

This article presents a new algorithm: ampycloud, deployed by MeteoSwiss at different aerodromes within the country. As the authors state, given the opaqueness of ceilometer manufacturers' cloud layer height and amount algorithms, using these at a given aviation meteorological institute remains difficult. The article makes a cloud layer algorithm available, along with a suite of test cases for interested users to apply. The author present test cases and statistical studies in order to validate their algorithm. These, together with the algorithm deserve publication.

However, changes in the manuscript must be made in order to improve its clarity. Specifically, more explanation is needed for certain parts. For example, the algorithm is tunable through many parameters, yet, the authors do not explain the reason for their choices of various constants. Also, their choice of integration time is not well argued and it is not immediately clear why vertical visibility is not reported with ampycloud. These comments are detailed in the following sections

General comments:

Line 32: Please define auto metreport before using this name

Line 55-56: It also prevents combining different types of ceilometers. Combining different types of ceilometers might be done for e.g. financial reasons.

Introduction or section2:

Some background reasoning or framing on why ampycloud is designed the way it is -> It uses rather advanced statistical methods compared to for example the Larsson algorithm used by the KNMI (see e.g. Wauben et al. 2002). What are the advantages/disadvantages of ampycloud compared to e.g. the algorithm described by Wauben?

line 94: I find (pitfall 1) difficult to understand

Line 103: "Obscuration of the sky by fog or snow must be handled by a separate algorithm". Please clarify as to why this is the case.

Line 111: "relies on ceilometer hits". Perhaps a brief description of what a ceilometer hit is, together with a few words on how a ceilometer calculates cloud base heights would help the reader.

Line 120: what is meant by relative time (It is described later, but perhaps better to describe it here)?

Line 123-125: "One should note that for the moment, ampycloud cannot use full back-scatter profiles to derive cloud base hits independently from the ceilometers' proprietary software." Since deriving your 'hits' from backscatter profiles would just mean that you're clustering different cloud base hits, does this affect the quality of the algorithm?

Line 130: definition of AUTO METREPORT should be done earlier e.g. in the introduction

Line 132: Discussion Figure 2 should be in the main text.

Line 134-135: "The specification of hcid for every hit is necessary to compute a correct estimation of the sky covering fraction under the assumption that cloud layers have a unique base per time step per ceilometer line-of-sight.". Could you explain this further? Why is the hcid necessary for the correct estimation of the sky covering fraction given that all hits are treated equally, pooled together and sorted?

Line 149: a delta t of 15 and 6 minutes seems arbitrarily chosen: Could you provide more details as to why 15 and 6 minutes were chosen?

Line 155: I was under the impression that compliant METAR cloud layer codes were derived sequentially, meaning that first S, then G then L were computed in order to finally give up to 3 layers. With the current wording, it seems that a METAR code is generated for the 'S step', the 'G step' and the 'L step'. Could you please clarify?

Line 156: Does that mean that a cloud layer is considered at the 4 hit? Does that mean that the cloud information other than NCD would be available to the AUTO METAR generation after 1 minute of detecting the beginning of a cloud layer (given a time resolution of 15 seconds for the ceilometer)? Could you say a few words on the operational suitability of this?

Line 170-172: "If a user were to prefer the cloud base heights to be more representative of the most recent hits within a slice/group/layer, the B_t parameter can be set to lower values, e.g. B_t = 30%. Note that changing the value of _t does not have any effect on the cloud amounts." As far as I understand, this is not the same as doubly weighing the most recent cloud hits. When doubling the most recent cloud hits, the oldest cloud hits still have a single weight, and so are still considered in an algorithm (e.g. Larsson's algorithm). In ampycloud, setting B to 30% equates to only using 30% of the data within a slice. Is this correct? If not, could you please add a few words to clarify?

Line 164-165: "as commonly done by other cloud height algorithms" Are there any references to this?

Line 209: why is alpha_s chosen as 0.2?

Line 220: perhaps refer again to the over-slicing as the top layer of figure 1 to add clarity

Line 229: Please give a reason for choosing epsilon as 10%

Line 245: how is fb chosen?

Line: 250 how is Lfrac chosen?

Line 330: Could you please give more information on how a human observer collects/validates data for his METAR: they derive a cloud layer height and amount based on the clouds in an area above and in the vicinity of the aerodrome. But do they also use raw ceilometer cloud hits? Do they use the results of another algorithm? Are they trained to more or less estimate cloud heights without ceilometer information?

Line 332: how were the cases selected? Did you only select cases where human observers reported something operationally significant? Or did you select cases where either a human or the algorithm detected something important? Or did you select cases where both the algorithm and the human observers agreed that something of operational significance was occurred? Or was there another criteria used? Please specify as this could have consequences on the number of misses and false alarms. Also, can you say if using a larger sample (for example by using the whole 5 years) would affect the statistics?

Line 336: What is meant with operationally significant cloud events: are these ceilings? Low clouds? Low ceilings? Situations where at least 2 layers are present with at least a SCT in the second layer? And since these represent such a small sample compared to the whole 5 year dataset, does that mean that ampycloud is designed to be most accurate for operationally significant events at the expense of other events? Please clarify this.

Line 338-344: what is meant by a density of a cloud? Is it the number of oktas?

Line 346: lines 333-336 seem to suggest that the 2128 cases are the high impact cases. In line 346 a further selection is made? Could you please clarify this?

Line 367-369: "Understanding the exact behavioral differences between ampycloud and other algorithms with similar purpose would require a dedicated comparison of their respective performances against a reference dataset, which is outside the scope of this article." Please provide more detail as to why such an analysis is not done. It is true that algorithms from manufacturers are kept secret, but in the case of Vaisala it is possible (against payment) to have the outcome of their algorithm in terms of cloud layer heights and cloud amount. Could you explain why a comparison of this output against ampycloud is not done? Furthermore, what is meant with a reference data set in this case?

Line 415-417: Should this be in the results section?

Line 419-429: In the method section?

Figure B10: low visibility procedures should be introduced

Figure B12: information about VV being decided by a separate algorithm should be in the main text in the introduction or methodology section, together as to why meteoswiss/the authors decided this that way. Also, as VV is not within ampycloud, it should be emphasized that ampycloud is only partly compliant with ICAO's rules for cloud reporting.

Minor comments

Please use 'inputted' instead of 'fed'

Lines 94-95: starting the sentences with 'It' may make the sentences more readable

Lines 114-115: Better to write: "The longer the time interval or the larger the wind speed, the better the spatial representativity of the dataset, but the worse is the view of the current state of the sky"

Line 127: use 'apply' instead of 'run'

Line 353-355: "But it cannot guarantee that the "correct" cloud layers are being identified when they have the correct density." Perhaps better to say that detecting the right number of oktas for a cloud layer cannot guarantee that the cloud layer has the correct height?

Line 390: small amount of information implies that you are talking about a small number of cloud hits, which is not the case. In this sentence, please consider saying that it relies on cloud hits only.

Line 399: use the word "detect" instead of "see"

Line 489: Please define "scientific stability"