast, R. (2019): How Can Existing Ground-Based Profiling Instruments Improve European Weather Forecasts? *Bulletin of the American Meteorological Society*, 100(4), 605-619. https://doi.org/10.1175/BAMS-D-17-0231.1.

International Civil Aviation Organization (2018): Meteorological Service for International Air Navigation, Annex 3 to the Convention on International Civil Aviation, 20th Edition, AN 3, ISBN 978-92-9258-482-5.

Luebke, A. E., Ehrlich, A., Schäfer, M., Wolf, K., and Wendisch, M. (2022): An assessment of macrophysical and microphysical cloud properties driving radiative forcing of shallow trade-wind clouds, *Atmos. Chem. Phys.*, 22, 2727–2744, https://doi.org/10.5194/acp-22-2727-2022.

Nadolski, V. L. (1998): *Automated Surface Observing System (ASOS) User’s Guide*, Tech. rep., 112 pp. National Oceanic and Atmospheric Administration, Department of the Air Force, Federal Aviation Administration, and United States Navy, USA.

Vaisala (2009): *User’s Guide Vaisala Ceilometer CL31*, Tech. rep., M210482EN-D, 142 pp., Vaisala, Finland.

Wauben, W. M. F. (2002): Automation of Visual Observations at KNMI: (II) Comparison of Automated Cloud Reports with Routine Visual Observations, in: Symposium on Observations, Data Assimilation and Probabilistic Prediction, AMS Annual Meeting, Orlando (FL), USA.

Werner, C., Streicher, J., Leike, I., Münkel, C. (2005). Visibility and Cloud Lidar. In: Weitkamp, C. (eds) Lidar. Springer Series in Optical Sciences, vol 102. Springer, New York, NY. https://doi.org/10.1007/0-387-25101-4_6.

World Meteorological Organization (2017): *International Cloud Atlas*, 2017 Edition, https://cloudatlas.wmo.int.