Dear Anette

We hereby submit the revised manuscript (both raw and annotated) together with reply letters to the reviewers. To include almost all comment from the reviewers to improve the manuscript, particularly the readability to non-auroral scientists and further explanation to the relation to the machine learning method (Random Forest is also mentioned), we made extensive revision, particularly the in the introduction, entire section 2 and beginning of section 4, with 4 new figures (Figures 1-4) and one new table. Most of the old figures and tables are also remade (e.g., much less all-sky images and streamlined table format).

Sincerely
Yama

M. Yamauchi
Swedish Institute of Space Physics, Kiruna

Reply to Reviewer #1

Thank you for your encouraging and detailed comments, and for reminding that potential readers are much wider than the auroral observation community.

#comment 1#
The submitted Manuscript presents an approach for Aurora real-time detection based on hemispheric RGB camera images. In contrast to other approaches using deep learning, the introduced method uses spectra indices calculated from RGB image data.
Unfortunately, the authors assume exhaustive preknowledge on the detection / measurement of aurora that impedes an easy and exciting entry into the topic.

#Answer#
To make the content for wider range of readers, we made considerable revision with much more text (particularly §1 and §2), four new figures (ASC image with different category, Keogram, image showing N2 red line, and picture of the camera setting), and one new table (example of actual classification from an image). Also all ASC images in figures are re-made.

#comment 2:
The method is very complex and hard to understand due to missing figures that would definetely help to grasp the matter. Furthermore, it sounds like the approach is only applicable to one study area using one specific camera setup where the authors developed the method. It was not stated that the approach was tested / validated at some else location or using some other camera configuration.
The evaluation is based on manual observations that AFAIK could be very subjective.
Unfortunately it is not stated if the reference data is based on observations of several operateurs to have some quality control on the Groud Truth data.

#Answer#
For the validation, we added explanation in §2.5. The need for the "expert system" is also explained in the introduction. Unlike big telescope images that many people look at the same images, auroral judgement is quite primitive (so, only auroral scientist can do) and what we can calibrate is only cameras but not the judgement.

In the auroral observation, no classification method (including machine learning) can be applied to different cameras/location without changing parameters (for machine learning method, training set should be different between different cameras, and definition of training set is the subjective part). Since the limitation of the camera and the location (light pollution) problem are so common, we did not explain in detail but simply mention that we have version 0 for different camera (this is already used in a different team).

#comment 3:
Due to the mentioned points I recommend to resubmit the paper after extensive rework as the method itself sounds interesting.

#Answer#
We did an extensive revision as summarized above and below.

#comment PDF
Some more comments are given directly in the attached PDF.

--- abstract ---
#line 10 "(which means saturated or mixed with N2 red line at 670 nm)":
What is a N2 red line at 670 nm?
I expect the explanation in the following text but in the abstract it confuses the reader

#Answer#
We added Figure 4 (example of N2 line) and related explanation in §2.1.

#line 15 "were compared to the manual (eye-)identification of the auroral activity during the rest of the auroral season":
Is the number of camera-based and manual observations nearly the same? This info could be given already in the abstract.

#Answer#
We manually examined all images as stated in §2.5, and this is now mentioned in the introduction.

--- 1. Introduction ---
#line 31 "are being nowcasted":
Perhaps you could shortly explain at the beginning what the reader has to understand by the term "nowcast".

#Answer#
We added explanation "(e.g., webcasting the real-time data)"

#line 32 "keograms of":
Please explain the term "keogram" together with allsky image in one sentence and provide an example image. I could imagine that the interested reader is not familiar with these terms

#Answer#
We added Figure 2 and relevant explanation just for the keogram.

#line 40 "Helmholtz Centre": => Helmholtz - German Research Centre for Geosciences(GFZ)

#Answer#
We fixed.

#line 59 "each pixel of about 150000 active pixels has a 256x256x256 colour values":
This part of the sentence is difficult to understand. Is the statement that you have 8 Bit RGB images with a resolution of 15 MPix? If so, please keep it simple like this
=> "However, todays computers should easily deal with such data. I was wondering why special large computing abilities are required? "

#Answer#
We rephrased the phrase and removed "large" from all relevant places. We also mentioned the ability of the processor that we suppose is PC level.

#line 61 "we judge":
do you mean => "classify" with "judge"?

#Answer#
We changed to "classify"

--- 2. Algorithm and source data ---
#line 80 "detail is described":
"step is detailed?

#Answer#
We changed to "The classification scheme is detailed"

#line 85 "The ASC is located in a 30 cm heated dome at the roof of the optical laboratory in Kiruna at 68 degree north":
A picture of the study area (map + photo of installed camera) would be nice

#Answer#
Camera geometry and camera installations are added in Figures 2 and 3.

#line 92 "we do not have to define one-to-one relation between the colour and category as long as we can define the activity level properly":
but this requires a lot of manual work, doesnt it?

#Answer#
We need some manual work to apply to the other location, but such modification (mainly considering different artificial light) is not very tough. The reason is explained in §2.3, §2.4 and §4.

#Figure 1 (new Figure 5):
Please provide additionally a figure where the reader can see the selection of the patterns exemplary

#Answer#
We added new Figure 1 to show the example of the classification. We also added Table 1 as the introduction to this figure (Figure 5).

--- 2.1. Classification of each pixel ---
#line 98 "G values are":
Furthermore, consumer cameras do not caputure individual bands of wavelengths but implement colour filters like the common Bayer Pattern to separate the incoming light in RGB with double times green than red and blue.

#Answer#
We added a paragraph to explain how the green aurora (558 nm) is modified in the RGB system. The filter difference between cameras is simply mentioned as "camera difference".

#line 109 "that H criterion fluctuates depending on the auroral luminosity and on the effects of moon, cloud, and artificial light."
I imagine that varying exposure of the camera is a huge disadvantage regarding this point

#Answer#
The dynamic exposure is always the problem in identifying the aurora (bigger problem than difference in the JPEG compression scheme) and is the priority task in the future, as we mentioned in §4.1 (we added words for actual numbers of the subdivision). We also added the exposure time in all images such that one can visualize its effect, and Table 1 will help that this is actually feasible.

#Tables 1-3 (now Tables 2):
please explain the differencing of the arc, e.g. with images where strong green arc 1-6 could be seen

I didn't get the point how you defined the conditions, please provide more detailed information on this. I got that you calculated ratios and defined min/max values to be fullfilled to assign a pixel to a specific class but how you have defined the limits? empirically?

#Answer#
We revised the table format, and also added Table 1 to explain how we constructed the table (conditions). To enhance the explanation, we divided §2.1 into §2.1-§2.3.

#line 120 "or like the arc mixed with N2 red line (670 nm)"
more figures please that show N2, N1, ...

generally, you assume a lot of prior knowledge. A few more background information would be very helpful in order to understand your processing steps and to improve the reading flow. You have to consider that the wide readership of egu sphere may not be experts in aurora monitoring.

#Answer#
We added Figure 4 (example of N2 line) and related explanation in §2.1.

#Table 2 and Table 3 (now Tables 2):
Is there a "or" between the conditions to be fulfilled or a "and"? (Question is valid for the other ratio/condition tables too)

Table 2 and 3, please see my comments above. A few more information as well as images depicting the classified regions would be very helpful.

#Answer#
We added the explanation in the text "AND logic between columns, and OR logic between rows, and these OR-logic separated conditions are numbered by S1, S2, S3, so on. "

#Table 5 (new Table 4):
I assume that consumer cameras are slightly different in their color spaces, e.g. due to different sensor sensitvity, color filters, ... are the mentioned values valid for other camera setups too or only for Sony + Nikkor?

#Answer#
Different camera has different criterions, and we mentioned this limitation in the Introduction and §2. Idea of the fist step (1a)+(1b) is to get as similar values as possible for the pixel-base value, such that final "index" becomes similar with 1 digit accuracy. This task is much easier than aurora identification from raw image for different cameras.

--- 2.2. (new 2.4.) Calculation of the aurora ASC index ---
#line 162 "which means saturation with the N2 red lines)":
again, a few more images would be helpful

#Answer#
We added Figure 4 (example of N2 line) and related extensive explanation in §2.1.

#line 163 "0.1-5% ":
0.1 - 5% of what?

#Answer#
Occupancy ratio, and this is now stated. We even added the actual numbers in pixels.

#line 178 "pixels of the "strong aurora" exceeds 4900 pixels (about 3% of the image), we simply use the pixel data for the most luminous 4900 pixels";
very hard to imagine ... pictures / figures would help!

#Answer#
New Figure 1 would help (strong aurora in Figure 1 is 2.4%).

#line 193 "classification criterions are different at different location (city light and latitude)":
how to adapt your approach to other cities?

#Answer#
Our old statement "latitude" was misleading and removed here (instead, we added explanation in §4. The difference is only the artificial light, but that does not affect the final "activity level" judgement, unless the ASC is highly polluted by the city light with strong green color

#line 196 (end of §2.2): The entire approach sounds interesting but very complex and unique for one single location. At the beginning, the authors stated that deep learning approaches are also used but expose uncertainities and black boxes. But after reading the method I¬¥m wondering if the given approach with all the unique ratios outperforms the black box solution?

#Answer#
As long as the camera setting is the same, the criteria for aurora (particularly "strong" and "arc") are not vey much different between different location, and we revised the text to make sure this point. With rather stable the criterion for the critical classes (strong and arc), we found (at least ourselves) this "criteria" method is much easier to translate how we judged the aurora in the sky. In fact, version 0 (using HSL only) method is now applied by another team.

--- 2.3. (new 2.5.) Evaluation of the activity level from the aurora ASC index ---
#line 207 "images by eye (traditional method) .... somewhat subjective.":
how sensitive is the traditional approach with regards to the operator? Or is there a "4-eye" redundancy?

#Answer#
We added more words on this in §2.5. In addition, we actually examined the magnetometer data (this is now mentioned). We believe that our "subjective" judging is quite stable.

#Table 10:
This is hard to catch and to understand. What is the key message? Could table 10 (and maybe 9) be transfered to a nice diagram?

#Answer#
We revised the Table and relevant explanation.

--- 4. Discussion and future tasks ---
#line 295 "Adding a third category between two different phenomena a common method in practical classification of data when the criterion is not very much discrete (i.e., somewhat fizzy).":
Agreed but I miss the clear distinction between category I and II.

#Answer#
We added Figure 1 and enhanced the text in §2.2.

#line 314 "category. Therefore, one can examine pixel by pixel in any case of categorizing wrong.": and this is not possible using deep learning for image classification?

#Answer#
Judging the color-code (step 1a) is actually a possible part that may be replaced to the machine-learning method in the future. We now mentioned this possibility in various places.

#line 320 "just have to adjust the criterion for different ASCs":
"Just" sounds easy but looking at the given tables in the methodology part it seems like many many many indices must be reworked by hand and running new observations to generate these criterions to be compared. Perhaps I haves't got the simplicity point in this section. Please revise

#Answer#
We added many text of how to define the criteria (old §2.1 is divided into §2.1-2.3). With new explanation, we hope that the "improvement" does not sound as complicated as the old text.

--- 4.1. Correction using exposure time information and UT information (ver. 1.1) ---
#line 330 "exiftool"
or other libs

#Answer#
We fixed.

--- 4.2. N2 red (670 nm) line (ver. 1.2) ---
#line 337-346 "§4.2":
I like these short and precise subsections. Maybe this would help to improve the understanding of the Methodology section a lot.

#Answer#
We subdivided §2.1 into §2.1-2.3. Also we combined Tables 1-3 and Tables 4-6, with an improved table format.

--- 5. Summary and conclusion ---
#line 407 "150000 active pixels"
what is, btw, the image format? (width x height)

#Answer#
We now added information in Figure 1 (new) caption.

Reply to Reviewer #2 (same as by community comment Christian Kehl)

Thank you for your encouraging and detailed comments, and for reminding that potential readers are much wider than the auroral observation community.

 # preface#
That said, next to the line-by-line review attached, there are some general comments that need addressing in a potential revision.

#comment 1:
The manuscript makes the data appear as very complex, multi-dimensional and challenging. Sadly, this only appears so due to ambiguous writing. In the end, after careful review, it appears the images are rather small (150,00 pixel) 3-band images, which is hidden behind confusing writing and misleading illustrations (fig. 1). This needs fixing.

#Answer#
To make the content for wider range of readers, we made mahjr revision with much more text (particularly §1 and §2), four new figures (ASC image with different category, Keogram, image showing N2 red line, and picture of the camera setting), and one new table (example of actual classification from an image). Also all images in the figures are re-drawn. We also use the word of "6-bit color information" to avoid the misunderstanding. Accordingly, explanation around old Figure 1 (now Figure 5) is enhanced.

#comment 2:
The manuscript makes the developed method appear as intricate-yet-powerful, which is related to how the data is presented. In the end, the presented method is a multi-level, double-bound thresholding with 5 different thresholding levels (i.e. categroies). This method is indeed simple, but introduces a considerable amount of fine-tuned parameters due to the lack of in-depth prior image processing. This also makes the formulation ambiguous and the method very difficult to replicate for other observatories.

#Answer#
We added explanations toward possible application to the other places (§2). Latitude difference is not the problem and light contamination is not as bad as it sounds in the old text. We also made it clearer that the presented two step method is the first trial of "translation" of how auroral scientists actually judge "onset" of auroral activity in the sky: first evaluate the color information to judge if it is aurora or not (just using green color cannot distinguish diffuse aurora or cloud because the morphology is similar to each other, and this is why we need all R, G, and B values), and then evaluate the activity level from both intensity and area within the field-of-view.

#comment 3:
The manuscript requires more context information to make it accessible and comprehensible to researchers outside the expert aurora observation community. Special terms and phenomena such as the N2 red line or local arc breaking are neither explained in-text or referenced in literature. It is very hard to understand the message and research objectives of the manuscript for a common EGU

audience. There is still more than sufficient space available for some added references to give the reader information on where to find further details on the assumption baseline of the manuscript.

#Answer#
As answered above (#1) we added many figures and associated explanation for general readers in both introduction and §2.

#comment 4:
The manuscript lacks references to make some context information understandable and to make certain design decisions of the method easier to reason and comprehend.

#Answer#
Akasofu's textbook (1977) is the best reference for this. Instead of adding more reference, we added considerable amount of explanation as a short summery of this textbook.

#comment 5
The manuscript lacks a details discussion of the imaging parameters (camera resolution, quantisation, etc.). The authors use JPEG images with small resolutions, dynamic exposure times from the camera, hence automatic white balancing, and an ingrained 3-band 8-bit pixel quantisation. The paper lacks any in-depth discussion of the actual influence of those instrumentation parameters. The paper lacks any comparison and impact assessment of the image compression error from JPEG. Overall, the manuscript treats the actual imaging influences superficially.

#Answer#
I added the description of the camera setting and parameters in §2. We also added some explanation on the color change by moon and twilight, while no significant change in the compression scheme.

#comment 6:
The manuscript lacks a comparison with alternative, more up-to-date methods of image pattern recognition.

#Answer#
We added the text stating that there is no other method that quantifies the aurora activity (only eye identification is used in all auroral science paper).

#comment 7:
The manuscript needs considerable language revision, as minor and intermediate grammar mistakes are frequent.

#Answer#
We went through the language check.

#comment 8:
The actual presentation of the ASC images needs revision as the displayed metrics within the image are not indicative and certain features in the images require expert explanation. Furthermore, fewer-but-larger images would support reader comprehension.

#Answer#
All ASC images are enlarged and numbers of images are reduced. We also added Figure 1 and 4 to explain the auroral features.

#comment PDF

#Typos#
We fixed types

--- abstract ---
#line 14 "When Level 6 is detected, automatic alert E-mail is sent out to the registered addresses immediately":
Why is this relevant ? Lack of context information about the dangers of auroras. For most people, it's a beautiful colour spectacle in the sky - no need for alerts ...

#Answer#
We enhanced the explanation in both the abstract and introduction.

--- 1. Introduction ---
#lines 20-25:
style - very lengthy sentence - split - reader can loose context.

#Answer#
We divided and also rephrased.

#line 40 http://wdc.kugi.kyoto-u.ac.jp/ae_realtime/presentmonth/index.html
#line 41 https://isdc.gfz-potsdam.de/kp-index/
#line 66 https://tromsoe-ai.cei.uec.ac.jp/#/
#line 88 https://www.irf.se/sv/observatorieverksamhet/firmamentkamera/
#line 186 https://www.irf.se/alis/allsky/nowcast/
Please no extensive URL's in text - either make a reference or a footnote.

#Answer#
We will do it at the final editorial stage. We wrote the address explicitly for draft such that reviewers can check easily.

#line 46 "auroral arc with expanding motion (we call it "Local-Arc-Breaking" hereafter)":
OK, you as authors refer to this phrase now for the third time including the abstract - please show a picture of what you mean!!!

The non-expert readership as well as me as reviewer has no idea or figure of what this expanding auroral arc looks.

#Answer#
In the new Figure 2 where the keogram is explained, we also added two images that are good example of the expansion of the aurora during typical Local-Arc-Breaking.

#lines 58-60 "This part requires a large computing ability for real-time processing because a large number of input data (each pixel of about 150000 active pixels has a 256x256x256 colour values) should be processed 60 simultaneously":

With all due respect, but processing what essentially boils down to a 256^3 image is NOT computationally challenging.

The phrasing of the part in brackets is perhaps also misleading:
- each pixel (N=150000) has values
- each values is a 256^3 colour (M=2^24 ~ 16,000,000)

*so, does the author want to tell us that 1 image has N*M = 2.4 * 10^14 values ? Yes, that would be really big in terms of data and then, even 1 image would be a challenge to handle. That said, I don't really know if that is the intended message.*

Illustration would help here a lot

#Answer#
We revised this misleading phrase. Each pixel has 6 byte color information, and we now removed the claim of "large data". We also revised the relevant parts in §2 and §4.

#line 61 "each pixel independently":
Ok, I am intrigued how that's going to proceed - per-pixel evaluation just from 1 pixel and without any global information, that would be challenging in terms of model stability and replicability,

#Answer#
It is not as challenging as it sounds, and to show it, we added many texts in §2.1 - §2.3 (old §2.1 is divided into three).

#line 62 " occupies solid small area":
In what way is that area 'solid' ?

#Answer#
Most of the pixels within a radius of >5 pixels are "moon". We added this information in §1 and §2.3.

#lines 68-69 "Therefore, we take a different approach using sets of solid criteria so that we can identify the reason for any error in the judgement, which is inevitable in improving the algorithm for the future version 2, version 3, etc.":

From my expert perspective, this is now the actual challenge:

for 30 years or more, we as the pattern recognition community tried to develop process-based, mathematical models and algorithms to transform raw-value data into interpretable information. 30 years of highs and lows, of success stories such as kMeans, decision trees, feature detection, etc. And for 30 years, we see that domain experts use those rigid models to 'fix' the pattern recognition scores of those algorithms with 'expert tuning'. And for 30 years, we saw that those tunings come with big limitations as the tuning makes every algorithm ungeneralisable - no knowledge gain for anyone 'outside of the bubble'.

We have seen in the past 10 years how deep neural networks outperform the traditional process-based modelling approach - and often even expert human judgement (!) - with a data-driven learning approach. I am by now personally not aware of any application area where a handcrafted approach and often even the expert opinion is better than a top-notch neural net, iff the data is there to train it properly. Going now back to a process-based modelling approach is a strongly questionable decision and mandates - and it has to be said so strictly - that then the developed approach needs to be compared fairly to a data-driven modelling approach.

Conversely, and I agree with the authors, the issue of neural networks is their very poor interpretability. The weights and feature maps don't tell a big story. Showing feature maps of neural net layers is like showing an MRI of a thinking person - lots of activations and so forth going on, but just from the scan of the thinking process, one doesn't know how the person (or the network) 'thinks'.

Where I disagree is now the conclusion of "because neural nets are hard to interpret, we don't use them but use an inferior method that is at least explainable" - that's like skipping the use of making fire and waiting for daylight because one didn't understand how fire works (personal opinion). A more valuable approach would be to say "we use a neural network to do the training, but then find a method that allows us to explain what the network does. The explainable method then gives us the information to refine the training and reason about why the network performs the way it does".

TL;DR: out of subjective but comprehensible reasons, I disagree with the approach. Out of objective reasons from literature, the shown approach does not represent the state-of-the-art in terms of classification expectations. From objective reason, it can be concluded that the presented method needs to be compared to a data-driven method in the end of the paper to provide for a relevant discussion of the research.

#Answer#
We added many sentences to explain why we took (we had to take) the expert system.

In many areas, pattern recognition is a strong tool, but not for aurora yet. Just identifying aurora "exist or not" (this is what existing machine learning is aimed for) was done long time ago using photometer with green filter and magnetometer monitor.

Problem is beyond it, i.e., "quantifying the aurora activity". While there are many Deep Learning of aurora classification, none of them went into the quantification, and our trial (including old version 0) is the only one that tries it. One of the reasons why the pattern recognition has not reached the quantification stage is probably that aurora morphology changes to almost all types of the shape much (ASC can over only a small part of the aurora which is a global phenomena) and "partial cloud" alters the original morphology quite a lot.

The present method is very "classic", but such a classic approach is very important for auroral community scientists who are not expert of machine learning classification. The very high success rate (§3) indicates that this is valid approach and, if limiting to now, the most successful approach. Actually, version 0 is already applied to Finland team.

We also mentioned that we do see the usefulness of the machine leaning, for partial application (step 1a and step 2, respectively). This is now mentioned in several places.

#line 70 "acronym; stands for "Joint Photographic Expert Group'"

#Answer#
We added

#line 70 "users" => "user circle"
one cannot 'widen the users' (lingual error in this context)"

#Answer#
We rephrased

--- 2. Algorithm and source data ---
#line 76 "(256^3=1.7x10^7 colour values for 1.5x105 pixels each)":
#Figure 1 (new Figure 5):
At this point, the method and especially the starting point needs clarification.

(1) yes, RGB space is the same as 3D real space - each value has a 3D coordinate. Meaning that each stored item has 3 values - 1 per dimension. If the RGB space referred to here is exactly that, then the calculation of 256^3 values is WRONG - it does not mean that one need to store 16.7 million colour values per pixel (which is what that bracket calculation suggests). Instead, each pixel just stores 3 values. The point where the 256 'levels' per channel (i.e. R-G-B; bands) come□ into play is the process called _quantization_. Quantization here means that the continuous photonic intensity for the CCD chip needs to be mapped to e.g. 256 (equi-distant) values. 256 values are stored, as binary number, in an 8-bit number. Thus, each pixel has 3 x 8-bit numbers, i.e. 3 x 1 byte, to store the values (in memory or on disk).

*The case (2) suggested by this 256x256x256 calculation and the figure would be that each spot in the 3D RGB space itself occupies not just 1 bit (i.e. that value being set or not), but that each spot in RGB stores multiple levels - just as if each value would store a full-form intensity curve. That one would indeed store per image 150,000 pixel * 256 * 256 * 256 values, and that indeed would be massive (!). That said, this is not what the text describes and what the vagueness of the figure also insufficiently expresses.*

Now, the text and the figure are equally ambiguous about which of the 2 situations in is. The figure would be clear if it was case (1) if the parantheses annotation is left out. Also, the shape of those blocks should be more binary - either a quantization step in that channel (i.e. band) is SET or NOT SET. Furthermore, the text need to be edited to the point saying that each of the 150,000 pixels stores 3 values, 1 for each band in RGB, with each band being quantized into 8 bit (i.e. 256 values).

What happens here is that - to my understanding and expecting case (1) to be what the method is based on - the method is a plain multi-level RGB double-bound thresholding of a regular RGB image. This is a simple-yet-effective method, it is not the state-of-the-art in terms of image processing (but: simple), and it is very much explainable. That said, the way it is explained is very bad. It is also a case of 'over-selling' the method as being 'intricate' and 'complex'. That is in-and-of-itself not wrong, but because of the over-selling the reader get the impression that the explanation is ambiguous, and thus inhibits the comprehension of the method - and that, in the end, is bad.

If the authors are in contrast actually dealing with case (2) then this certainly doesn't come through the explanations at all. The explanations are inconcise and need to be revised to a simpler and less ambiguous point.

#Answer#
Thank you for clarification. Yes, our writing was very misleading and we revised the statement (and no longer saying "large").

#lines 90-91 "The camera is configured for night-time observations and it operates with dynamic exposure time in order to be able to observe during twilight conditions.":

More information please on the setup of the exposure!

Because of the nighttime recording and the stark contrasts in brightness, it is important to detail how to setup exposure limits, white balancing, and so forth, so that other researchers can replicate the method. Furthermore, I expect an EXPANDED discussion in the end of the paper on the JPEG format as information carrier. JPEG is a lossy compression format. Even a compression-level 0 (lossless) compresses and culls the data because the DCT values within JPEG for YCrCb are culled to an 8-bit range. This quantization error possibly affects the any image analyses here. So, do you as the authors have a comparison in terms of RSME between raw and JPEG images ? Or are there other literature items that the authors could cite for such a comparison.

#Answer#
We added description of the camera setting and output format. The camera is "automatic exposure" with actual exposure time 1-30 sec (now stated). The JPEG compression difference is much less different than the camera difference and it is not an obstacle for the present purpose (we added this information at the end of the introduction).
We do not have any raw data from the camera, but we adjust the parameter (camera setting) by comparing the JPEG image and actual aurora (it took 4 months to optimise it), as is commonly done with ASC teams. This is why "eye identification" is important for validation (validation part explanation in §2.5 is enhanced accordingly).

#line 93 "as long as we can define the activity level properly."
comment: does that really work with those 'fixed' threshold intervals in 2.2 ?

#Answer#
Yes. This is the entire idea behind (and well known in the auroral community, with some "fuzzy" difference between individual scientists).

#Figure 1 (second comment)
Why are these blocks trapezoidal instead of neat blocks ? Does the shape of those blocks have any meaning in your method ?

#Answer#
The criterion uses R/G ratio and B/G ratio in different manner, and it is not a simple block. To make this clearer, we revised the table format for Table 2-4.

--- 2.1. Classification of each pixel ---
#line 123 "N2 red line":
need reference here - what is that 'N_2 red line' ?
#line 131 "For example, strong N2 red line with very little green line can be added after separating its colour code from the twilight. So far, we have not included such strong N2 red line pixels as the aurora pixels ": please add illustrating images

#Answer to both#
We added Figure 4 (example of N2 line) and related extensive explanation in §2.1.

#lines 139-143 "In these classifications, some pixels meet ...., as long as 0.24<H<0.34":
Choosing the fuzzy classification according to expert priority is fine. As a reader I am interested in two things:
(1) is there something in the literature that validates the priority ? How do you know about the significance in this priority list ?
(2) have the authors also looks on a statistical manner in priority selection ? What's the prior and posterior precision and recall of this classification priorisation ?

#Answer#
The priority is a direct consequence of the definition of each class. We added sentences explaining the reasoning of the priority.

#line 157 "present version of dealing with the moon is still effective.":
I believe that the dealing is effective due to the unique image sample the author's observatory has. Still, from a method perspective, the method is not stable and hard to replicate for observatories closer to the equator.

#Answer#
We added the explanation on the method and its stability (because of our restrictive definition) in §2.3. It is characterized by high values of all of R, G, and B, because of moon's wide color spectrum, and hence dependent only on camera.

--- 2.2. (new 2.4) Calculation of the aurora ASC index ---
#line 164 "0.1-5%":
How is this determined ? Expert judgement is fine, but it needs to be backed by some reference in literature, otherwise it's (a) hard to replicate and (b) difficult to judge for another expert or novice.

#Answer#
After looking at all values over operated 5 months. It is now explicitly stated.

--- 2.3. (new 2.5) Evaluation of the activity level from the aurora ASC index ---
#Figure 2 (new Figure 6):
#(1) Why are those ASC images environment-mapped, rendered hemispheres with specular highlights ?
Rendering the images on a sphere isn't entirely practical here, but rendering the spheres via Phong shading with specular highlights skews the images themselves. Furthermore, with the lighting of the sphere, new 'shadows' are introduced. Lasting, the spherical rendering stretches the images (most apperent on the 4 central images), which obfuscates the data.
CORRECTION: after looking at figure 7, the bright spot is the moon.
#(3) please tell the reader here that the bright spot is the moon ... it looks artificial and isn't that straight-forward to recognize.

#Answer#
We indicated the moon in the images in Figures 1, 2, and 4.

#Figure 2 (new Figure 5):
#(2) On the additional information here: the universal time attached to the images is fine, but just the L3 value doesn't tell the reader a lot, because the classification depends on more information.

I think, as a ready, it would help more to show fewer images, but the ones that are shown than double the size (i.e. the space now used for an image quartet rather to be used for 1 image), and therein highlighting the few pixels taht contribute to the classification (i.e. the pixels covered by the 0.x%, as discussed above). Without it, the L-value doesn't add lots of information to me as a reader.

#Answer#
We replaced L3 to "Alert or not" of the Local-Arc-Breaking. Also, we enlarged the ASC images (by reducing the total numbers of the image) for all figures. We also added new Figure 2 to explain the keogram.

--- 4.1. Correction using exposure time information and UT information (ver. 1.1) ---
#lines 328-330 "However, if limited to Kiruna, we can use two extra information that is available. One is UT information that is tagged to the jpeg image. The other is exposure time given as hidden information of the jpeg file (can be extracted with "exiftool" command for python program":

Why would these information be limited to Kiruna ? The specific values may differ - sure - but the whole method shouldn't be just a collection of magic numbers anyway.

Every place on earth has UTC and a local time that can be included in the evaluation, if it boost local arc breaking detection. Every camera noways stores the exposure time in the metadata of the image - in particular for camera's that perform a dynamic exposure. If knowledge of the exposure improves the detection score, it should be used.

#Answer#
Thank you. We removed "Kiruna", and revised the phrases.

--- 5. Summary and conclusion ---
#lines 405-406 "Unlike the other automatic identification such as using the neural network, we used a set of simple criterions and calculations (expert system)":

yes, issue is: at no point has this (possibly over-simplistic and possibly magic-number driven) method be compared to other, statistically stronger pattern recognition approaches. This comparison is mandatory in the case of this manuscript to validate that the method works and it outperforms the data-driven approaches, especially in corner-cases where automatic methods fail.

With the presented method, the author oversell the method, claim it's simple, effective, flexible (due to the many parameters) and comprehensible. As reviewer, I see the presented method as (a) hard to reproduce, (b) non-generalisable to any other setup and location of observation, (c) very handcrafted with no comparison to up-to-date procedures, and (d) ambiguously explained. There is a mismatch in understanding here.

#Answer#
As described in the introduction (and §2), no machine-learning method reached the "quantification" stage that the present method aimed and reached. The machine-learning method requires well-defined learning set (no "purely automated classification by machine" does not yet exists because aurora can take any form and because partial cloud alters the visible aurora into any from), but that requires the same level of "subjective" examination of images as the present. For the reproduction and application to the other places, we added extensive texts to explain, such that it is not as difficult as it sounds from the old manuscript. Finally, we made it clear that the present method is not excluding future usage of the machine-learning method part by part.

#line 408 "256x256x256 colour values"
to be reviewed by the authors - potentially strongly incorrect statement.

#Answer#
We rephrased to "6-bytes color information"

#lines 410-411 "real-time when the index values satisfy a criterion "Level 6". The alert system started 5 November, 2021":
positive outcome of the study that I acknowledge: the scheme is in operation and appears to be effective, judging from annotary feedback.

#Answer#
Thank you. We made major revision such that this outcome is rather natural rather than surprising.

Reply to Reviewer #3

Thank you for your encouraging and detailed comments, and for reminding that potential readers are much wider than the auroral observation community.

#preface#
The paper presents a fast approach for real-time Aura detection by using RGB images acquired by an All-Sky-Camera (ASC) (i.e., full frame camera with a fish-eye lens, pointing towards the sky), located in an a heated-dome at the Kiruna Observatory. The aim of the authors is to identify intensification of auroral arc with expanding motion, named Local-Arc-Breaking (LAB), and send email alerts when the "Level 6" criterion for LAB aurora detection is satisfied.

The proposed approach is composed of two steps: 1) perform a pixel-wise classification of all the image pixels into different categories (including three aurora categories), based on the colour information of the pixel itself; 2) compute a series of indexes based on the percentage of pixels detected for each category and the average luminosity of the most intense aurora pixels. Based on the computed indexes, the alert system can detect most relevant aurora events and trigger an alert.

The topic is interesting and the proposed solution for a real time alert is relevant, especially because it is a fast and not computationally expansive approach (which is crucial for real-time applications). However, there are some critical aspects that should be solved (or at least discussed):

##comment 1 on Computation of the thresholds and transferability of the method.
The pixel-wise classification is basically based on thresholding on RGB data, and it relies on a large number of thresholds. This makes the classification fast and computationally easy. However, it requires a proper fine tuning of the threshold levels, that is crucial for an accurate classification. I believe that threshold calibration is mostly based on the authors,Äô experience, as very little information about this is provided in the paper. The clear consequence of this, is that the procedure may not be transferable to another site, with different environmental conditions and different instruments (as the authors stated in the paper). With this idea in mind, it would be useful if the authors can discuss er their choices and provide some more considerations that have led to the derived threshold values, rather than just providing the values themselves.

#Answer#
We enhanced the explanation of pixel-level color judgement in both the introduction and §2.1-§2.3 (dividing into three subsections), e.g., how to contain the shift color range from 5577Å particularly when N2 line is contaminated. We also added new Table 1 just for this, and also added Figures 1 and 4 for the classification of aurora.

##comment 2 on Camera geometry
Is the camera geometry considered in your study? I understand that you are not trying to reconstruct the location of the aurora (which may be an interesting extension of your study for the future), but the "occupancy" of the pixel classified as aurora (within the different classes), among the total amount of pixels. However, you are using a fish-eye lens (Nikon Nikkor 8 mm), with strong geometrical

distortions, that require a proper camera model for correcting them (please, see e.g., https://docs.opencv.org/4.x/db/d58/group__calib3d__fisheye.html). I think that this aspect should be taken into account (e.g., by undistorting the images before any image processing) or at least discussed to see if geometrical distortions have a relevant impact on the accuracy of the classification.

#Answer#
Including camera geometry is a part of future task in §4.5 (considering field-of-view). As explained there, pixels at higher azimuth will less valued than pixels at zenith. In addition, in §2.4, we added the explanation why we treat them equally.

##comment 3 on Comparison with different methods
It would be useful to have some comparison with different state-of-the-art approaches (or also widely established methods, if any) to fully grasp the potential of the proposed solution.

#Answer#
We added the text explaining that no state-of-art method has "quantified" for auroral activity from the image (as long as we know) in the introduction and §4. All machine-learning schemes (actually Learning) only classify the entire picture. There is a geomagnetic method using geomagnetic deviation only, but success rate is not very high for Kiruna (we validated over the same 2021-2022 winter season).

##comment 4 on Learning
In the paper, the authors state that Neural Networks (NN) are black boxes, difficult to debug, and strongly dependent on the training data. I partially agree with the authors on these statements. However, I believe that Learning (DL) is a powerful tool for identifying the presence of aural events and to classify them, according to a-priori defined classes (as it is done in this paper) and ground truth data (manually classified images, or images classified with the algorithm proposed in this paper). Moreover, if the NN layers are trained with datasets from different locations and cameras, and with proper data augmentation, this may result in a more transferable and generalized approach, that can be applied in different observatories. Additionally, NN is basically based on sequences of filter convolutions. Therefore, it may overcome the limitation of considering each pixel independently from its neighborhood for the classification (of course DL is not the only possible solution for this: spatial filters, Markov Random Fields are few other examples). Eventually, DL may be combined with the step 2 of the proposed solution to build a real time alert framework.

All this thing considered, I understand that the use of DL is out of the scope of this paper (I suggest considering it for future works, though), but I believe that some more references and discussions on recent works that involve DL for aurora image classification may improve the quality and completeness of this work.

#Answer#
Thank you for explaining the potential of Deep Learning, and we actually consider using Deep Learning or other machine learning for step 2 (from index values to alert level). We also see the possibility of using Random Forest for step 1a. We added more explanation in the introduction, §2, and §4, on the possible use of the machine-learning method as the part of the present scheme.

#comment 5 on Figure
I believe that figures and tables can be improved. In my opinion, Figure 1 is not clear, and it should be revised (please, see comments in the pdf supplement). Additionally, figures with ASC images are often placed far from the text in which they are cited (even 3 or 4 pages afterwards), and this prevent a smooth read of the paper. I suggest revising figure placement in the text, reducing the number of ASC images to the really informative ones and making them bigger.

#Answer#
In addition to adding four new figures and one new table, explanation for Figure 5 (old Figure 1) is enhanced to make the intension of the figure clear. The placement of the figure and size of figure are Latex program problem and we make sure they are located correct pages and size before during copy editing (see our annotated manuscript made by Word document).

#comment 6 on Tables
Moreover, I believe that the tables containing the thresholding conditions are too verbose and don‚Äôt add relevant information in the text. Moreover, they are splitted in different pages, making even harder to read the tables. I suggest summarizing the main concepts of the tables (e.g., the different classes and the main classification criteria) in just one or two tables to be included in the main text of the paper. On the other hand, all the thresholds can be condensed in one table in the appendix.

#Answer#
We combined Tables 1-3 (new Table 2) and Tables 4-6 (new Table 4). We also re-formatted these tables. With this new format, thresholds are aligned to each column.
For the location of the Table, we keep the present place because we added more explanation.

#comment 7 on Background knowledge
As commented by other reviewers, I also believe that some background knowledge should be included in the text to allow a wider community to understand your work.

#Answer#
We added large amount of explanation on the aurora (colour, ASC, keogram) in the Introduction and §2, with four new figures.

#comment PDF
Please, refer to the comments included in the pdf supplement for other minor comments.

#Typos#
We fixed types

--- abstract ---
#line 10 "N2 red line at 670 nm)"
Please, introduce the topic (or at least, provide some references).

#Answer#
We added Figure 4 (example of N2 line) and related extensive explanation in §2.1.

--- 1. Introduction ---
#line 59 "because a large number of input data"
This kind of images are definetely not a "large number of input data" for modern workstations, which can include tens, or even hundred of processor units and powerful GPUs for image processing. Unless the goal is to refer also to extremely low power processor units (e.g. Rasperry Pi, which is never mentioned in the paper), I suggest removing this sentence.

#line 60 "each pixel of about 150000 active pixels has a 256x256x256 colour values"
This sentence makes no sense. Please, revise it.
Maybe you mean "each image of about ...". In any case, I suggest to completely rephrase it.

#Answer to both#
We rephrased and removed "large" from all relevant places. We also mentioned the ability of the processor is PC level.

--- 2. Algorithm and source data ---
#line 75 "256^3=1.7x10^7 colour values for 1.5x10^5 pixels each)"
As before, this is not a large amount of data. I suggest to hedge all your sentences that refer to this topic, because they sounds just as "artificial".

#Answer#
We rephrased in the same way as the previous answer.

#Figure 1 (new Figure 5) caption "The criterion are thus"
Which criterion? As they are not explained yet at this point of the paper, they should be shortly explained, otherwise the figure is not self-explanatory.

#Answer#
The explanation for Figure 5 (old Figure 1) in the text is enhanced to make the intension of the figure clearer.

--- 2.1. Classification of each pixel ---
#line 108 "coordinate system"
Which "coordinate system" are you referring to?

#Answer#
We rephrased the entire paragraph (removed the confusing word "coordinate").

#line 112 "Furthermore, values (0-255) for R, G, B can take only limited discrete numbers"
The meaning of this sentence is not clear.
Are you referring to the radiometric depth of digital images, in which the color information is stored as 8bit integer number?

#Answer#
We rephrased the entire paragraph (we generally used "6 bytes color information").

#Table 1 (new Table 2) title "one of them"
What does "one of them" mean?

#Answer#
We changed to "AND logic between columns, and OR logic between rows."

#Table 1 (new Table 2)
Please, give some more information on this categories (rather than giving all the conditions, which may be included in one appendix)

#Answer#
We added explanation of "how to define" in the text with new Figure 1.

#line 120 "Either of them"
What are you referring to with "them"? Please, be explicit.

#Answer#
The sentence is revised

--- 2.2. (new 2.4) Calculation of the aurora ASC index ---
#line 190 "keogram"
I know that for experts in the field, it is cleat what a keogram is. However, not all the readers are experts. Therefore, I suggest introducing it the first time you mentioned it.

#Answer#
We added Figure 2 and relevant explanation just for the keogram.

--- 3.1. Successful cases
#Figure 4a (new Figure 7b)
Why there is a break line in the graph? Is it the L3 threshold? It is not immediately clear and, in case, it should be explained in the caption.

#Answer#
We added explanation that we combined different scales for 0-10 and 10-100

#Figure 4b (new Figure 8b)
please, place the graph closer to the relative text (that is 2 pages before) and to Fig. 3.

#Answer#
This is LaTex template problem, and we make sure right location during copy-editing.

#Figure 5 (new Figure 9)
please, use this stile also for Fig 3 and 4.
Please place Fig. 5 closer to the text (3 pages before).
Please, explain the "break line" in the graph.

#Answer#
We combined old Figures 3 and 4 to make Figure 7.
Location of the figure is Latex problem, and we make sure right location during copy-editing.
For Figure 5 (new Figure 8), we simply forgot to add Y-axis for (c) and (d), and we fixed

--- 4. Discussion and future tasks
#lines 294-296 "Adding a third category ... somewhat fizzy)."
Please, revise this sentence, as it is not clear (verb missing)

#Answer#
Yes, we revised.

#line 309 "this"
Please, be explicit. You are not mentioning "your software package" nearby here.

#Answer#
We revised the sentence.

#line 313 "filter matrix"
Why are you labelling the classification matrix as "filter"? I don't agree, as usually a filter is a small window (e.g., 3x3, 5x5) that is convolved all over the image.

#Answer#
We could not find a better name, and to compromise we use "masking matrix" for each category whereas "filter matrix" for individual small region listed in the Table 2-4.

#line 320 "Another advantage of the present method"
In my opinion, this is a drawback, not really and advantage (please, see major comments).

#Answer#
We added explanation why we believe an advantage: the entire "two-step" scheme can be reserved while upgrading each step (1a, 1b, and 2) can be done one by one, and such upgrade can be done with a help of machine-learning method.

--- 4.2. N2 red (670 nm) line (ver. 1.2)
#lines 342-343 "This is not impossible .... been defined."
Please, revise this sentence.

#Answer#
Yes, we revised.