

Interactive comment on “A comparative study of auroral morphology distribution between northern and southern hemispheres based on automatic classification” by Qiuju Yang and Ze-Jun Hu

Anonymous Referee #2

Received and published: 28 November 2017

Review of GI-2017-49

Yang and Hu, "A comparative study of auroral morphology distribution between northern and southern hemispheres based on automatic classification"

–

The authors describe experiments with data from two auroral cameras with one being in the northern hemisphere and the other in the southern hemisphere. The central idea is to use a KNN classifier to compare the auroral type and to construct occurrence distributions from the data. The topic is very relevant to the GI journal and a welcome contribution to data analysis methods in space research.

C1

In general, I feel that the manuscript would benefit from a language check by a native English speaker to improve the "flow". There are a few spots where interpreting what the authors want to say becomes difficult.

My main criticism is that the classification results are "too good". Earlier studies, including the listed references (Wang et al., 2010, Syrjasuo and Donovan, 2004) have highlighted the uncertainty of auroral categories. The issue is whether a fixed number of auroral classes such as "arc", "drapery", "radial" and "hot-spot" is sufficient for all auroral shapes one encounters in the data. A simple method like the KNN will always select K neighbours and thus always provide a label regardless of input data. From the results (table 2) I do get a feeling that the four auroral types can be separated quite well. But what happens if one tries to classify an image that does not belong to any of these classes? An obvious example would be diffuse aurora, which does meet the criteria of not being taken during bad weather and having aurora in the sky (although not necessarily daytime...). This is very important from the practical point of view: if we can rely on the classification, we can also consider, for example, higher time resolution, if something "interesting" happens in the sky.

So, I recommend a major revision with more depth in the analysis of the retrieval and classification itself.

1) Image retrieval (Fig 3): rather than showing only the closest match, which always exists, it would be much better to show, for example, five or more closest matches with their Chi-squared distances. What are your observations, are all images "matches"? At what distance do you think the images are not any more similar?

For this analysis to work, you most likely need to consider the capture time of each retrieved image to separate the similarity due to capture time rather.

2) Machine learning algorithms such as KNN are known to be rather bad in extrapolating and then there is the infamous "curse of dimensionality" (sampling density). So, my main question is whether one can always rely on the classification result. Answer-

C2

ing this does require some additional work. KNN is a special case of kernel density estimation and there are thus built-in uncertainties which should be examined.

a) Your dataset ARD contains four types of aurora. What is the variability/distribution of pair-wise distances within one class? I would expect arcs to be a "tight group" but all of the other types may be more scattered. What is your observation and conclusion about how it may affect your results?

b) What is the variability/distribution between the distances between different types of aurora? Are the types of aurora well separated? If not, should that be taken into account in KNN by, e.g., ignoring far away neighbours? While requiring all neighbours to be from the same class does improve the reliability, it still chooses the closest class (whether it belongs there or not). For example, I do not see any obvious difference between the "radial" (Fig 3g) and "hot-spot" (Fig 3j) types of aurora.

3) How does the viewing geometry affect your classification? And does it matter for this study?

=====

Minor comments:

4) Abstract, line 9: the authors were perhaps thinking of "important" rather than "significant"?

5) Introduction, line 30: please clarify the sentence containing "the inversion of magnetic layer structure and dynamic process"

6) Introduction, line 33: replace "our country" with, for example, "Chinese research institutes".

7) Introduction, line 52: what do you mean with "multiple-wavelength intensity distributions"? Is it relevant for this manuscript?

C3

8) Data and methodology, line 76: I am unsure what exactly you mean with "MLT~UT-3.6 hrs."

9) SPS data and YRS data: in my opinion, the description of preprocessing steps and the instruments would benefit from a table which would list the technical details for each instrument. As far as I see the process starts with first manually removing data without aurora and then rotating and scaling the images for classification. For YRS, there is also subtraction of dark current (which is assumed to be the same for the whole image?). So, the steps are almost identical for both cameras and just the rotation angles etc. are different.

10) SPS data, line 84: your statement "cannot work normally in cloudy or foggy weather" is incorrect as the instruments work fine in those conditions – you just cannot see the aurora! If I interpret the manuscript correctly, you (manually) select only those images that contain aurora. Please clarify.

11) SPS data, line 89; YRS data, lines 564 and 117: what are the dynamic ranges of these cameras? Providing a number "4000" does not mean anything in itself. My understanding of intensity stretching uses a minimum pixel value (black point) and the maximum pixel value (white point). Is this the maximum and you start from zero (black)? Please clarify.

12) SPS and YRS data: why are you carrying out the image stretching? LBP uses relative intensity values and should thus be insensitive to brightness changes as long as the order of brightness is maintained. Please clarify.

13) ASC image representation, lines 119-135: Wang et al. (2010) used an improved LBP and carried out experiments to determine best parameters for LBP neighbourhood size (and number of samples). They concluded that the parameter choice was rather insensitive but they are essentially using a 5x5 neighbourhood, from which 8 pixels are sampled at a radius of 2 pixels from the centre pixel. On the other hand, you are using 3x3 neighbourhood. Could you add some text to explain why you chose to do it your

C4

way and what the advantages/disadvantages are, please?

14) Classification mechanism, line 156: how well does this 1-min time interval work with arcs that may be present for a longer time (also in the daytime)?

15) Supervised classification: how is this different from what Wang et al. (2010) did? Are the labels now "W", "M", "H" and "D"? What about those images that have aurora whose type is not arc, drapery, radial or "hot-spot"?

16) Supervised classification, line 192: Please clarify your choice $k=1,3,25$. Why 25 rather than sampling closer to 3?

17) Table 1. Please revise the caption of table ("image number" means the number of images?)

18) Table 2. There is a bias in all error analysis methods and crossvalidation is no exception: please reconsider the precision of classification accuracies (is 98.37% really meaningful or should it be 98% due to errors in estimating the error?)

19) Occurrence distribution, line 250: you indicate (lines 81 and 103) that both cameras can operate 24-h in the wintertime. Yet this is clearly not fully correct as there are fewer images around the noon due to sunlight. Please be more specific and revise the text accordingly.

20) Occurrence distribution, line 260: is this description of "fast changing" hot-spot auroras coming from the data? You were analysing single images without temporal analysis, weren't you?

21) Conclusions, lines 313-314: your sentence "all dayside coronal auroras can be classified into three auroral types" is a very strong statement. Given my concerns about the number of possible (correct) auroral types, I don't quite agree yet. I leave it to the the authors to consider whether this statement belongs to "discussion" rather than conclusions.

C5

An observation, not a comment on the manuscript:

The authors might want to consider using their method to study the lifetimes of dayside auroral forms in a future study: it would be interesting to use Chi-squared distance to locate time periods when the forms are stable (or changing rapidly). A simple method would be to create a time-series of the distance to the previous image to study the evolution of similarity in time.

Interactive comment on Geosci. Instrum. Method. Data Syst. Discuss.,
<https://doi.org/10.5194/gi-2017-49>, 2017.

C6