Introduction

We thank referee #1 for his/her careful reading, comments and suggestions which we address in the following. The authors' answers are printed in italics.

Remark: The figure numbers in the referee comments and the page numbers in the authors' answers are corresponding to the original manuscript. If not stated otherwise, figure and equation numbers in the authors' answers are referring to the revised, marked-up manuscript version (showing the changes made) which can be found at the end of this text.

General comments

– The submitted manuscript deals with the 3-D reconstruction of clouds via the structure-from-motion technique using image data obtained from a downward-looking camera installed at a research aircraft. The goal is to provide 3-D cloud top geometry information and geolocation that can be used to improve retrievals by other remote sensing methods, e.g. derivation of cloud droplet radii from spaceborne hyper-spectral imagery. While airborne observation of clouds is a costly enterprise and delivers data only for the flight period, it provides a quite complete and still rare view on the cloud top geometry.

  The article describes in detail the methodology and evaluation of the proposed airborne reconstruction, including camera calibration, feature tracking and 3-D reconstruction. Besides an empirical evaluation with an onboard lidar system, the article discusses related challenges of such an approach, such as synchronization with the aircrafts navigation system or the effect of cloud evolution and motion during the sequence of photographs. The article proves that the structure- from-motion technique can be successfully applied to obtain the 3-D cloud top geometry of clouds and should be published after dealing with the following remarks.

  → *Thank you for your helpful and supportive review. We generally agree with your comments and are confident that we could improve the manuscript quite a bit with your support. At the end of this text you will find a diff for the revised manuscript.*

Specific comments

– For the purpose of evaluation, the article yields a case study of tracked features (Fig.2) and an illustration of the retrieved data (Fig.3). While Fig.3 shows that the method allows to detect cloud evolution (arrows), the missing spatial reference, height information and the large dataset makes a proper interpretation difficult.

  It would be helpful if the reader would be able to connect the shown 3-D data with the cloud scene shown in in Fig.2. Maybe it is possible to exclude the more distant 3-D data and introduce some regions of interest, such as individual cloud turrets, that could be marked in Fig.2 and then used in Fig.3 to provide a direct connection. Also, the shown arrows could encode the mean height by an appropriate color code, as done in Fig. 5. This might have the advantage that the reader can estimate the cloud geometry directly.

  → *Indeed, it is difficult to connect the 3-D data with 2-D images. While it is possible to draw a 2-D image into the 3-D plot, this does not yield much benefit. Showing this data from a different perspective than the camera perspective is almost as hard to understand as the figures presented in the discussion manuscript. Thus we chose to present only a single figure showing the camera's perspective as in the previous figure 2 but added color coded height information as well as cloud movement vectors as you suggested. This way,*

the reader can estimate the cloud geometry more easily and relate it to the actual image. Additionally, we marked the location of this image in figure 7, which shows the wind field in the larger area.

- Fig.5 gives a nice overview of the techniques capabilities on a large scale. Two points of critique here: First, the figure encodes the height as color, but lacks a legend. Second, the figure shows the dataset over a quite large extent. It might help to add a detail view of a specific region of interest contained in the large-scale view, such as a local two-layer situation.

  → *We have added a legend and a magnification of the central part of the scene. It highlights two cloud layers and a small cloud patch above both of them.*

- Fig 6 and 7 may be combined into one figure as both intend to to show (among others) the challenge of a proper comparison between lidar and stereo data.

  → *We have combined the figures and added an arrow to mark the relevant region of the comparison plot.*

Technical Corrections / Suggestions

- **P1, 5:** "...relatively simple installation on an aircraft..."

  Maybe simple in case of a dedicated research aircraft, but most probably not in general.

  → *We've added "(research)" before "aircraft". We agree that it is certainly easier to install the system on research aircraft. On general purpose aircraft, the lack of apertures might prevent an easy installation but still, the discussed single camera system requires very little space and the only additional requirement is an accurate navigation system, which should be available on most bigger aircraft anyways. If that is not available, such a navigation system mostly consists of a box to be attached statically somewhere on the airframe and connected to a GPS antenna. So compared to other, especially bigger or active sensors, this system is indeed simpler to install on a general aircraft as well.*

- **P1, 7:** „However we will show that to some extent usable wind information can also be recovered."

  More precise ("to some extent").

  → *We now refer to the filtering of outliers.*

- **P2, 21:** „...a big advantage when observing moving and changing clouds."

  Maybe better: „...., so that cloud evolution and motion does not affect the 3-D reconstruction."

  → *Changed accordingly.*

- **P3, 16:** „For geometric calibration of the camera we use a common approach."

  Which approach? More precise.

→ *Has been changed to "For geometric calibration of the camera we use a common approach of analyzing multiple images of a known chessboard pattern to resolve unknown parameters of an analytic distortion model." For further details, the reader is referred to the appendix.*

– **P8, 6:** „After all filtering...“

Delete „all“.

→ *Changed accordingly.*

– **P8, 8:** „Such a point cloud is shown in figure 5.“

Maybe just put the figure reference at the end of the previous sentence and delete this sentence („... relative to a point on the earth’s surface (figure 5).“)

→ *Changed accordingly.*

– **P8, 8-10:** „This point cloud can then be used as a starting point...“

Maybe better: „The point cloud can then serve as reference for other distance measurement techniques...“ (Which?) „ ...or allow for a 3-D surface reconstruction.“

→ *Changed to "This point cloud can be used on its own, serve as a reference for other distance measurement techniques (e.g. oxygen absorption methods (Zinner et al., 2018) distances derived by a method according to Barker et al. (2011)) or allow for a 3D surface reconstruction."*

– **P9/10, 20/1:** „Generally, there is a good agreement...“

Maybe better: „The measured distances between the aircraft and clouds as obtained from the WALES lidar and the stereo method show a good agreement...“ (typical errors?) „... .The automated comparison between lidar and the stereo method, however, typically includes a significant number of outliers in multi- layer cloud situations.“

→ *We reviewed the data for this comparison in order to better quantify the typical errors. Still, we were not able to find a sensible method of removing clear outliers due to comparing different clouds without manual filtering. We prefer not to introduce an artificial bias into the comparison by adding subjective criteria. Therefore we added a reference to Stevens et al. (accepted) and additional explanation about the difficulties in comparing the sensors (different sensors see different clouds). We decided that out of this reasons, quantitative comparison with lidar data is only useful for bias, not for spread. On the other hand, for homogeneous cloud decks, as investigated in the across track stability section, the internal spread of the stereo method can be quantified (*$47.3\ \mathrm{m}$ *standard deviation in this case). We swapped the order of sections 4.1 and 4.2 to support this argument.*

– **P10, 27:** „...have been binned in 1 min bins...“

Maybe better: „....have been binned in time intervals of 1 minute...“

→ *Changed accordingly.*

References

Barker, H. W., Jerg, M. P., Wehr, T., Kato, S., Donovan, D. P., and Hogan, R. J.: A 3D cloud-construction algorithm for the EarthCARE satellite mission, Quarterly Journal of the Royal Meteorological Society, 137, 1042–1058, https://doi.org/10.1002/qj.824, http://dx.doi.org/10.1002/qj.824, 2011.

Stevens, B., Ament, F., Bony, S., Crewell, S., Gross, S., Hirsch, L., Mayer, B., Wendisch, M., Wirth, M., Bakan, S., Brück, H.-M., Ehrlich, A., Ewald, F., Farrell, D., Forde, M., Gödde, F., Grob, H., Hagen, M., Hansen, A., Jacob, M., Jäkel, E., Jansen, F., Klepp, C., Klingebiel, M., Kölling, T., Konow, H., Mech, M., Peters, G., Rapp, M., Wing, A., and Wolf, K.: A high-altitude long-range aircraft configured as a cloud observatory - the NARVAL expeditions, Bulletin of the American Meteorological Society, accepted.

Zinner, T., Schwarz, U., Kölling, T., Ewald, F., Jäkel, E., Mayer, B., and Wendisch, M.: Cloud geometry from oxygen-A band observations through an aircraft side window, Atmospheric Measurement Techniques Discussions, 2018, 1–20, https://doi.org/10.5194/amt-2018-220, https://www.atmos-meas-tech-discuss.net/amt-2018-220/, 2018.

Introduction

We thank referee #2 for his/her careful reading, comments and suggestions which we address in the following. The authors' answers are printed in italics.

Remark: The figure numbers in the referee comments and the page numbers in the authors' answers are corresponding to the original manuscript. If not stated otherwise, figure and equation numbers in the authors' answers are referring to the revised, marked-up manuscript version (showing the changes made) which can be found at the end of this text.

General comments

– This manuscript showcases a novel technique of using well known computer vision techniques to reconstruct cloud geometry and is a valuable contribution to science. This referee suggests this paper should be published, following some revisions, see below for the major and minor issues. The overall content of the paper is well formed, but the introduction and concluding sections require multiple typo corrections. The included comparison to lidar is well received, although the choice of a large area of cloud top height comparison should be revisited.

→ *Thank you for your very helpful comments. Based on your suggestions, we've had new insights into details of our measurement system and the described method. While the comment on spectral aberration lead to an intensive and interesting re-investigation of our data, we finally decided not to perform any changes as the effects are comparably small.*

Major issues

– Verification of the method uses a dubious assumption of cloud homogeneity within 150m of the lidar measurement, refinement should be done, and subsequent conclusions of the lidar representing higher clouds is put into question.

→ *Short version: while the choice of the radius of the comparison cylinder is to some extent arbitrary, no systematic error should be introduced by any particular choice. The differences in cloud representation between stereo and lidar method are expected because of general reasons. Our conclusions merely state that our observations agree with these general reasons.*

Longer version: the authors are aware that there is no reason, why a cloud should be homogeneous within any chosen radius for a physical reason. This is also not the point to be made in the comparison. The assumption is rather that clouds or parts thereof which are co-located horizontally are also co-located vertically. As discussed later in the corresponding section, this assumption is not valid in areas of multi-layer clouds. In areas of single layer clouds, this assumption should however hold. Comparing data from within a vertical cylinder should therefore result in results of similar height. While the individually measured data pairs may scatter broadly due to cloud inhomogeneities, apart from systematic differences in the measurement principle, there is no reason to believe that the mean or median deviation of all data pairs should be different from zero.

The size of the cylinder is rather arbitrary but the particular choice has reasons: the aircraft moves at a speed of approximately $200\ \mathrm{m/s}$ *and the data of the lidar system is available at* $1\ \mathrm{Hz}$ *and averaged over this period. Any comparison between both systems should therefore be in the order of* $200\ \mathrm{m}$ *horizontal resolution. Furthermore, data derived from the stereo method is only available where the method is confident that it worked. Thus not every lidar data point has a corresponding stereo data point. Increasing*

the size of the cylinder increases the count of data pairs, but also increases false correspondences. The general picture however remains unchanged.

The statement that lidar is representing higher (parts of) clouds is not really a conclusion but rather an expectation due to the measurement principles. First, the stereo method is using slanted observation directions in addition to the nadir direction, thus and because it uses regions of high image contrast, the method tends to favour cloud sides, which are below the top of a cloud. The lidar observes any part of the cloud which is visible from directly above. On average, the parts of the cloud which are observed by stereo thus should be below of the parts which are observed by the lidar method. Second, the lidar has been designed to be able to observe barely visible or even invisible parts of the atmosphere, so it is clearly more sensitive than the camera. As the algorithms used select signals closest to the aircraft, a more sensitive system should prefer higher clouds.

Our conclusion states that our analysis is in accordance to this expectation.

– "cloud surface" has not been defined, yet it underpins this manuscript. Cloud surface is not what the feature selection algorithm is used, but rather cloud surface edges. Clarification should be done.

→ *We have added a paragraph describing our definition of the cloud surface to the introduction. We also changed the following paragraph, which now reflects that we are searching for points on the surface, rather than generating a full surface model.*

Minor issues

– Title of the manuscript is slightly misleading, common wording for this methodology is 'Structure-from-Motion', see Westoboy et al., 2012 (amongst others)

→ *We have thought about this potentially misleading naming and think that for our method, stereo is slightly more applicable as Structure-from-Motion (SfM). As Westoby et al. (2012) describe, a typical feature of SfM methods is to reconstruct both, the observed structure as well as the motion of the camera only from an image sequence. Due to the unknown movement and deformation of the observed clouds, this approach is less feasible in our case. Also our own experiments show that the commonly and particularly used SIFT feature matching algorithm (Lowe, 2004) does not work very good on non-solid surfaces as clouds and ocean provide. Due to these limitations, we depend on highly accurate position and orientation information of the aircraft, which is in our case provided by the aircraft's basis instrumentation. Furthermore, SfM methods tend to use more than two images simultaneously to perform object recognition. This also does not happen in our method, but rather the results of processed image pairs are combined in a later step. Thus the actual image processing step is very close to classical stereographic methods. So while it is true that we uses the camera's motion to derive structural information, we do not use typical SfM techniques. This is why we choose not to use the term Structure-from-Motion.*

– Point selection algorithm choice has not been described. Some description of these selection points, for finding the corners would be a welcomed addition to this manuscript.

→ *We have reformulated the corresponding paragraph and explain the point selection algorithm and the reasoning behind choosing this algorithm in more detail.*

– Figure 3 is nearly useless without a better frame of reference. Please include a frame of reference marker. It may be useful to put and 'x-y-z' axis in Fig. 2, and the rotated version of which in Fig. 3.

→ *This has also been pointed out by referee #1. We agree that this figure is barely understandable. Displaying the corresponding data from a different than the camera's perspective remains to be hard to understand, so we decided to combine figures 2 and 3 into an improved new figure 2. This way, the points can be associated visually to the image and height information is now included as color codes.*

– Figure 5 should have a colorbar to denote the color scheme of the cloud height.

→ *Changed accordingly*

– Last paragraph of section 3 describe transformation of a point cloud to cartesian 3D, but does reference the use of the aircraft navigation, or potential sources of errors from it.

→ *In the beginning of section 3, we explain that the transformation into a geocentric reference frame is performed using the aircraft's navigation system. Together with the addition about landmarks (see next item), our procedure should now be more clear to the reader. By adding the paragraph you proposed in the second next item to the end of section 3, the reader is pointed to the requirement of having an accurate time synchronization (at the order of tens of milliseconds) between navigation system and sensors, which we see as the biggest challenge in transforming the points to a geocentric reference frame.*

– Section 4.2 is using data from a status cloud deck to infer cross track stability of the measurement. Further evidence of the status cloud deck's vertical stability should be presented to reinforce this point. If no other is available, is it possible to use a ground target instead of the cloud to cross track stability? Related remarks in the conclusion should be amended

→ *In fact, the orientation of the camera with respect to the aircraft has been determined independently of any clouds by aligning camera images of landmarks (taken on multiple flights) with satellite images. We added an according note to the first paragraph of chapter 3. In consequence, the orientation of the cloud deck used to determine the across track stability of the presented method has not been used to determine the camera's orientation. We thus are quite confident that potential systematic deviations would be visible in the presented plot. Remaining inhomogeneities should already be included as detrimental effects in our conclusions. We avoid using ground targets to assess across track stability, first because we want to include effects which are specific to observing clouds (if any) and second, because most of our measurements have been conducted over ocean or clouds, leaving only a few ground targets, which also have been used to align the camera and thus should not be used for assessment to avoid circular logic. For further clarification, a note has been added to the stability analysis as well.*

– Last paragraph of the conclusions should be inserted in the methods as well, and references to the appendix.

→ *A modified version of the last paragraph of the conclusions has been inserted at the end of the methods section.*

– A note on the spectral aberrations (if any) would be useful in the appendix A.

→ *It is indeed a very good idea to look at spectral aberrations, which we did not prior to your comment. To further investigate effects of spectral aberration, we rerun our calibration procedure for all color channels separately and reprocessed parts of the measurement data. A comparison of different calibration data reveals that choosing a different calibration changes an observed viewing angle difference in the order of* $1-2‰$*. For a cloud distance of* $10\ \mathrm{km}$ *as in the calibration example, this translates to uncertainties of about* $10\ \mathrm{m}$*, which is also confirmed by rerunning the analysis of the horizontal cloud deck. Interestingly, despite the same order of magnitude, we did not find that the curvature of the cloud deck analysis can be removed by considering spectral aberrations. Due to effectively using fewer pixels when doing the camera calibration procedure on a single channel image, the reprojection error is increased accordingly. In the end, we choose not to use different calibration data for different color channels, as the effects turn out to be smaller than other sources of error and using only a single calibration facilitates data handling. We did not consider effects of spectral aberrations within a single color channel, but assume that these effects should be even smaller and thus can be neglected as well. We have added a note on this analysis to appendix A.*

Here are some specific points to be addressed

- **P.1 line 10:** typo: “comparson” should be “comparison”

  → *Changed accordingly*

- **P.1 line 16:** what the authors describe is unclear: “where observed clouds and observer are at different locations,... “

  → *The sentence has been reworded.*

- **P. 1 line 21:** why is the term “Finally, ...” used at the start of the sentence? Flow of the entire paragraph should be reevaluated.

  → *The sentence starts an additional use case. We have split the paragraph into two changed the wording.*

- **P. 1 line 25:** “by Ewald (2016); Ewald et al. (2018)” should be “by Ewald (2016) and Ewald et al. (2018)”

  → *Changed accordingly*

- **P.1 line 26:** Unclear grammar to what “it is shown [...]” is referencing, Is it “Ewald et al. (2018) showed that [...]” ?

  → *Changed to "In particular, Ewald (2016) and Ewald et al. (2018) have shown ...".*

- **P. 6 caption of figure 2:** Unknown symbol of ’^’ on top of ’=’, please define or use more widely known character.

  → *Due to reworking the previous figures 2 and 3, the mentioned symbol is not present anymore.*

- **P. 8 line 3:** grammatically unsound “because due to the [...]”, please rephrase.

  → *Has been rephrased.*

- **P. 8 line 5:** please be more precise in this sentence “these clouds can still be tracked in the presence of sunglint.” - related conclusion remarks should also be ammended

→ *We now describe the situation more precisely, both in the marked section, as well as in the respective part of the conclusions.*

– **P. 12 line 3:** "active remote sensing in the nadir perspective" seems odd, maybe: "nadir pointing active remote sensing"

→ *Changed accordingly*

– **P. 12 line 5:** please remove capitalization of "Because"

→ *Changed accordingly*

– **P. 12 line 7:** typo: "requirment" should be "requirement"

→ *Changed accordingly*

– **P. 12 line 20:** "in stead" should be "instead"

→ *Changed accordingly*

References

Ewald, F.: Retrieval of vertical profiles of cloud droplet effective radius using solar reflectance from cloud sides, Ph.D. thesis, LMU München, http://nbn-resolving.de/urn:nbn:de:bvb:19-205322, 2016.

Ewald, F., Zinner, T., Kölling, T., and Mayer, B.: Remote Sensing of Cloud Droplet Radius Profiles using solar reflectance from cloud sides. Part I: Retrieval development and characterization, Atmospheric Measurement Techniques Discussions, 2018, 1–35, https://doi.org/10.5194/amt-2018-234, https://www.atmos-meas-tech-discuss.net/amt-2018-234/, 2018.

Lowe, D. G.: Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60, 91–110, https://doi.org/10.1023/B:VISI.0000029664.99615.94, http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94, 2004.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., and Reynolds, J. M.: 'Structure-from-Motion' photogrammetry: A low-cost, effective tool for geoscience applications, Geomorphology, 179, 300–314, https://doi.org/https://doi.org/10.1016/j.geomorph.2012.08.021, http://www.sciencedirect.com/science/article/pii/S0169555X12004217, 2012.

Aircraft based Stereographic Reconstruction of 3D Cloud Geometry

Tobias Kölling[1], Tobias Zinner[1], and Bernhard Mayer[1]

[1]Ludwig Maximilians Universität, Meteorologisches Institut, München, Germany

Correspondence: T. Kölling (tobias.koelling@physik.uni-muenchen.de)

Abstract. This work describes a method to retrieve location and geometry of clouds using RGB images from a video camera on an aircraft and data from the aircraft's navigation system. Opposed to ordinary stereo methods where two cameras with fixed relative position at a certain distance are used to match images taken at the exact same moment, this method uses only a single camera and the aircrafts movement to provide the needed parallax. Advantages of this approach include a relatively simple installation on ~~an~~ a (research) aircraft and the possibility to use different image offsets, even larger than the size of the aircraft. Detrimental effects are the evolution of observed clouds during the time offset between two images as well as the background wind. However we will show that ~~to some extent usable~~ some wind information can also be recovered and subsequently used for physics based filtering of outliers. Our method allows the derivation of cloud top geometry which can be used, e.g., to provide location and distance information for other passive cloud remote sensing products. In addition it can also improve retrieval methods by providing cloud geometry information useful for the correction of 3D illumination effects. We show that this method works as intended by ~~comparson~~ comparison to data from a simultaneously operated lidar system. The stereo method provides lower heights than the lidar method, ~~on average by~~ the median difference is $126\ \mathrm{m}$. This behaviour is expected as the lidar method has a lower detection limit (leading to greater cloud top heights for the downward view) while the stereo method also retrieves data points on cloud sides and lower cloud layers (leading to lower cloud heights). Systematic errors across the measurement swath contribute less than $50\ \mathrm{m}$.

1 Introduction

~~In cloud remote sensingapplications, where observed clouds and observer are at different locations~~As implied by the name of remote sensing, the observer is located at a position different from the observed objects. Accordingly, the location of a cloud is not trivially known in cloud remote sensing applications. Thus, cloud detection, cloud location and cloud geometry are parameters of high importance for all consecutive retrieval products. These parameters themselves govern characteristics like cloud mass or temperature and subsequently thermal radiation budget and thermodynamic phase. Typically passive remote sensing using spectral information is used to retrieve cloud properties including cloud optical thickness, effective droplet radius, thermodynamic phase or liquid water content. However, these methods ~~can not~~ cannot directly measure the cloud's location. ~~Finally~~To put the results of such retrieval methods into context, the location must be obtained from another source.

Additional to a missing spatial context, unknown cloud location and geometry are the central reason for uncertainties in microphysical retrievals because of the complex impact of 3D structures on radiative transport (e.g. Várnai and Marshak, 2003; Zinner and Mayer, 2006). The classic method of handling complex, inhomogeneous parts of the atmosphere (e.g. typical MODIS retrievals) is to exclude these parts from further processing. This of course can severely limit the applicability of such a method. As shown by ~~Ewald (2016); Ewald et al. (2018)~~ Ewald (2016) and Ewald et al. (2018) the local cloud surface orientation affects retrieval results. In particular, ~~it is~~ Ewald (2016) and Ewald et al. (2018) have shown that changes in surface orientation and changes in droplet effective radius produce a very similar spectral response. Thus an independent measurement of cloud surface orientation would very likely improve retrieval results on droplet effective radius.

As location and geometry information is of such a great importance, a couple of different approaches to get this information can be found. Among these are active methods using lidar or radar. Fielding et al. (2014) and Ewald et al. (2015) show how 3D distributions of droplet sizes and liquid water content of clouds can be obtained by the use of a scanning radar. Ewald et al. (2015) even visually demonstrate the quality of their results by providing simulated images using the retrieved 3D distributions as input and comparing them to actual photographs. Major downside of this approach is the limited scanning speed. Consequently these methods are especially difficult to employ on fast moving platforms. For this reason, the typical implementations of cloud radar and lidar on aircraft only provide data directly below the aircraft.

Passive methods are often less accurate but can cover much larger observation areas in shorter measurement times. They typically either use spectral features of the signal or use observations from multiple directions. MODIS (Moderate Resolution Imaging Spectroradiometer) cloud top height for example uses thermal infrared images to derive cloud top brightness temperatures (Strabala et al., 1994). Using assumed cloud emissivity and atmospheric temperature profiles, cloud top heights can be calculated. Várnai and Marshak (2002) used gradients in the MODIS brightness temperature to further classify observed clouds into "illuminated" and "shadowy" clouds. Another spectral approach has been demonstrated amongst others by Fischer et al. (1991) and Zinner et al. (2018) using oxygen absorption features to estimate the travelled distance of the observed light through the atmosphere. Assuming most of the light gets reflected at or around the cloud surface, this information can be used to calculate the location of the clouds surface.

Other experiments (e.g. Beekmans et al., 2016; Crispel and Roberts, 2018; Romps and Öktem, 2018) use multiple ground based all-sky cameras and apply stereophotogrammetry techniques to georeference cloud fields. Due to the use of multiple cameras, it is possible to capture all images at the same time~~ a big advantage when observing moving and changing clouds~~, so that cloud evolution and motion does not affect the 3D reconstruction.

Spaceborne stereographic methods have been employed e.g. for the Multi-angle Imaging SpectroRadiometer (MISR) (Moroney et al., 2002) and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) (Seiz et al., 2006). MISR features 9 different viewing angles which are captured during 7 minutes of flight time. During the long time period of about one minute between two subsequent images the scene can change substantially. Clouds in particular are transported and deformed by wind, which adds extra complexity on stereographic retrievals. The method by Moroney et al. (2002) ~~adresses~~ addresses this problem by tracking clouds along all the perspectives and derivation of a coarse wind field at a resolution of about 70 km. ASTER comes with only two viewing angles but still takes about 64 s to complete one image pair. Consequently,

the method by Seiz et al. (2006) uses other sources of wind data (e.g. MISR or geostationary satellite data) to correct for cloud motion during the capturing period.

Parts of the introduction refer to the cloud surface, a term which comes with some amount of intuition, but is hard to define in precise terms. This difficulty arises because a cloud has no universally defined boundaries but rather changes gradually between lower and higher concentrations of hydrometeors. Yet, there are many uses for a defined cloud boundary. Horizontal cloud boundary surfaces are commonly denoted as cloud base height and cloud top height, which by their correspondence to the atmospheric temperature profile and subsequently thermal radiation largely effects the energy balance of clouds. Another such quantity, namely cloud fraction, is often used for example in atmospheric models to improve the parametrization of cloud-radiation interaction. Still, defining a cloud fraction requires to discriminate between areas of clouds and no clouds, introducing vertical cloud boundary surfaces. Stevens et al. (accepted) illustrate what Slingo and Slingo (1988) already said: cloud amount is "a notoriously difficult quantity to determine accurately from observations." Besides the difficulties in defining a thing like the cloud surface, it is a very useful tool to describe how clouds interact with radiation. This in turn allows us to do a little trick: we define the cloud's surface as the visible boundary of a cloud in 3D space. This may or may not correspond with gradients of microphysical properties, but clearly captures a boundary of interaction between clouds and radiation. This ensures that the chosen surface is relevant, both to improve microphysical retrievals which are based on radiation from a similar spectral region, as well as to use it in investigating cloud-radiation interaction. Additionally, by definition, the cloud surface is located where an image discriminates between cloud and no cloud, which is a perfect fit for the observation with a camera.

In this work, we present a stereographic method which uses 2D images taken from a moving aircraft at different times to find the georeferenced location ~~and structure of~~ of points located on the cloud surface facing the observer. This method neither depends on estimates of the atmospheric state nor does it depend on assumptions on the cloud shape. Contrasting to the spaceborne methods, our method only takes 1 s for one image pair. Due to the relatively low operating altitude of an aircraft compared to a satellite, the observation angle changes rapidly enough to use two successive images without the application of a wind correction method. As we employ ~~an~~ a 2D imager with a wide field of view, each cloud is captured from many different perspectives (up to about 100 different angles, depending on the distance between aircraft and cloud). Due to the high number of viewing angles, it is possible to derive geometry information of partly occluded clouds. Furthermore, this allows to simultaneously derive an estimate of the 3D wind field and use it to improve the retrieval result.

We demonstrate the application of our method to data obtained in the NARVAL-II and NAWDEX field campaigns (Stevens et al., accepted; Schäfler et al., 2018). In these field campaigns, the hyperspectral imaging system ~~specMACS~~ specMACS has been flown on the HALO aircraft (Ewald et al., 2016; Krautstrunk and Giez, 2012). The deployment of ~~specMACS~~specMACS, together with other active and passive instrumentation, aimed at a better understanding of cloud physics including water content, droplet growth, cloud distribution and geometry. The main component of the ~~specMACS~~ specMACS system are two hyperspectral line cameras. Depending on the particular measurement purpose, additional imagers are added. The hyperspectral imagers are operating in the wavelength range of $400 - 1000$ nm and $1000 - 2500$ nm at a spectral resolution of a few nanometers. Further details are described by Ewald et al. (2016). During the measurement campaigns discussed in this work, the two sensors ~~where~~ were looking in nadir perspective and have been accompanied by a 2D RGB imager with about twice

the spatial resolution and field of view. In this work, we focus on data from the 2D imager, because it allows observing the same cloud from different angles.

In ~~section~~Section 2 we briefly explain the measurement setup. Section 3 introduces the 3D reconstruction method and ~~section~~Section 4 presents a verification of our method. For geometric calibration of the camera we use a common approach of analyzing multiple images of a known chessboard pattern to resolve unknown parameters of an analytic distortion model. Nonetheless, as the geometry reconstruction method is very sensitive to calibration errors, we provide a short summary of our calibration process in ~~appendix~~Appendix A. We used the OpenCV library (Bradski, 2000) for important parts of this work. Details are listed in ~~appendix~~Appendix B.

2 Measurement Setup

During the NARVAL-II and NAWDEX measurement campaigns ~~specMACS was deployed on board~~ specMACS was deployed on-board the HALO aircraft. As opposed to Ewald et al. (2016), the cameras have been installed in a nadir looking perspective. The additional 2D imager (Basler acA2040-180kc camera + Kowa LM8HC objective) has been set up to provide a full field-of-view of approximately 70° with 2000 by 2000 pixels and data acquisition frequency at 1 Hz. To cope with the varying brightness during and between flights, the camera's internal exposure control system has been used.

Additionally, the WALES lidar system (Wirth et al., 2009), the HALO Microwave Package HAMP (Mech et al., 2014), the Spectral Modular Airborne Radiation measurement sysTem SMART (Wendisch et al., 2001) and an AVAPS dropsonde system (Hock and Franklin, 1999) was part of the campaign specific aircraft instrumentation. The WALES instrument is able to provide an accurate cloud top height and allows to directly validate our stereo method as described in section 4.

3 3D reconstruction

The goal of our 3D reconstruction method is to find georeferenced points which are part of a cloud surface at a specific time in an automated manner. Input data are geometrically calibrated images from a 2D camera fixed to the aircraft. As the aircraft flies, pictures taken at successive points in time show the same clouds from different perspectives. A schematic of this geometry is shown in ~~fig~~Fig. 1. The geometric calibration of the camera and the rigid mounting on the aircraft allows to associate each sensor pixel with a viewing direction in the aircraft's frame of reference. The orientation of the camera with respect to the aircraft's frame of reference has been determined by aligning images taken on multiple flights to landmarks also visible in satellite images. Using the aircraft's navigation system, all relevant distances and directions can be transformed into a geocentric reference frame in which most of the following calculations are performed. The reconstruction method contains several constants which are tuned to optimize its performance. Their values are listed in table 1.

In order to perform a stereo positioning, a location on a cloud must be identified in multiple successive images. ~~We start by identifying points which are likely to be detectable in another image. The surrounding of these points must show a contrast in both directions~~ A location outside of a cloud is invisible to the camera, as it contains clear air, which barely interacts

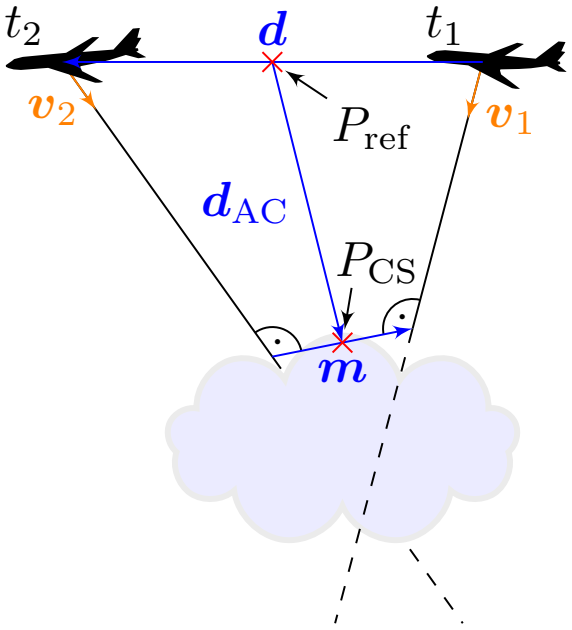

Figure 1. Schematic drawing of the stereographic geometry. Images of clouds are taken at two different times from a fast moving aircraft. Using aircraft location and viewing geometry, a point P_{CS} on the clouds surface can be calculated. Note that the drawing is not to scale: $\boldsymbol{d}$ is typically around 200 m, $\boldsymbol{d}_{\mathrm{AC}}$ in the order of 5 km and $\boldsymbol{m}$ is a description of mis-pointing and in the order of only a few meters.

with radiation in the observed spectral range. Locations enclosed by the cloud surface do not produce strong contrasts in the image, as the observed radiation is likely scattered again before reaching the sensor. Thus, a visible contrast on a cloud is very likely originating from a location on or close to the cloud surface as defined in the introduction. This method starts by identifying such contrasts. If such a contrast is only present in one direction of the image ~~plane. Ideally, these points should be evenly distributed over the image. For the identification of these points, we use the method by Shi and Tomasi (1994). Candidate~~ (basically, we observe a line), this pattern is not suitable for tracking to the next image due to the aperture problem (Wallach, 1935). We thus search one image for pixels of which the surroundings show a strong contrast in two independent directions. This corresponds to two large eigenvalues (λ_1 and λ_2) of the Hessian matrix of the image intensity. This approach has already been formulated by Shi and Tomasi (1994): interesting points are defined as points with $\min(\lambda_1, \lambda_2) > \lambda$ with λ being some threshold. We use a slightly different variant and interpret $\min(\lambda_1, \lambda_2)$ as a quality measure for each pixel. In order to obtain a more homogeneous distribution of tracking points over the image, candidate points are sorted by quality~~, points~~. Points which have better candidates at a distance of less then r_{min} are removed from the list and the remaining best N_{points} are taken. For these initial points, matches in the following image are sought using the optical flow algorithm described by Lucas and Kanade (1981). In particular, we use a pyramidal implementation of this algorithm as introduced by Bouguet (2000). If no match can be found, the point is rejected.

The locations of the two matching pixels define the viewing directions $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in ~~fig~~Fig. 1. The distance travelled by the aircraft between two images is indicated by $\boldsymbol{d}$. Under the assumption that the aircraft travels much faster than the observed clouds, an equation system for the position of the point on the cloud's surface P_{CS} can be found. In principle, P_{CS} is located at the intersection of the two viewing rays along $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, but as opposed to 2D space in 3D space there is not necessarily an intersection, especially in presence of inevitable alignment errors. We relax this condition by searching for the shortest distance between the viewing rays. The shortest distance between two lines can be found by introducing a line segment which is perpendicular to both lines. This is the mis-pointing vector $\boldsymbol{m}$. The point on the cloud's surface P_{CS} is now defined at the center of this vector. If for further processing a single point for the observer location is needed, the point P_{ref} at the center of both aircraft locations is used.

This way, many points potentially located on a cloud's surface are found. Still, these points contain a number of false correspondences between two images. During turbulent parts of the flight, errors in synchronization between aircraft navigation system and camera will lead to errors in calculated viewing directions. To reject these errors, a set of filtering criteria is applied (the threshold values can be found in ~~tab~~Tab. 1). Based on features of a single P_{CS}, the following points are removed:

- P_{CS} position is behind the camera or below ground
- absolute mis-pointing $|\boldsymbol{m}| > m_{\mathrm{abs}}$
- relative mis-pointing $|\boldsymbol{m}|/|\boldsymbol{d}_{\mathrm{AC}}| > m_{\mathrm{rel}}$

Figure 2 shows long tracks ~~starting at~~ corresponding to a location on the cloud surface. These tracks follow the relative cloud position through up to 30 captured images. The tracks are generated from image pairs by repeated tracking steps originating at the t_2 pixel position of the previous image pair. Using these tracks, additional physics based filtering criteria can be defined.

Each of these tracks contains many P_{CS} points which should all describe the same part of the cloud. As clouds move with the wind, the points P_{CS} do not necessarily have to refer to the same geocentric location~~. In fact, figure ?? shows similar tracks to figure 2, but now these tracks consist of the corresponding cloud surface points~~, but should be transported with the local cloud motion. For successfully tracked points, it can indeed be observed that the displacement of the P_{CS} points in a 3D geocentric coordinate system ~~. These tracks do not jump randomly around some center as they would do if they are only influenced~~ roughly follows a preferred direction instead of jumping around randomly, which would be expected if the apparent movement would just be caused by measurement errors. ~~Instead, they roughly follow some preferred direction~~The arrows in Figure 2 show the average movement of the P_{CS} of each track, reprojected into camera coordinates.

For the observation period (up to 30 s) it is assumed that the wind moves parts of a cloud on almost straight lines at a relatively constant velocity (which may be different for different parts of the cloud). Then, sets of P_{CS} can be filtered for unphysical movements. The filtering criteria are

- *velocity jumps*: the fraction of maximum to median velocity of a track must be less than v_{jump}
- *count*: the number of calculated P_{CS} in a track must be above a given minimum N_{min}

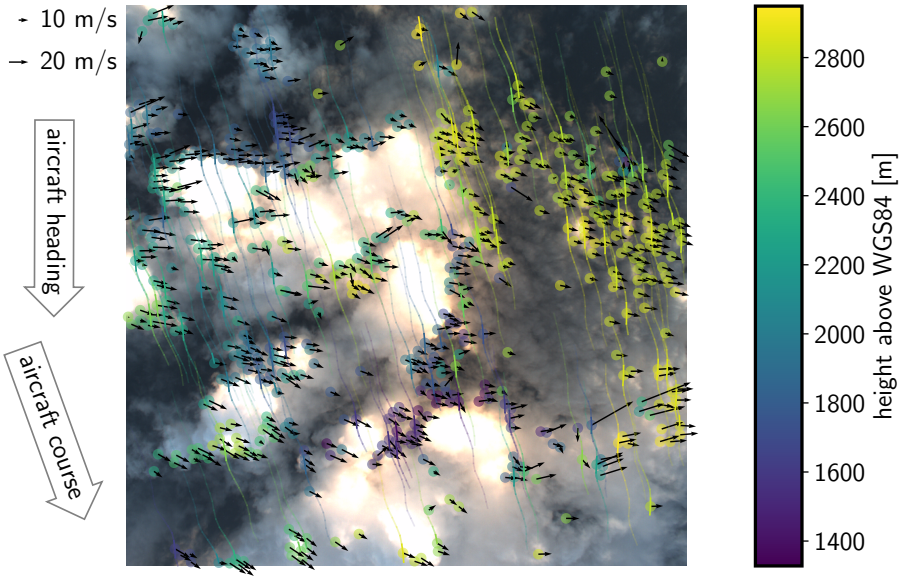

Figure 2. Image point tracking, every line in this image represents a cloud feature which has been tracked along up to 30 images. The images used have been taken on the NAWDEX flight RF07 (2016-10-06 09:32:15 UTC, location indicated in Figure 7) in an interval of 1 s. ~~The color~~ Transparency of the tracks indicates ~~if~~ time difference to the ~~given thresholds have been met as follows: green ≙ OK; yellow ≙ mis-pointing exceeded; pink ≙ velocity jumps too large; blue ≙ not enough~~ image. Color indicates retrieved height above WGS84, revealing that the larger clouds on the left belong to a lower layer than the thin clouds on the right. The arrows indicate estimated cloud movement. Due to the wind speed at the aircraft location, its course differs significantly from the heading and the tracks are tilted accordingly. The number of points ~~; orange ≙ distance uncertainty too large~~shown has been reduced to include at maximum 1 point per 20 px radius in the image. Tracks are only shown for every 5th point.

~~This figure shows a closeup of the retrieved data from the scene in figure 2. The view is not from the camera perspective, but sideways, with a slight downwards tilt. In dark red, P_{CS} of a track are connected to a line. The arrows show a local average of the resulting point movement vectors.~~

– *distance uncertainty*: the distance $\boldsymbol{d}_{AC}$ between aircraft and cloud may not vary more than d_{abs} or the relative distance variation with respect to the average distance of a track must be less than d_{rel}

During measurements close to the equator, typical during the NARVAL-II campaign, the sun is frequently located close to the zenith. In this case, specular reflection of the sunlight at the sea surface produces bright spots, known as sunglint and illustrated in ~~figure~~Figure 3. Due to waves on the ocean surface, these regions of the image also produce strong contrasts. It turns out that such contrasts are preferred by the Shi and Tomasi algorithm for feature selection, but are useless in order to estimate the cloud surface geometry. To prevent the algorithm from tracking these points, the image area in which bright sunglint is to be

Table 1. Filter thresholds

name	value
N_{points}	1000
r_{min}	5 px
m_{abs}	20 m
m_{rel}	1.5×10^{-3}
v_{jump}	3
N_{min}	5
d_{abs}	250 m
d_{rel}	7×10^{-2}

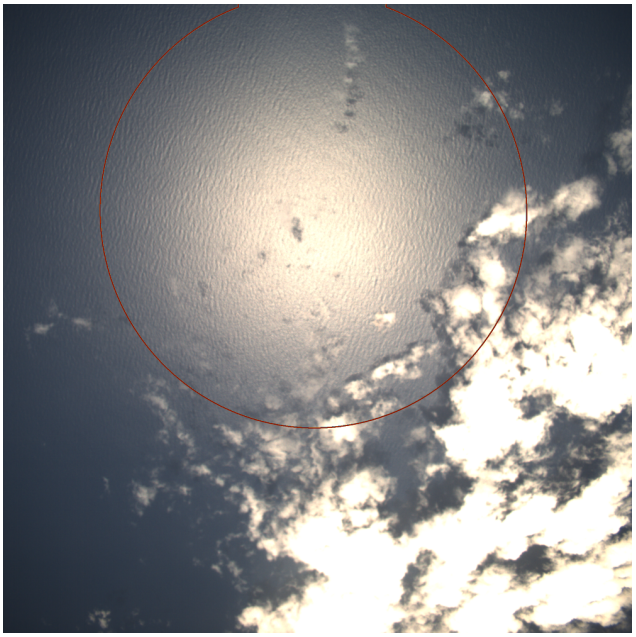

Figure 3. At low latitudes, close to the local noon as on the NARVAL-II flight RF07 (2016-08-19 15:06:13 UTC), the specular reflection of the sun on the ocean surface (sunglint) produces bright spots and high contrasts on the waves tails. While the bright spots can visually hide clouds, the contrasts create useless initial tracking points. The latter are mitigated by calculating the region of a potential sunglint (shown as red contour) and masking that region before the images are processed.

expected is estimated using the bidirectional reflectance distribution function (BRDF) by Cox and Munk (1954) included in the libRadtran package (Mayer and Kylling, 2005; Emde et al., 2016). The resulting area (indicated by a red line in ~~fig~~Fig. 3) is masked out of all images before any tracking is performed. Masking out such a large area from the camera image seems to

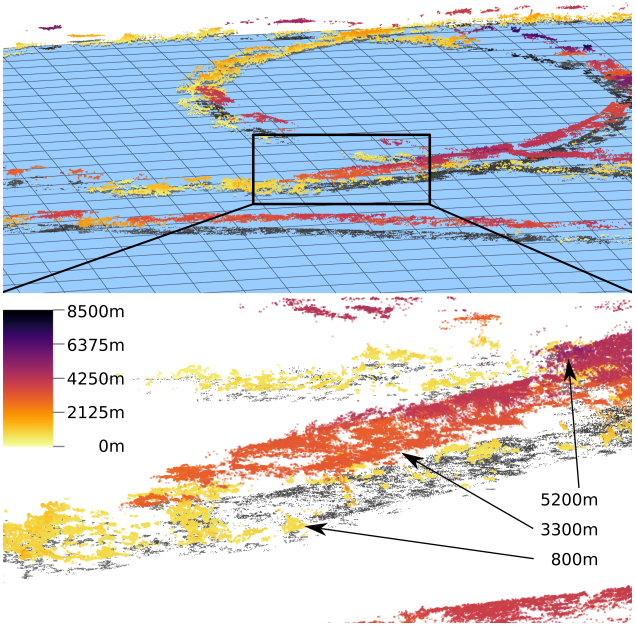

Figure 4. The collection of all $\bar{P}_{\mathrm{CS}}$ form a point cloud. Here, a scene from the second half of the NARVAL-II flight RF07 is shown. The colors indicate the ~~points~~ point's height above the WGS84 reference ellipsoid (indicated as blue surface). Below, ~~which also highlights~~ a part of the ~~ability to detect multiple~~ scene is shown magnified, displaying two main cloud layers: one at ~~once~~about 800 m in yellow and the other at about 3200 m in orange. On the right, a small patch of even higher clouds is visible at 5200 m. The gray dots are a projection of the points onto the surface to improve visual perception.

be a wasteful approach. In fact~~it is acceptable, because~~ , this is acceptable: due to the large viewing angle of the camera, all masked-out clouds are almost certainly visible at a different time in another part of the image. Therefore, these clouds can still be tracked ~~in the presence of~~ using parts of the sensor which are not affected by sunglint, even if a large part of the sensor is obstructed by sunglint.

After ~~all~~ filtering, a final mean cloud surface point $\bar{P}_{\mathrm{CS}}$ is derived from each track as the centroid of all contributing cloud surface points. The collection of all $\bar{P}_{\mathrm{CS}}$ form a point cloud in a Cartesian 3D reference coordinate frame which is defined relative to a point on the earth's surface ~~. Such a point cloud is shown in figure 4~~(Figure 4). This point cloud can ~~then be used as a starting point for further processing, including usage as reference points to verify~~ be used on its own, serve as a reference for other distance measurement techniques ~~as well as a starting point for~~ (e.g. oxygen absorption methods (Zinner et al., 2018), deriving distances by a method according to Barker et al. (2011)) or allow for a 3D surface reconstruction.

4 Verification

3.1 Lidar comparison

Comparison of cloud top height, measured with the WALES lidar and the stereo method. The most prominent outliers, present in the region of high lidar CTH and low stereo height can be attributed to thin, mostly transparent cirrus layers and cumulus clouds below. While the lidar detects the ice clouds, the stereo method retrieves the height of the cumulus layer below.

Thin cirrus layer above cumulus cloud field during NAWDEX RF10 (2016-10-13 10:32:10 UTC). Scenes like this are the typical cause for large outliers in the lidar comparison. The lidar detects the signal reflected by the cirrus layer while the stereo method uses the contrast of the lower cumulus clouds.

Cloud top height information derived from the WALES lidar (Wirth et al., 2009) is used to verify the accuracy of the described method. While the stereo method provides P_{CS} at arbitrary positions in space, the lidar data is defined on a fixed grid ("curtain") beneath the aircraft. To match lidar measurements to related stereo data points, we collect all stereo points which are horizontally close to a lidar measurement. This can be accomplished by defining a vertical cylinder around the lidar beam with 150 m radius. Every stereo derived point which falls into this cylinder with a time difference of less than 10 s is considered as stereo point related to the lidar measurement. As the (almost) nadir pointing lidar is observing cloud top heights only, we use the highest stereo point inside the collection cylinder. Figure 6 compares the measured distance between aircraft and clouds from the WALES lidar and the stereo method. Generally, there is a good agreement between both methods. Because of the way data is collected for this comparison, the automated comparison over many whole flights still suffers from the possibility that lidar and stereo method derived heights are related to different cloud layers in a multi-layer situation. Due to its high sensitivity, A precise camera calibration (relative viewing angles in the lidar method strongly favors the top height of the highest cloud while the stereo method is able to look at lower cloud layers along a slanted path. Thus, we expect that false correspondences typically relate larger lidar heights to lower stereo heights. The most prominent instance of this effect is order of 0.01°) is crucial to this method, which can be achieved by the region of WALES CTH $>$ 6000 m and low stereo height in fig. 6. This region can be attributed almost exclusively to parts of the flight in which high thin cirrus clouds are found above a low layer of cumulus clouds, like the scene shown in figure **??**. Additionally, as the stereo method relies on contrasts in the images, it works best at cloud edges which are often lower than cloud tops. The median bias is approximately 126 m for all compared flights with lower stereo heights as expectedcalibration process as described in Appendix A. A permanent time-synchronization between the aircraft position sensors and the cameras, accurate at the order of tens of milliseconds is indispensable as well. It should be noted that this does involve time stamping each individual image to cope with inter-frame jitter as well as disabling any image stabilization inside the camera. As this involves generating data which is only available during the measurement, this must be considered prior to the system deployment. For the system described in this work, we used the network time protocol (Mills et al., 2010) with an update interval of 5 min.

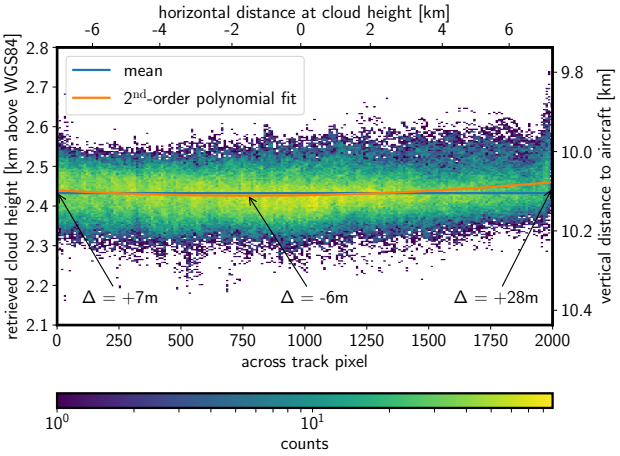

Figure 5. In the time from 09:01:25 to 09:09:00 UTC during NAWDEX flight RF11 on 2016 Oct 14, a stratiform cloud deck has been observed. The parabolic fit shows that a small systematic variation can be found beneath the noise (which is due to small scale cloud height variations and measurement uncertainties). Compared to the overall dimensions of the observed cloud (≈ 14 km) and the uncertainty of the method, these variations are small. It may still be noted, that data from the edges of the sensor (≈ 50 px on each side) should be taken with care.

4 Verification

4.1 Across track stability and signal spread

Errors in the sensor calibration could lead to systematic errors in the retrieved cloud height with respect to lateral horizontal distance relative to the aircraft (perpendicular to flight track). In order to assess these errors, data from a stratiform cloud deck observed between 09:01:25 and 09:09:00 UTC during NAWDEX flight RF11 on 2016 Oct 14 has been sorted by the average across track pixel position. While the cloud deck features a lot of small scale variation, it is expected to be almost horizontal on average. Note that as the orientation of the camera with respect to the aircraft has been determined independently using landmarks, deviations from the assumption of a horizontal cloud deck should be visible in the corresponding data and are counted as additional retrieval uncertainty in this analysis. During the investigated time frame, 260360 data points have been collected using the stereo method. The vertical standard deviation of all points is 47.3 m, which includes small scale cloud height variation and measurement error. Figure 5 shows a 2D histogram of all collected data points. From visual inspection of the histogram, apart from about 50 px at the ~~sensors~~ sensor's borders, no significant trend can be observed. To further investigate the errors, a 2nd-order polynomial has been fitted to the retrieved heights. This polynomial is chosen to cover the most likely effect of sensor mis-alignment which should contribute to a linear term and distortions in the optical path which should contribute to a quadratic term. The difference between left and right side of the sensor of 21 m corresponds to less than $0.1°$ of absolute camera misalignment and the curvature of the fit is also small compared to the overall dimensions of the observed clouds.

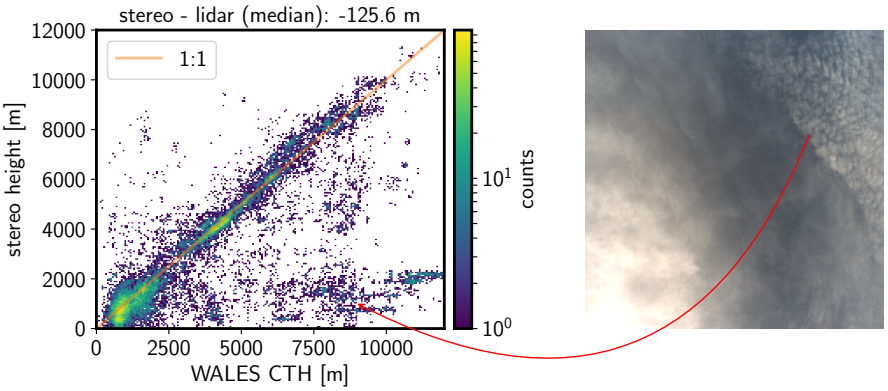

Figure 6. Comparison of cloud top height, measured with the WALES lidar and the stereo method. The most prominent outliers, present in the region of high lidar CTH and low stereo height can be attributed to thin, mostly transparent cirrus layers and cumulus clouds below, illustrated by a scene from NAWDEX RF10 (2016-10-13 10:32:10 UTC). While the lidar detects the ice clouds, the stereo method retrieves the height of the cumulus layer below.

4.2 Lidar comparison

Cloud top height information derived from the WALES lidar (Wirth et al., 2009) is used to verify the bias of the described method. While the stereo method provides P_{CS} at arbitrary positions in space, the lidar data is defined on a fixed grid ("curtain") beneath the aircraft. To match lidar measurements to related stereo data points, we collect all stereo points which are horizontally close to a lidar measurement. This can be accomplished by defining a vertical cylinder around the lidar beam with 150 m radius. Every stereo derived point which falls into this cylinder with a time difference of less than 10 s is considered as stereo point related to the lidar measurement. As the (almost) nadir pointing lidar is observing cloud top heights only, we use the highest stereo point inside the collection cylinder. The size of the cylinder is rather arbitrary but the particular choice has reasons: the aircraft moves at a speed of approximately 200 m/s and the data of the lidar system is available at 1 Hz and averaged over this period. Any comparison between both systems should therefore be in the order of 200 m horizontal resolution. Furthermore, data derived from the stereo method is only available where the method is confident that it worked. Thus not every lidar data point has a corresponding stereo data point. Increasing the size of the cylinder increases the count of data pairs, but also increases false correspondences. The general picture however remains unchanged.

Figure 6 compares the measured cloud top height from the WALES lidar and the stereo method, visually showing a good agreement. However its quantification in an automated manner and without manual (potentially biased) filtering proves to be difficult. Part of this difficulty is due to the cloud fraction problem, which is explained by (Stevens et al., accepted), basically stating that different measurement methods or resolutions will always detect different clouds. This is also indicated in Figure 6 on the right: the stereo method detects the lower cumulus cloud layer due to larger contrasts while the lidar observes the higher cirrus layer, leading to wrong cloud height correspondences while both methods are supposedly correct. Filtering the data for

high lidar cloud top height and low stereo height, reveals that the lower right part of the comparison can be attributed almost exclusively to similar scenes. Further comparison difficulties arise from collecting corresponding stereo points out of a volume which might in fact include multiple (small) clouds. Considering all these sources of inconsistency, only a very conservative estimate of the deviation of lidar and stereo values can be derived from this unfiltered comparison. The median bias between lidar and stereo method is approximately 126 m for all compared flights, indicating lower heights for the stereo method. As the lidar detects cloud top heights with high sensitivity and the stereo method relies on image contrast which is predominantly present at cloud sides, this direction is expected.

Further manual filtering indicates that the real median offset is likely in the order of 50 m to 80 m, however this cannot be shown reliably. Quantifying the spread between lidar and stereo method yields no meaningful results for the same reasons.

4.3 Wind data comparison

An important criteria that we use to identify reliable tracking points is based on the assumption that the observed movement of the points can be explained by a smooth transport due to a background wind field. The thresholds for this test are very tolerant, so the requirements for the accuracy of the retrieved wind field are rather low. However, a clear positive correlation between the observed point motion and the actual background wind would underpin this assumption substantially. In order to do so, we compare the stereo wind against an ECMWF reanalysis in a layer in which many stereo points have been found.

In the following, the displacement vectors of every track have been binned in ~~1 min bins~~ time intervals of 1 minute along the flight track and 200 m bin in the vertical. To reduce the amount of outliers, bins with less than 100 entries have been discarded. Inside the bins, the upper and lower 20% of the wind vectors (counted by magnitude) have been dropped. All remaining data has been averaged to produce one mean wind vector per bin. In ~~figure~~Figure 7 the horizontal component of these vectors is compared to ECMWF reanalysis data at about 2000 m above ground with horizontal sampling of 0.5°. The comparison shows overall good agreement according to our goal to consider a quantity in the stereo matching process which roughly behaves like the wind. The general features of wind direction and magnitude are captured. Deviations may originate from multiple sources including the time difference between reanalysis and measurement, representativity errors and uncertainties of the measurement principle. These results corroborate the assumption that the observed point motion is related to the background wind and filtering criteria based on this assumption can be applied.

5 Conclusions

The 3D cloud geometry reconstruction method described in this work is able to produce an accurate set of reference points on the observed surface of clouds. This has been verified by comparison to nadir pointing active remote sensing~~in the nadir perspective~~. Using data from the observation of a stratiform cloud field, we could verify that no significant systematic errors are introduced by looking in off-nadir directions. Even for sunglint conditions cloud top heights can be derived: as clouds move through the image while the sunglint stays relatively stable, we can choose to observe clouds when they are in unobstructed

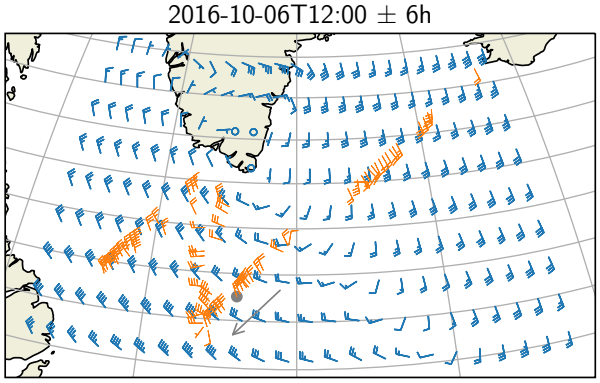

Figure 7. Horizontal wind at about 2000 m above ground. Comparison between ECMWF reanalysis (blue) and stereo derived wind (orange). Comparing grid points with co-located stereo data, the mean horizontal wind magnitude is 15.1 m/s in ECMWF and 13.4 m/s in the stereo dataset. This amounts to a difference of 1.7 ± 4.5 m/s in magnitude and $6.0 \pm 33°$ in direction. The shown deviations are standard deviations over all grid points with co-located data. The gray dot and arrow mark the location and flight course corresponding to Figure 2.

parts of the sensor. Because of the wide field-of-view of the sensor, there are always viewing directions to each cloud which are not affected by the sunglint.

As a visible contrast suited for point matching is a central ~~requirment~~ requirement of the method, it is able to provide ~~position~~ positional information at many but not ~~every point~~ all points of a cloud. Especially flat cloud tops can show very little contrast and are hard to analyse using our method. In future, we will integrate other position datasets like the distance measurement technique using O_2A-absorption as described by Zinner et al. (2018) which is expected to work best in these situations. In combining multiple datasets, the low bias and angular variability of the stereo method can even help to improve uncertainties of other methods.

While the wind information derived as part of the stereo method constituted a byproduct of this work the results look promising. After some further investigations about its quality and possibly additional filtering, this product might further add valuable information to the campaign dataset.

During the development of this method, it became clear that a precise camera calibration (relative viewing angles in the order of $0.01°$) is crucial to this method. A permanent time-synchronization between the aircraft position sensors and the cameras, accurate at the order of tens of milliseconds is indispensable as well. It should be noted that this does involve time stamping each individual image to cope with inter-frame jitter as well as disabling any image stabilization inside the camera. For upcoming measurement campaigns, improvements may be achieved by optimizing the automatic exposure of the camera for bright cloud surfaces ~~in stead~~ instead of relying on the built-in exposure system. Furthermore, it would be useful to re-investigate the proposed method with a camera system operating in the near-infrared which would most likely profit from higher image contrasts due to lower Rayleigh scattering in this spectral region.

Data availability. specMACS data is available at https://macsserver.physik.uni-muenchen.de after requesting a personal account.

Appendix A: Geometric Camera Calibration

As the distance between aircraft and observed clouds is typically much larger than the flight distance between two images, the 3D reconstruction method relies on precise measurements of camera viewing angles. To allow analysis as presented in this paper, frame rates of about $1\ \mathrm{Hz}$ are required. At that frame rate, a change in distance between cloud and aircraft of $100\ \mathrm{m}$ at a distance of $10\ \mathrm{km}$ results in approximately $0.01°$ difference of the relative viewing angle or about $1/3\ \mathrm{px}$. Consequently, achieving accuracies in the order of $100\ \mathrm{m}$ or below requires both, to average over many measurements in order to get sub-pixel accuracy and to remove any systematic error in the geometric calibration to less than $1/3\ \mathrm{px}$. This is only possible if distortions in the cameras optical path can be understood and corrected.

We use methods provided by the OpenCV library (Bradski, 2000) to perform the geometric camera calibration, our notation is chosen accordingly. Geometric camera calibration is done by defining a parameterized model which describes how points in world coordinates are projected onto the image plane including all distortions along the optical path. Generally, such a model includes extrinsic parameters which describe the location and rotation of the camera in world space and intrinsic parameters which describe processes inside ~~of~~ the camera's optical path. Extrinsic parameters can differ between each captured image while intrinsic parameters are constant as long as the optical path of the camera is not modified. After evaluation of various options for the camera model, we decided to use the following:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \boldsymbol{t} \tag{A1}$$

$$x' = x/z \tag{A2}$$

$$y' = y/z \tag{A3}$$

Where X, Y, Z are the world coordinates of the observed object, R and $\boldsymbol{t}$ the rotation and translation from world coordinates in camera centric coordinates and x, y, z are the object location in camera coordinates. x' and y' are the projection of the object points onto a plane at unit distance in front of the camera. The distortion induced by the lenses and the window in front of the camera is accounted for by adjusting x' and y' to x'' and y'':

$$r^2 = x'^2 + y'^2 \tag{A4}$$

$$x'' = x'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + s_1 r^2 + s_2 r^4 \tag{A5}$$

$$y'' = y'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + s_3 r^2 + s_4 r^4 \tag{A6}$$

Here k_1 to k_3 describe radial lens distortion and s_1 to s_4 add a small directed component according to the thin prism model. During evaluation of other options provided by OpenCV, no significant improvement of the calibration result was found using more parameters. Finally, the pixel coordinates can be calculated by a linear transformation (which is often called "camera

matrix"):

$$u = f_x x'' + c_x \tag{A7}$$

$$v = f_y y'' + c_y \tag{A8}$$

Here, f_x and f_y describe the focal lengths and c_x and c_y describe the principal point of the optical system. In this model, there are 6 extrinsic parameters (rotation matrix R and displacement vector $\boldsymbol{t}$) and 11 intrinsic parameters ($k_1 ... k_3, s_1 ... s_4, f_x, f_y, c_x, c_y$).

We use the well-known chessboard calibration method to calibrate this model which is based on Zhang (2000). The basic idea is to relate a known arrangement of points in 3D world space to their corresponding locations on the 2D image plane by a model as described above and solve for the parameters by fitting it to a set of sample images. The internal corners of a rectangular chessboard provide a good set of such points as they are defined at intersections of easily and automatically recognizable straight lines. Furthermore, the intersection of two lines can be determined to sub-pixel accuracy which improves the calibration performance substantially. While the extrinsic parameters have to be fitted independently for every image, the intrinsic parameters must be the same for each image and can be determined reliably if enough sample images covering the whole sensor are considered.

To evaluate the success of the calibration, the reprojection error can be used as a first quality measure. The reprojection error is defined as the difference of the calculated pixel position using the calibrated projection model and the measured pixel position on the image sensor. To calculate the pixel position, the extrinsic parameters have to be known, thus the images which have been used for calibration are used to calculate the reprojection error as well. This makes this test susceptible to falsely return good results due to an ~~over fitted~~ overffitted model. We use many more images (of which each provide multiple constraints to the fit) than parameters to counter this issue and have validated the stereo method (which includes the calibration) against other sensors to ensure that the calibration is indeed of good quality. Nonetheless, a high reprojection error would indicate a problem in the chosen camera model.

We have taken 62 images of a 9 by 6 squares chessboard pattern with $65\ \mathrm{mm}$ by $65\ \mathrm{mm}$ square size on an aluminium composite panel using the system assembled in aircraft configuration. The images have been taken such that the chessboard corners are spread over the whole sensor area, Figure A1 shows the pixel locations of all captured chessboard corners. After previous experiments with calibration targets made from paper and cardboard, it became clear that small ripples which inevitably appeared on the cardboard targets render the reprojection errors unusably high. Using the rigid aluminium composite material for the calibration target let the average reprojection error drop by an order of magnitude to a very low value of approximately $0.15\ \mathrm{pixels}$, which should be enough to reduce systematic errors across the camera to less than $50\ \mathrm{m}$. The per pixel reprojection error is shown in Figure A2 for every chessboard corner captured.

The calibration used for this work has been performed using gray scale versions of the captured chessboard images. To assess effects of spectral aberrations, the procedure has been repeated separately for each channel. A comparison shows that observed viewing angle differences vary up to the order of $1 - 2‰$ when switching between different calibration data. For the example used in the beginning of this appendix, this translates into cloud height differences of about $10\ \mathrm{m}$, which is considerably lower than the total errors achievable by this calibration method. Note that besides the same order of magnitude, these effects are

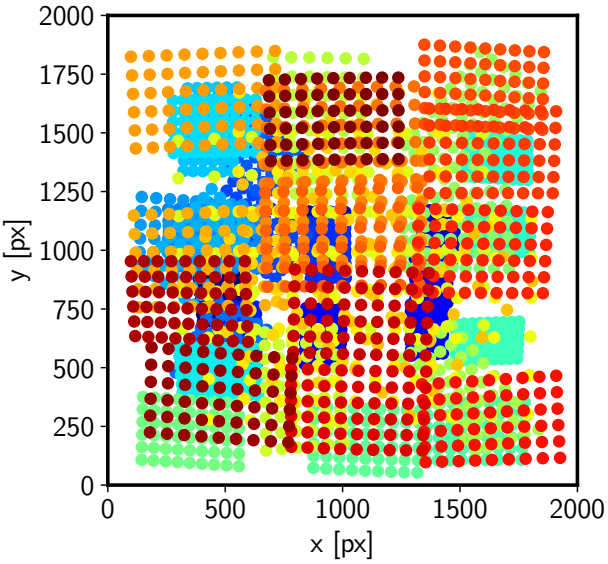

Figure A1. Locations of each chessboard corner during calibration on the image sensor plane.

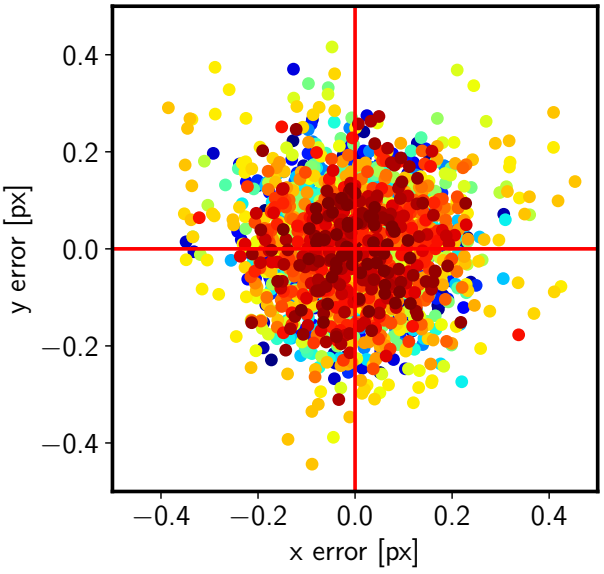

Figure A2. Reprojection error: Difference between the actual position of the chessboard corners and the calculated positions of the chessboard corners after applying the camera calibration. The average reprojection error is about 0.15 pixels. Note that the actual chessboard corner locations can be found far into sub-pixel accuracy by following the lines along the edges of the squares.

not able to explain the curvature in the analysis of Section 4.1. Due to effectively using fewer pixels when doing the camera calibration procedure on a single channel image, the reprojection error is increased accordingly. For these reasons, and in order to facilitate data handling, only one set of calibration data is used. Effects of spectral aberrations within one color channel have not been assessed, but are assumed to be smaller than effects between color channels.

Appendix B: OpenCV usage

The image processing of this work has been done with help of the OpenCV library (Bradski, 2000). The most important functions used are:

- `calibrateCamera` to calculate the camera calibration coefficients from a set of chessboard images
- `goodFeaturesToTrack` to find pixels which are most likely good candidates to be identified in the following image
- `calcOpticalFlowPyrLK` to find the corresponding pixel in the following image

Author contributions. T. Kölling was in charge of the presented measurements, developed the stereo reconstruction method and prepared the manuscript. T. Zinner and B. Mayer prepared the field campaigns, provided valuable input during the development of the method and contributed to the final version of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by the DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) through the Project 264269520 "Neue Sichtweisen auf die Aerosol-Wolken-Strahlungs-Wechselwirkung mittels polarimetrischer und hyper-spektraler Messungen". This work was supported by the Max Planck Society, and the DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) through the HALO Priority Program SPP 1294 “Atmospheric and Earth System Research with the Research Aircraft HALO (High Altitude and Long Range Research Aircraft)”. The authors thank Hans Grob for his help during the measurements.

References

Barker, H. W., Jerg, M. P., Wehr, T., Kato, S., Donovan, D. P., and Hogan, R. J.: A 3D cloud-construction algorithm for the EarthCARE satellite mission, Quarterly Journal of the Royal Meteorological Society, 137, 1042–1058, https://doi.org/10.1002/qj.824, http://dx.doi.org/10.1002/qj.824, 2011.

Beekmans, C., Schneider, J., Läbe, T., Lennefer, M., Stachniss, C., and Simmer, C.: Cloud photogrammetry with dense stereo for fisheye cameras, Atmospheric Chemistry and Physics, 16, 14 231–14 248, https://doi.org/10.5194/acp-16-14231-2016, https://www.atmos-chem-phys.net/16/14231/2016/, 2016.

Bouguet, J.-y.: Pyramidal implementation of the Lucas Kanade feature tracker, Intel Corporation, Microprocessor Research Labs, 2000.

Bradski, G.: The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2000.

Cox, C. and Munk, W.: Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter, J. Opt. Soc. Am., 44, 838–850, https://doi.org/10.1364/JOSA.44.000838, http://www.osapublishing.org/abstract.cfm?URI=josa-44-11-838, 1954.

Crispel, P. and Roberts, G.: All-sky photogrammetry techniques to georeference a cloud field, Atmospheric Measurement Techniques, 11, 593–609, https://doi.org/10.5194/amt-11-593-2018, https://www.atmos-meas-tech.net/11/593/2018/, 2018.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., and Bugliaro, L.: The libRadtran software package for radiative transfer calculations (version 2.0.1), Geoscientific Model Development, 9, 1647–1672, https://doi.org/10.5194/gmd-9-1647-2016, https://www.geosci-model-dev.net/9/1647/2016/, 2016.

Ewald, F.: Retrieval of vertical profiles of cloud droplet effective radius using solar reflectance from cloud sides, Ph.D. thesis, LMU München, http://nbn-resolving.de/urn:nbn:de:bvb:19-205322, 2016.

Ewald, F., Winkler, C., and Zinner, T.: Reconstruction of cloud geometry using a scanning cloud radar, Atmospheric Measurement Techniques, 8, 2491–2508, https://doi.org/10.5194/amt-8-2491-2015, https://www.atmos-meas-tech.net/8/2491/2015/, 2015.

Ewald, F., Kölling, T., Baumgartner, A., Zinner, T., and Mayer, B.: Design and characterization of specMACS, a multipurpose hyperspectral cloud and sky imager, Atmospheric Measurement Techniques, 9, 2015–2042, https://doi.org/10.5194/amt-9-2015-2016, http://www.atmos-meas-tech.net/9/2015/2016/, 2016.

Ewald, F., Zinner, T., Kölling, T., and Mayer, B.: Remote Sensing of Cloud Droplet Radius Profiles using solar reflectance from cloud sides. Part I: Retrieval development and characterization, Atmospheric Measurement Techniques Discussions, 2018, 1–35, https://doi.org/10.5194/amt-2018-234, https://www.atmos-meas-tech-discuss.net/amt-2018-234/, 2018.

Fielding, M. D., Chiu, J. C., Hogan, R. J., and Feingold, G.: A novel ensemble method for retrieving properties of warm cloud in 3-D using ground-based scanning radar and zenith radiances, Journal of Geophysical Research: Atmospheres, 119, 10–912, https://doi.org/10.1002/2014JD021742, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014JD021742, 2014.

Fischer, J., Cordes, W., Schmitz-Peiffer, A., Renger, W., and Mörl, P.: Detection of Cloud-Top Height from Backscattered Radiances within the Oxygen A Band. Part 2: Measurements, Journal of Applied Meteorology, 30, 1260–1267, https://doi.org/10.1175/1520-0450(1991)030<1260:DOCTHF>2.0.CO;2, https://doi.org/10.1175/1520-0450(1991)030<1260:DOCTHF>2.0.CO;2, 1991.

Hock, T. F. and Franklin, J. L.: The NCAR GPS Dropwindsonde, Bulletin of the American Meteorological Society, 80, 407–420, https://doi.org/10.1175/1520-0477(1999)080<0407:TNGD>2.0.CO;2, https://doi.org/10.1175/1520-0477(1999)080<0407:TNGD>2.0.CO;2, 1999.

Krautstrunk, M. and Giez, A.: The Transition From FALCON to HALO Era Airborne Atmospheric Research: Background – Methods – Trends, pp. 609–624, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-30183-4_37, https://doi.org/10.1007/978-3-642-30183-4_37, 2012.

Lucas, B. D. and Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision, in: Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'81, pp. 674–679, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1981.

Mayer, B. and Kylling, A.: Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use, Atmospheric Chemistry and Physics, 5, 1855–1877, https://doi.org/10.5194/acp-5-1855-2005, https://www.atmos-chem-phys.net/5/1855/2005/, 2005.

Mech, M., Orlandi, E., Crewell, S., Ament, F., Hirsch, L., Hagen, M., Peters, G., and Stevens, B.: HAMP – the microwave package on the High Altitude and LOng range research aircraft (HALO), Atmospheric Measurement Techniques, 7, 4539–4553, https://doi.org/10.5194/amt-7-4539-2014, https://www.atmos-meas-tech.net/7/4539/2014/, 2014.

Mills, D., Martin, J., Burbank, J., and Kasch, W.: Network Time Protocol Version 4: Protocol and Algorithms Specification, Tech. Rep. 5905, https://tools.ietf.org/html/rfc5905, 2010.

Moroney, C., Davies, R., and Muller, J. P.: Operational retrieval of cloud-top heights using MISR data, IEEE Transactions on Geoscience and Remote Sensing, 40, 1532–1540, https://doi.org/10.1109/TGRS.2002.801150, 2002.

Romps, D. M. and Öktem, R.: Observing clouds in 4D with multi-view stereo photogrammetry, Bulletin of the American Meteorological Society, 0, https://doi.org/10.1175/BAMS-D-18-0029.1, https://doi.org/10.1175/BAMS-D-18-0029.1, 2018.

Schäfler, A., Craig, G., Wernli, H., Arbogast, P., Doyle, J. D., McTaggart-Cowan, R., Methven, J., Rivière, G., Ament, F., Boettcher, M., Bramberger, M., Cazenave, Q., Cotton, R., Crewell, S., Delanoë, J., Dörnbrack, A., Ehrlich, A., Ewald, F., Fix, A., Grams, C. M., Gray, S. L., Grob, H., Groß, S., Hagen, M., Harvey, B., Hirsch, L., Jacob, M., Kölling, T., Konow, H., Lemmerz, C., Lux, O., Magnusson, L., Mayer, B., Mech, M., Moore, R., Pelon, J., Quinting, J., Rahm, S., Rapp, M., Rautenhaus, M., Reitebuch, O., Reynolds, C. A., Sodemann, H., Spengler, T., Vaughan, G., Wendisch, M., Wirth, M., Witschas, B., Wolf, K., and Zinner, T.: The North Atlantic Waveguide and Downstream Impact Experiment, Bulletin of the American Meteorological Society, 99, 1607–1637, https://doi.org/10.1175/BAMS-D-17-0003.1, https://doi.org/10.1175/BAMS-D-17-0003.1, 2018.

Seiz, G., Davies, R., and Grün, A.: Stereo cloud-top height retrieval with ASTER and MISR, International Journal of Remote Sensing, 27, 1839–1853, https://doi.org/10.1080/01431160500380703, https://doi.org/10.1080/01431160500380703, 2006.

Shi, J. and Tomasi, C.: Good features to track, in: 1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 593–600, https://doi.org/10.1109/CVPR.1994.323794, 1994.

Slingo, A. and Slingo, J. M.: The response of a general circulation model to cloud longwave radiative forcing. I: Introduction and initial experiments, Quarterly Journal of the Royal Meteorological Society, 114, 1027–1062, https://doi.org/10.1002/qj.49711448209, http://dx.doi.org/10.1002/qj.49711448209, 1988.

Stevens, B., Ament, F., Bony, S., Crewell, S., Gross, S., Hirsch, L., Mayer, B., Wendisch, M., Wirth, M., Bakan, S., Brück, H.-M., Ehrlich, A., Ewald, F., Farrell, D., Forde, M., Gödde, F., Grob, H., Hagen, M., Hansen, A., Jacob, M., Jäkel, E., Jansen, F., Klepp, C., Klingebiel, M., Kölling, T., Konow, H., Mech, M., Peters, G., Rapp, M., Wing, A., and Wolf, K.: A high-altitude long-range aircraft configured as a cloud observatory - the NARVAL expeditions, Bulletin of the American Meteorological Society, accepted.

Strabala, K. I., Ackerman, S. A., and Menzel, W. P.: Cloud Properties inferred from 8 - 12-µm Data, Journal of Applied Meteorology, 33, 212–229, https://doi.org/10.1175/1520-0450(1994)033<0212:CPIFD>2.0.CO;2, https://doi.org/10.1175/1520-0450(1994)033<0212:CPIFD>2.0.CO;2, 1994.

Várnai, T. and Marshak, A.: Observations of Three-Dimensional Radiative Effects that Influence MODIS Cloud Optical Thickness Retrievals, Journal of the Atmospheric Sciences, 59, 1607–1618, https://doi.org/10.1175/1520-0469(2002)059<1607:OOTDRE>2.0.CO;2, https://doi.org/10.1175/1520-0469(2002)059<1607:OOTDRE>2.0.CO;2, 2002.

Várnai, T. and Marshak, A.: A method for analyzing how various parts of clouds influence each other's brightness, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2003JD003561, http://dx.doi.org/10.1029/2003JD003561, 4706, 2003.

Wallach, H.: Über visuell wahrgenommene Bewegungsrichtung, Psychologische Forschung, 20, 325–380, https://doi.org/10.1007/BF02409790, https://doi.org/10.1007/BF02409790, 1935.

Wendisch, M., Müller, D., Schell, D., and Heintzenberg, J.: An Airborne Spectral Albedometer with Active Horizontal Stabilization, Journal of Atmospheric and Oceanic Technology, 18, 1856–1866, https://doi.org/10.1175/1520-0426(2001)018<1856:AASAWA>2.0.CO;2, https://doi.org/10.1175/1520-0426(2001)018<1856:AASAWA>2.0.CO;2, 2001.

Wirth, M., Fix, A., Mahnke, P., Schwarzer, H., Schrandt, F., and Ehret, G.: The airborne multi-wavelength water vapor differential absorption lidar WALES: system design and performance, Applied Physics B, 96, 201, https://doi.org/10.1007/s00340-009-3365-7, https://doi.org/10.1007/s00340-009-3365-7, 2009.

Zhang, Z.: A Flexible New Technique for Camera Calibration, vol. 22, pp. 1330–1334, https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/, 2000.

Zinner, T. and Mayer, B.: Remote sensing of stratocumulus clouds: Uncertainties and biases due to inhomogeneity, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006955, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006955, 2006.

Zinner, T., Schwarz, U., Kölling, T., Ewald, F., Jäkel, E., Mayer, B., and Wendisch, M.: Cloud geometry from oxygen-A band observations through an aircraft side window, Atmospheric Measurement Techniques Discussions, 2018, 1–20, https://doi.org/10.5194/amt-2018-220, https://www.atmos-meas-tech-discuss.net/amt-2018-220/, 2018.