Response to Anonymous Referee #2 (Minor comments, **The major comments will be addressed in a separate file.**)

We would like to thank Anonymous Referee #2 for his / her critical comments helping us to improve our paper.

The referee's original comments are written in italics, while our response is intended to the right.

The title could be made more specific, indicating that the method has been developed specifically for mountain gorges/valleys..

=> We agree. The title will be changed to "Automatic cloud top height determination **in mountainous areas** using a cost-effective time-lapse camera system". The beginning of the abstract will be changed to "A new method for the determination of cloud top heights ~~from~~ **using** the footage of a time-lapse camera, **that is placed above a frequently occurring cloud layer in a mountain valley**, is presented."

Although the method is mostly automated, "empirically derived thresholds" have been mentioned six or seven times throughout the manuscript. These thresholds probably need significant refining for each different mountain valley; so perhaps "fully automated" could be replaced with a weaker formulation.

We agree. Appearances of "fully automated" will be reformulated.

p. 2786, l. 19 – 21 will be changed to "Although it [The method by Bendix et al., 2008] still needs human interference, it shows the potential for further automation. The aim of this paper is to develop and validate a cost-effective~~, fully automated~~ method for zCT and zCB determination from camera footage in a cloud forest area of Taiwan **with a much higher degree of automation**."

p. 2805, l. 23 will be changed to "**Besides the necessary adaptations of the algorithm to different camera locations,** the data analysis is fully automated"

Page 2784, line 24: "If zDEM is equal to or below zCB..." Shouldn't it be above zCB?

=> Thanks for the correction. The sentence will be changed to: "If zCB is equal to or **above** zDEM "

Page 2785, line 8: "...that is far away from perfect." Far from perfect sounds better.

=> We agree. We will change that.

Page 2785, line 11: "...cloud radar devices." Cloud radars would suffice; "device" is not necessary. Also in line 13.

=> We agree. We will change that.

Page 2785, lines 15-17, first sentence of the paragraph: This sentence is convoluted and difficult to follow. Could you rephrase and simplify? E.g. The lack of cloud height data in remote regions also impedes the design of ground fog detection schemes/networks.

> => We will change the sentence using you suggestion. "**The lack of cloud height data in remote regions also impedes the design of ground fog detection schemes that shall be used to map fog frequencies**.
>
> This makes also changes in p. 2786, l. 7 – 10 necessary:
> "The necessity of manual evaluation of the photos makes the approach hardly applicable for any comprehensive statistical investigation **such as the analysis of the intradiurnal variability of cloud heights**. ~~the mentioned preliminary studies for the design of ground fog mapping techniques~~."

Page 2785, line 17, next sentence: "...inter-diurnal dynamic of cloud heights..." Do you mean diurnal variability of cloud heights?

> That was a typo. "**Intra**diurnal dynamic" was meant.

Page 2785, line 25: MODIS, spell out the acronym.

> We agree. We will do that.

Page 2785, line 27: "...the mentioned problems..." Aforementioned might be better.

> We agree and will use "aforementioned".

Page 2786, lines 7-10: The sentence starting with "The necessity of manual..." is confusing. You could just remove the part "for any comprehensive statistical investigation as the mentioned preliminary studies" to make it sound better.

> => We agree. The sentence will be reformulated to "The necessity of manual evaluation of the photos makes the approach hardly applicable for any comprehensive statistical investigation **such as the analysis of the intradiurnal variability of cloud heights**."

Page 2787, section 2.1: The first and the third sentences sort of say the same thing, that is, the Taroko Gorge is well suited for testing cloud top height retrieval techniques. Maybe you can combine the two sentences.

> => You are right. We will simply remove the second half of the first sentence: "The Taroko Gorge located in Eastern Taiwan is famous for a frequently (almost daily) occurring sea of clouds, which can be observed from higher terrain ~~and is therefore well suited for cloud top height determination~~. Since cloud forest is present on the slopes of the gorge, the frequency of ground fog will be mapped using satellite data in a future study. Therefore the area is ideally suited to test a technique that can be used to design and validate methods for ground fog retrieval from satellite data."

Page 2787, line 11: "...can be considered as suited for the installation.." as suited is not necessary.

=> Maybe we are a little pedantic here, but "some places can be considered for the installation" would mean, that some places STILL can be considered for the installation. This is not the case since the camera is already installed. In principle, however, these places are still suited.

Please let us know if you do not agree.

Page 2788, line 1: "usually form in different heights..." at different heights sounds better.

=> We agree and will use your suggestion

Page 2788, line 8: "...in a distance of about 200 m." at a distance of … sounds better.

=> We agree and will use your suggestion

Page 2788, lines 22-23: Spell out acronyms ASTER and METI.

=> We will use "Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)" instead of "ASTER".

We would, however, prefer not to spell out the acronym METI as we also would also have to spell out NASA then (at least if we try not to be western-centric) and the whole sentence would become a mess. METI and NASA are anyhow only mentioned because it is required according to the conditions of ASTER GDEM usage (https://www.jspacesystems.or.jp/ersdac/GDEM/E/3.html) and not to give additional information to the reader.

Please let us know if you do not agree.

Page 2789, line 23: "...in a temporal resolution of..." at a temporal resolution of

=> We agree and will use your suggestion

Page 2789, lines 24-28: The first two sentences ("Scenes are discarded if...") are awkward, what with the parenthesis and the word "therefore". Could you rephrase? Also, it should be "an undifferentiated image".

=> We agree that the overuse of parentheses makes the sentence hardly understandable and will rephrase it to:

"Scenes are discarded if the location of the main cam itself is cloud immersed. Cloud

immersion results in an undifferentiated mean image. Therefore the coefficient of variation of the brightness of all image pixels that are below the horizon of the reprojected DEM can be used to detect that condition. If it is below 0.8 for each color channel the camera is cloud immersed. Also scenes that are too dark to be analyzed (mean brightness of all pixels below the horizon is below 25) are excluded from further analysis. For all other scenes the image analysis algorithm delineated in Sect. 4.2.3 is used to detect cloud tops in the main cam imagery."

Page 2791: The discussion on adjusting the camera based on the virtual and true horizons is a bit difficult to follow. Could you describe verbally too what the fit quantity in equation (1) represents?

=> We agree. Considering also the remarks of reviewer #1 we will rephrase the paragraph as follows:

"To adjust the camera the fit between the virtual horizon and the horizon in the mean image is calculated. As the horizon can (if it is not obstructed by clouds or mist) be seen as an edge in the camera footage, a simple edge detection is applied: For each pixel of the mean image the sum of the euclidean distances in RGB space to its neighboring pixels is calculated and written to a new image further referred to as edge image. For each pixel pi of the scene that touches the virtual horizon the sum si of all edge image pixel values epxy in a 10 × 10 pixels window surrounding pi (with x, y ranging from −5, −5 to 5, 5) weighted by the reciprocal value of their distance in pixels to pi is calculated. The fit, which quantifies how well the virtual horizon matches with edges in the mean image, is then calculated as the average of all N values of si.

[formula 1]

The more horizon pixels of the mean image (= high edge image values) are near to the virtual horizon, the higher is the fit value."

Page 2792, the first bullet (-) point: You exclude terrain farther than 10 km from the main camera. Was this horizontal distance threshold derived by an error analysis, in order to limit the errors in cloud top height retrievals? [...]

=> The threshold of 10 km corresponds to a natural segmentation of the visible terrain: Every visible pixel in a distance of more than ~9.9 km is also at least 13 km away. In such a distance our method definitely causes problems caused by the atmospheric influence to the signal as well as the fact that each 5 m height interval has a height of below 1 pixel in the camera's image causing additional uncertainties.

We will add this information to p. 2792, l.5 - 8 as follows:

"These areas include sky, foreground objects that are included in a manually created JPEG image, terrain with a distance of more than 10 kilometers to the main cam **(This corresponds to a natural segmentation of the camera's view shed: Every visible pixel with a distance of more than 10 km to the camera is also at least 13 km away. In such a distance the accuracy with which zCT can be determined is certainly insufficient for most purposes)** and areas that are in a vertical buffer of 10 pixels around distinctive edges in the terrain."

[…] Also, you filter out slopes (edges) whose steepness exceeds 400 m and 200 m per pixel, respectively, for the northern and southern parts of the gorge.

> *=> Slopes are not filtered out by a certain steepness threshold. We agree that our description is insufficient here. We will rephrase p. 2792, l. 8-11 as follows:*
>
> "Edges in the terrain are defined as areas where **the difference between the distance from a pixel of the reprojected DEM to the main cam and the distance of its upper neighboring pixel to the main cam** exceeds a threshold of 400 (northern slope) or 200 (southern slope) meters [...]"

If the goal is to eliminate areas where cloud top height retrievals are sensitive to a small error in the camera parameters, why the different thresholds for the two different parts of the gorge? Why do you use a more conservative value for the southern parts of the gorge?

> => As we mentioned, both thresholds were empirically derived. This is necessary as certain difference in the distance to the camera between two neighbored pixel does not necessarily entail a big difference in the height between these two pixels (dependent on the topography of the terrain and the viewing angle). Therefore a lower threshold for the northern slope would detect terrain features as edges that do not lower the zCT retrieval quality. That would only decrease the amount of pixels which are used in the analysis. We will account for this by adding
>
> "(both thresholds are empirically determined **and are dependent on the slopes' topography and the viewing angle**)."
>
> to the brackets in p. 2792, l. 11.
>
> We could also have incorporated the distance in height between the two pixels that are compared to each other and will do this when adapting method to another location, but for the moment the approach with two separate thresholds works fine.

In addition, the word "adulterate" seems out of place here; you could say instead that near steep edges the retrieval of zCT is very sensitive to small errors in camera parameters, or something to that effect.

> => We agree that "adulterate" is out of place. P. 2792, l. 11-14 will be changed as follows:
>
> **Near those edges the presented method could provide fundamentally incorrect values of zCT** since a small misfit of the virtual camera parameters could drastically influence the height that is attached to a main cam pixel in these areas.

Page 2792, the third bullet (-) point: You mention k-means clustering, which requires setting the number of clusters k, which is an input parameter. How many clusters do you use for fine segmentation? The number of optimal clusters presumably depends on the local topography and, thus, is different for each valley. This seems to be another parameter that needs to be tuned on a case-by-case basis (cf. Fully-automated method).

> => For each slope 400 initial centroids are used. You are right, this number may need to be

adapted for other locations. We will ad this information as follows to p. 2793, l. 3-5

"A fine segmentation **(using 400 initial centroids for each slope)** as shown in Fig. 6 as well as a coarse segmentation with a drastically decreased number of classes (6 per slope) are performed to obtain both segment images. **Those numbers have proven to be suited for the location in the Taroko Gorge and may need to be adapted for different locations.**"

Page 2793, lines 14-15: "...that is degenerated by the factor of 4 in..." ...that is decreased by a factor of 4 in...

=> We agree. Thanks for mentioning.

Page 2794, lines 7-10: The sentence "Despite the fact that..." is not clear. Do you mean in addition to/besides the fact that..., rather than despite?

=> Yes, despite is wrong. We will use "In addition to" instead. Thanks for noticing.

Page 2794, line 17: "...the distance depended influence..." ...the distance dependent influence...

=> Yes. We will change that.

Page 2797, line 8: "...for each fine segments..." ...for each fine segment...

=> Yes. We will change that.

Page 2797, lines 17-18: "The weighted RMSD times 1.2 has proven to be a good size for the height interval in which zCT segment is determined." How have you determined this factor of 1.2? By visual validation? Could you elaborate?

=> The factor has actually been determined by visual validation. We will add this information:

"Visual evaluation has shown that the weighted RMSD times 1.2 is a reasonable size for the height interval in which $z_{CT\ segment}$ is determined."

Page 2797, line 25: "...on a fine segments base..." ...on a fine segment(s) basis...

=> We agree and will change that.

Page 2798, lines 2-3: "Since clouds are overall brighter than non cloud covered terrain ..." What about snow- or ice-covered terrain? Could you comment?

=> Even in winter snow is rare in Taiwan and limited to the highest peaks. Every pixel used for zCT determination in the Taroko Gorge is below an altitude of 3000 m, so that won't be a problem. For the adaptation of the method to other locations, however, this might be a problem. We will address that issue in the final paper as follows:

p. 2798, l. 2 – 3: "Since clouds are overall brighter than non cloud **(nor snow)** covered

terrain (if analyzed for each fine segment separately) [...]"

p. 2805, l. 9 - 11 "Since a valid cloud height determination depends on clouds touching the terrain, the approach does only work for selected locations, ideally with frequently occurring sea of cloud conditions. **Also the occurrence of snow, which is unlikely for the area used for zCT determination in this study, might cause problems as the presented method relies on differences in the brightness between clouds and terrain.**"

Page 2798, last line: Spell out CSV, presumably comma-separated values, or simply leave out the file type, which is irrelevant to the discussion.

=> You are right. We will write "output file" instead of CSV file

Page 2800, lines 21-23: "Thus the validation results show to what extent the presented method is suited to determine zCT with a precision of 50 m." More precisely, you have validated the method at one particular height, that of the validation camera (2377 m). In Figure 12 you show retrieved cloud top heights ranging from 2000 to 2600 m. Strictly speaking, the method has not been validated over the full range of possible cloud top heights, although it might be reasonable to assume that your error estimates obtained at the validation camera height hold over the entire range.

=> To be 100 % precise this is also not correct although our formulation wasn't correct either. Even if the cloud top is way above the validation cam one could still tell from the validation cam footage (that would be cloudy) that it is above. In this case, however, the resolution would be drastically lowered. We will rephrase the sentence to

"Thus the validation results show to what extent the presented method is suited to determine **whether zCT is above or below the height of the validation cam** with a precision of 50 m."

Page 2802, line 11: "...cannot be reasonably be derived from..." remove the first be

=> We will do that.

Page 2802, section 4.3.2: This section validates cloud detection, rather than cloud top height retrievals. For clarity, you could change the section heading to "Visual validation of (automated) cloud detection" and contrast it to section 4.3.1, which could be "Validation of retrieved cloud top heights using the validation camera".

=> We do mostly agree. Since out method detects cloud top we would, however, prefer *"Visual validation of cloud **top** detection" for section 4.3.2*

Page 2802, last sentence of section 4.3.1: "After this exclusion the validation results can be interpreted as the answer to the question if the top height of clouds is derived correctly if they have been detected." This sentence is awkward, could you rephrase it? E.g. After this exclusion, the validation results indicate the fraction of detected cloud tops with correctly retrieved heights. Or something to that effect.

=> We agree and will use your suggestion: “After this exclusion, the validation results indicate the fraction of detected cloud tops with correctly retrieved heights”

Page 2804, line 19: “...HKD and the POD are quite high and the POD and FAR are low.” Shouldn’t it be POFD?

=> Yes, it should. Thanks for mentioning. We will correct that mistake.

Page 2804, lines 23-25: “The presence or absence of clouds can already be determined from satellite data with a high degree of certainty (Reuter et al., 2009).” I would disagree with such a definitive statement, although this point has little bearing on the merits of the presented algorithm. The GEWEX Cloud Assessment has revealed significant differences between the existing satellite cloud climatologies. For example, the global total cloud amount varies between 0.56 and 0.74, depending on satellite sensor. A more in-depth analysis is given by Stubenrauch et al. [2013].

=> You are right. Thick clouds as they usually form in the Taroko Gorge are easy to detect but for optical thin clouds this is not the case. To make our point clear without using that disputable statement we will change the paragraph as follows:

“A method that is designed to provide validation data for another method should, however, be as near to perfect as possible. For this reason the presented method should only be used to validate satellite derived cloud heights in scenes where there is no doubt about the presence of clouds. False positives and false negatives of the camera approach would be ignored in that way.

Page 2805, line 26: “...satellite derived cloud tops heights...” ...cloud top heights...

=> We will change that.

Fig. 2: The red lines are rather difficult to see, especially in the hard-copy version. Could you use a brighter color that gives more contrast (yellow, bright green, etc.)?

=> Since red is a really eye-catching color, we assume you are referring to black-and-white hard copies. To make the lines more outstanding in b/w we added an outline:

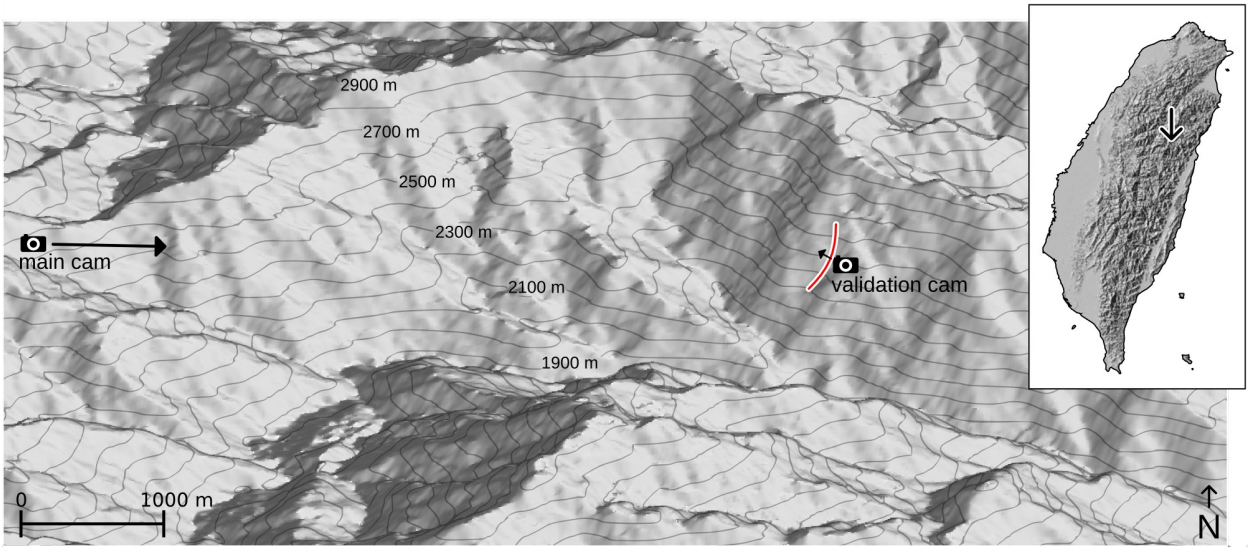

Fig. 6: Please specify in the caption what the reddish/greenish colors refer to (southern/northern slopes). What do color shades correspond to, distance classes? Also, "...fines segments..." should be "fine segments", I suppose. The masked out white areas are mainly steep slopes?

=> The color shades corresponds to fine segments, which are distance classes. The white area corresponds to every part of the image that is masked out (terrain edges, sky, foreground). To make things clearer we will change the caption to:

"Figure 6. Segmentation of the terrain into northern (greenish area) and southern (reddish area) slope as well as into fine segments (color shades). The white areas correspond to parts of the image that are masked out."

Fig. 7: The red lines are more visible here, but a brighter color might still be better.

=> We added an outline to make the red line more outstanding:

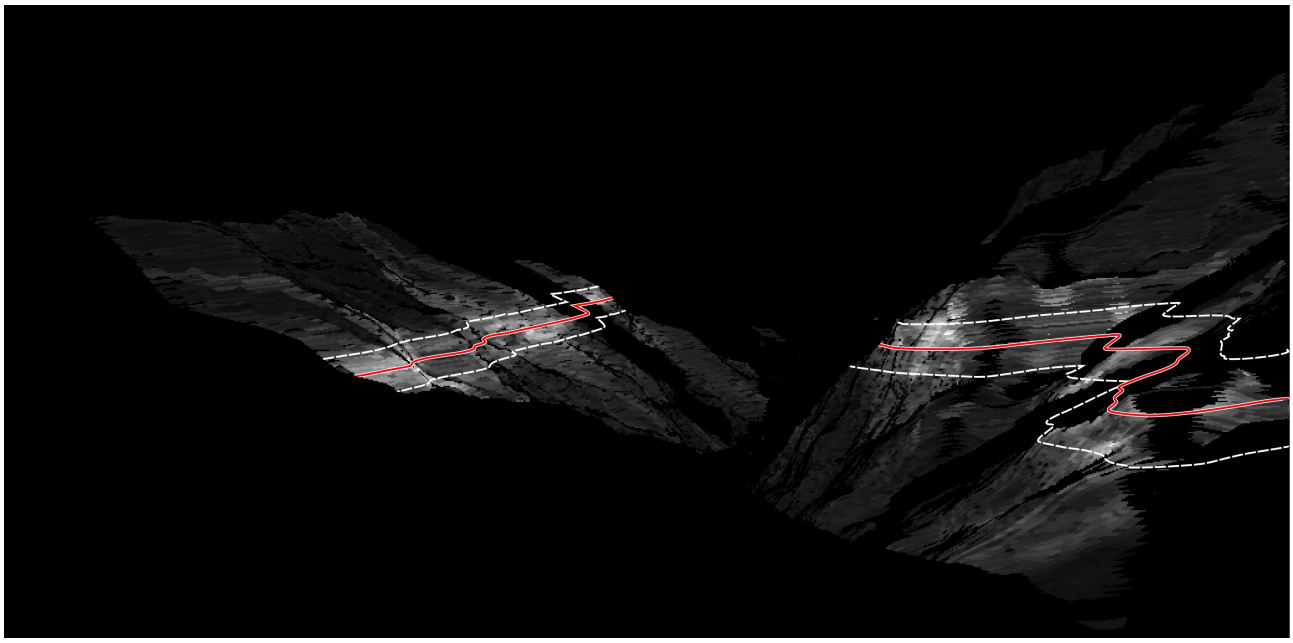

Fig. 11: A color palette would be better here. It’s rather difficult to read numerical cloud top height values from this gray scale plot. And “Cloud tops heights...” should be “Cloud top heights...”

=> “Cloud tops heights” will be changed to “Cloud Top Height”.

Regarding the color palette we respectfully disagree. RGB color schemes have several disadvantages for continuous data (cf. Especially 3. & 5. in http://blog.visual.ly/rainbow-color-scales/). We also could have used contour lines, but that would suggest a high precision that the interpolated data, of course, do not have. Illustrating the trend in zCT from the south to the north using a black-and-white color scheme, however, does not have to deal with these problems, is easy to understand and works well also in black-and-white hard-copies.

Please let us know if you do not accept this explanation.