Authors' Response to Referee Comment #2

Sedlak et al., 2023: Analysis of 2D airglow imager data with respect to dynamics using machine learning

We would like to thank Anonymous Referee #2 for his valuable comments.

Referring to the general comments:

- Out of curiosity, is there a specific reason for pointing the imager toward azimuth 204 degrees?

The azimuthal direction results from aligning the camera system with the window in our laboratory.

- l. 95-98: hasn't fladfielding already removed reflection of the lens for example?

It has for the major part but in practice, there are light remnants of a pattern left which we get completely rid of by subtracting the average image.

- Do you know if the star field has any effects on texture-based parameters like homogeneity or uniformity? Or on the FFT analysis?

Up to now we have not evaluated the effects of the star field on the texture-based parameters, but this might be interesting to investigate. We keep this in mind for our next analyses. We did investigate the influence of the star field on the PSD using quite comparable data captured at Otlica, Slovenia – we did the analysis both with and without removing the stars from the image. It turned out that there is hardly no difference in the course of the integrated PSD except a small offset. The temporal courses of IPSD with and without star removal show a correlation of 0.99, which led us to the assumption that the effect of the star field is negligable for our purpose.

- A table summarizing the 8 features would be helpful to remember what they are, and in which way they characterize the images.

We agree and included such an overview as Table 1.

- l. 268: How did you select the value -0.2? Does it come from another study or is it empirical?

It was empirical, but we paid attention to the range of learning that is covered. And that shouldn't be too high, because otherwise the training becomes ineffective. However, we wanted to have scheduling because it gives better results than without scheduling.

- l. 309-316: The authors say that only the psd feature group shows some significant effect on the precision. Would it be worthy to remove the basic feature group, for example (faster calculations with still good results)?

Would be an option

- Tables 1 and 2 could be combined in only one table XXXX (0.YY)

We appreciate the suggestion but we decided to display them separately for better readability.

- You realized that a lot of the mispredicted "calm" or "dynamic" images are in fact misclassified. How could you improve the classification?

First of all, we were surprised that there were so many misclassifications in the manual classification by hand, but it's maybe reasonable if you consider the following two aspects. First, distinguishing between " calm " and " dynamic " is not like recognizing a clear object. The transitions are smooth, and you are probably (unintentionally) biased by the overall night. For example, on a very calm night you are more likely to consider low dynamic activity as dynamic than on very dynamic nights. On the other hand, it is very hard to stay focused all the time if you are watching these videos for many hours. It is therefore very good news that a large proportion of the mispredicted sequences are in fact misclassified, despite the fact that the model is trained with probably many misclassified images in the training dataset. The aforementioned human bias with respect to single nights is also something that does not play a role in the predictions of the trained model. Therefore, you can improve the classifications by using the model instead of classifying the images by hand.

- The mean average precision is 0.82, but only 0.63 for "dynamic" images (which is what you are looking for). It doesn't seem that good!

Yes, it doesn't look good at first, but we tried to discuss it in detail.

1. There are many misclassified images that affect the result.
2. Distinguishing between calm and dynamic (which is the reason for the low average accuracy) is not like distinguishing between clearly defined objects. Very often the prediction is formally classified as wrong, but one could discuss who is wrong, the manual classification or the model prediction.
3. Nevertheless, the model still confuses a lot of sequences, so more research is needed to find additional features or other ways to minimize this confusion.

- The authors say that only using 8 features in the neural network is faster than the full images, but calculating these features requires already a lot of calculations (like for the FFT).

Indeed, these features require a lot of computation, but you only need to do this computation once because you can store features (the results of these computations). The difference with training a neural network is that the computations for training are needed in every epoch, and in addition they have to be done for whole sequences and not for single frames. These two aspects significantly increase the computational cost compared to our approach.

- Would it be possible to apply this method, or a similar method, on raw data? OF course, it would be a problem for the psd group! So, maybe not possible.

As the psd-based features turned out to be of great importance for the algorithm, we agree that it seems unlikely for this to work. However, we will give this a try in the next analyses. Since the algorithm turned out to perform this will, maybe it might surprise us at analysing images that are unrecognizable for the human eye.

- Only 13 dynamic features in 8 months (~1.5 per month). How does it compare with previous studies from the same authors?

Comparing this to our study of data captured at Otlica, Slovenia (Sedlak et al., 2021), where we could analyse 25 turbulence events in 19 months (~1.3 per month), this rate does not appear to diverse. The measurement setups were also quite comparable as concerns field-of-view and spatio-temporal resolution. With ongoing measurements it will be interesting doing longer-term analyses at different locations and to investigate, whether there are systematic differences in the occurence / intensity of turbulence related to local peculiarities (e.g. the strong bora winds near Otlica). The AI algorithm presented here provides a valuable instrument to approach this research.

Referring to the minor edits:

Line 26: Changed to emission.

Line 31: Corrected.

Line 43: Comma inserted.

Line 52: Comma removed.

Line 59: Corrected.

Line 62: Corrected.

Line 69: Corrected.

Line 72: Unfortunately we cannot comprehend what shall be suggested here. No changes made.

Line 80. Parenthesis omitted.

Line 90: Changed.

Line 91: Changed.

Line 94: Already deleted following the suggestions of Referee #1.

Line 92 (we believe that you mean line 97): We indeed mean the window of the building which can cause reflections.

Line 122: Corrected.

Line 129: Changed.

Line 140: Already omitted while following the suggestions of Referee #1.

Line 150: There is no space in our original DOCX file. This seems to be an artifact of the PDF export.

Line 155: Corrected.

Line 237: Already corrected to „of neurons“ while following the suggestions of Referee #1.

Line 315: Changed.

Line 321: Changed.

Line 336: Corrected.

Line 365: Corrected.

Line 393: Already changed to „In a further step“ while following the suggestions of Referee #1.

Line 419: We changed "even though the labelling was incorrect" to "as the labelling was incorrect".

Line 430: Corrected.

Line 440: Corrected.

Line 492: Changed.

Line 499: Changed.

Line 504: Removed.