Response to referee #2

Thank you for your thorough review. Here we explain how we have changed the manuscript in response to your comments and suggestions. We hope you agree they have improved the paper.

__Abstract__ This technique has only been demonstrated to be "fully automatic" on single arcs, which should be emphasised at appropriate points throughout this paper (including the abstract).

We have added the word “single” to the abstract and at other points in the paper.

Phrases like "more suitable" and "less consistent" should be replaced with precise statements using terms such as "bias" and "variance". This might help clarify why "it would be advantageous to apply both methods".

The last sentences of the abstract have been changed to:
“It is found that the new method measures the peak emission height regardless of the shape of the volume emission rate profile, unlike the other recent method. However, the new method is less suitable than the other method for analysis of very wide auroral arcs (>30 km) or for aurora in the magnetic zenith of one of the images.”

__Introduction__ Typo: 555.7 nm -> 557.7 nm

This has been corrected.

__Instrumentation__ "several all-sky cameras" <- how many exactly?

We have added “(between 5 and 8, depending on year)” to this sentence.

"narrow passband" <- 1 nm?

The camera filters have passbands of about 2 nm. This has been clarified in the text.

Are background (non-auroral) images collected?

Dark frames are collected for all cameras, but only the EMCCD cameras are equipped with filters to observe background emission, at 540 nm, 438 nm, and 620 nm. However, neither of the methods presented in this work require absolute intensity calibrated images, and normal background emission will not significantly distort images of auroral structures. We use unlit portions of the images (the four corners) to correct for varying dark current and detector bias voltage.

What are typical integration times?

We have added a sentence: “The exposure time of the ICCD images has usually been 1~s for green and panchromatic images and 2~s for blue and red images.”

__Previous work__ It is stated here that "The new method does not give information about the variation in height along an arc". Actually, it seems to provide estimated height profiles at multiple locations along the arc, but then averages away all this information. This may not be critical for synthetic data corresponding to a single height, but is likely to be important for real arcs. Even for synthetic data it seems likely that the set of profiles could be used, for example, to identify and remove outliers.

In most cases good test field-lines do not lie directly on the arc, but very close to it. By averaging the altitude profiles from all good field-lines (which will lay on both sides of the arc) we are able to obtain a representative average profile for the arc. It may indeed be possible to determine the height variation along an arc, but this should be done very carefully and on a case-study basis, which is beyond the scope of this paper. This paper focusses on applying the technique quickly and automatically.

It is not obvious that "The new method presented in this work measures the height of peak emission". Estimates do seem to be centered around the peak height. However, for the nearly symmetric height profile used here the mean and peak heights are effectively indistinguishable. It would be interesting to try a more skewed profile (or multiple peaks) in order to determine what this technique is responding to.

The paragraph in question describes previous work which examined the lower border of an auroral arc, and discusses the differences between measurements of the lower border of an auroral arc and measurements of its peak emission height. The last sentence emphasises that the disadvantages of measuring the lower border are not an issue with the new method we present. We have replaced "measures the height of peak emission" with "is designed to measure the height of peak emission", to avoid confusion.

The mean height of the profile is 118 km, which is significantly different to its peak height of 110 km. We have now tested the methods on 5 different profiles (see response to a similar comment below).

__Section 4.2__ The description of method#2 could be made more clear. I would suggest starting with the concept of mapping a very thin arc onto a single field line, followed by a discussion of how outcomes will differ for the correct choice versus other footprints. This would help motivate the use of tests 1-3 and provide a better foundation for understanding how/why this method can fail.

We believe this is what we have shown in figures 2 and 3 (now figures 4 and 2). Note that it is not an arc which is projected onto a field-line, but a 2d image projected onto a field-line (as a cut across the image). We have rearranged and edited the description of the method, to explain the concept of mapping an image onto a single field-line first, to emphasise why the three tests are needed, and to clarify the description.

It would also be informative to show how the method responds to increasingly thicker arcs, as this may be an important source of variance.

We have added a figure (figure 11) and the following text to the paper: "The width of an auroral arc has a substantial effect on the ambiguity between the vertical and horizontal dimensions within the 2-dimensional images of the arc. In this work the two methods have been tested using simulated arcs with widths ranging from 5 km to 35 km, as described in section 5. Figure 11 shows how the performance of a) method 1 and b) method 2 varies with arc width. Each panel shows seven histograms, one for each arc width tested. The shaded bins are coloured according to the number of image pairs (the scale on the right of the figure). It is clear from panel a) that method 1 performs consistently regardless of arc width up to the tested limit of 35 km. The histograms show an almost constant bias (estimated height - true height) of about -2 km. However, panel b) shows that the bias of method 2 does vary with arc width. In most cases the bias is about zero, but the variance and bias of the results increase with increasing arc width, particularly for arcs wider than 25 km. For arcs 35 km wide the bias of method 2 is +1.4 km (an overestimation of the height). For the same arcs the bias of method 1 is -2.4 km."

The next step would be to introduce the use of multiple randomly selected field lines. In that discussion you should consider using the word 'set' or 'selection' rather than 'grid' which implies a certain degree of regularity.

We have replaced the word "grid" with "set", and have edited the description of the method.

It would also be helpful to briefly discuss *why* to choose random over regular sampling. Using a regular grid (at least for the first pass) would allow pre-calculation of a large number of field line traces. This could significantly reduce computation time when analyzing large numbers of images.

We have added this sentence to the text: "The positions are random because it was found that if they conformed to a regular grid, the result could be biased to either lower or higher altitudes if the arc lay approximately parallel, but slightly away from, one row of the grid."

While it does seem reasonable to select field lines from "the brightest feature within the overlapping portions of the images" this could be a source of bias towards lower altitudes, and deserves some discussion.

This sentence was in fact slightly erroneous and has now been corrected – it now reads “the most significant bright feature within the overlapping portions of the images”, and the explanation of this selection has been expanded (see next comment below). The “most significant bright feature” means the structure (cluster) with the most pixels amongst the brightest 5% of pixels in the image. In principle the selection could still introduce a bias towards lower altitudes, but usually all auroral structures contain pixels in the brightest 5%. We have added the following paragraph to the manuscript:
“The clustering process used in method 2 to select a region of the image for analysis could bias the results, as it restricts the analysis to a bright structure within the image. It may be expected that the peak emission height of the aurora is related to its brightness (although this is not yet clear). However, this bias will only become apparent for images with aurora covering a very large portion of the image, for example diffuse aurora or a system of many parallel arcs. Under these conditions it is likely that the method would fail anyway, as the ambiguity between the horizontal and vertical dimensions inherent to a 2-dimensional image would be large. It is also unlikely that method 1 would produce a strong correlation between images in these cases.”

If the "clustering algorithm" is well known then provide a reference, otherwise provide a brief explanation.

The text has been changed to: “A set of magnetic field lines is generated across the most significant bright feature within the overlapping portions of the images, which is identified using a connectivity-based clustering algorithm. The brightest 5% of pixels within one of the images are grouped into clusters such that the euclidean distance between neighbouring pixels in the same cluster is no more than twice the shortest distance between any two of the bright pixels. The cluster containing the most bright pixels is selected to represent the most significant structure within the image.”

For test#3 the phrase "centres of mass" should probably be replaced with something like "luminosity height profile average".

This has been changed throughout the text.

Presumably the final step for figure#2 of "averaging together the altitude profiles for all [good] field lines" is used for all subsequent analysis and discussion. If so, I would strongly suggest that the authors explore the variance around that average. At the very least this should provide a useful estimate of noise or uncertainty. It could motivate the introduction of additional tests to reject profiles that are significant outliers.

We have added a new figure (figure 12) to show the relationship between the estimated height - true height for an arc and the standard deviation of heights over all good field-lines. This is described in a new paragraph in the results and discussion section:
“Method 2 involves taking an average profile over multiple field lines, each of which can provide an estimate of the peak emission height. The standard deviation of the height over these field lines is related to the accuracy of the method, and to the actual variation in height across the auroral structure. Figure 12 shows the relationship between this standard deviation and the difference between the estimated and true heights for the set of 9310 synthetic images described above as a 2-dimensional histogram. The shaded bins are coloured according to the scale on the right of the figure. The red dashed line highlights where the standard deviation is equal to the difference between estimated and true height. For the majority of test images this difference is less than the standard deviation (85% of images, histogram bins below the red dotted line), showing that the standard deviation across “good” test field lines can provide a useful estimate of the uncertainty of the method result. When examining real images, it is important to note that a large standard deviation could result from an auroral structure which does not have one constant peak emission height.”

We have also modified figure 12 (now figure 16) to include bars indicating the standard deviation over all good field lines.

It might also provide information about systematic height variation along the arc. Obviously there should be none for the synthetic data, but it would be reassuring to check.

As mentioned in response to a comment above, examining height variation along an arc is not as simple as it may appear, and is beyond the scope of this paper.

__Section 5__ The authors clearly state here that synthetic data are constructed using a single arc with midpoint located midway between the two sites. They do consider a range of different lengths, widths, angles, and "bendiness". It appears that exactly the same (nearly symmetric) height profile is used for all cases. Although the range of explored parameters is admirable, I can't help but wonder what the effect would be of moving arcs north/south or vertically, It would be interesting to explore what these changes would do to the bias and variance of methods 1&2. Note that we might expect systematic effects given the results obtained by Romick and Belon 1967a.

We have added figure 14 and the following paragraph to the paper:
While this study concentrates on analysing images of an arc centred at the midpoint between the Kilpisjärvi and Sodankylä stations, a short series of tests on arcs located to the north or south of this point have also been performed. The results are shown as a set of histograms in figure 14 for method 1 (panel a) and method 2 (panel b), with shaded bins coloured according to the number of image pairs (scales on the right). Seven points were tested between 67.47 N and 68.97 N, all at the same longitude of 23.59 E. The locations of the Sodankylä (SOD) and Kilpisjärvi (KIL) stations are marked with dashed lines. For each latitude tested a set of 1000 random synthetic image pairs was analysed using both methods, with arc properties (length, width, bendiness) in the same ranges as described in section 5. All of the synthetic arcs were approximately east-west aligned. The results of these tests show that method 1 performs consistently for all arc locations which have been tested. However, while method 2 is successful for arcs located approximately halfway between the two stations, it does not perform well for arcs located a few tenths of a degree to the south of either station. In these cases the arc is in the magnetic zenith as viewed from one of the stations, and therefore many auroral rays appear as dots or very short lines in the corresponding image. It is not possible to extract accurate emission profiles from test field lines close to the magnetic zenith, and so method 2 fails to provide a result or performs poorly in these cases. The results of method 2 appear to be biased for arcs located directly above the Kilpisjärvi station, but the reason for this is unclear. It is possible that the arcs are so close to the horizon of the Sodankylä images that the pixel resolution is too low to produce reliable altitude profiles from test field lines. It is not possible to analyse arcs located far to the north or south of both stations, as the separation of the stations is too large. The tests performed here are not exhaustive, and further analysis is required to fully assess the performance of the methods for extreme arc locations. However, in most cases images suitable for analysis will contain arcs located between the stations.

The effect of varying height profiles is also worth considering, as the one used for this study is nearly symmetric, with the mean very similar to the peak. Using a skewed height profile would test whether the analysis is actually responding to peak or mean brightness.

We have added a figure (figure 13) and the following paragraph to the paper:
To check that the methods respond to the peak emission height (rather than another property such as mean emission height) synthetic images of the same auroral arc were generated using five different volume emission rate profiles and were analysed using both methods. For each profile the footprints of the auroral rays used to generate the synthetic images were identical, and formed an east-west aligned arc 350 km long, 10 km wide and with a bendiness of 0.1. All of the profiles have a peak emission height of 110 km, but each has a different mean emission height. The five profiles are shown in figure 13. The results from the two methods are shown on the right of the figure as diamonds, coloured to match the height profiles to which they correspond. The same results are shown in the magnified panel (far right) for clarity, where the standard deviations over "good" test field lines used in method 2 (as discussed above) are plotted as error bars with a total length of twice the standard deviation. Both methods respond quite well to the peak emission height for all profiles, but method 2 is clearly less susceptible to variations in the shape of the profile and produces the best results. For all profiles the peak emission height is within the error estimate for method 2. The dark blue coloured profile was used in the generation of the 9310 synthetic image pairs described in section 5. For this profile method 1 underestimates the peak emission height, which is consistent with the results discussed earlier, but for most other profiles it produces an overestimate.

__Results__ Method#1 has lower variance and higher bias than method#2. If the bias is either constant or has a known dependence on geometry then it can be subtracted to produce a method#1a that has lower bias and lower variance than method#2. The authors need to do (at least) one of three things: 1) show that the bias in method#1 is so complicated as to be uncorrectable 2) reduce the variance in method#2 3) do not claim that method#2 is always superior to method#1 There are many applications where a constant 2km bias would be less troubling than a >5km uncertainty.

We did not intend to claim that method 2 is always superior to method 1, rather that the new method is likely to be most appropriate for a large statistical study. We have substantially changed the wording of the conclusions to emphasise this more clearly and to consider the additional analysis we have now performed.