

Ann. Geophys. Discuss., author comment AC3
<https://doi.org/10.5194/angeo-2021-33-AC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Maria Elena Innocenti et al.

Author comment on "Unsupervised classification of simulated magnetospheric regions" by
Maria Elena Innocenti et al., Ann. Geophys. Discuss.,
<https://doi.org/10.5194/angeo-2021-33-AC3>, 2021

We thank the reviewer for their comments, that helped improve our work. We reply here
in details, reporting the reviewer's comments in bold for easy reading.

Major comments:

1. The authors do not discuss the stochastic nature of artificial neural networks, their sensitivity to initial conditions and convergence to local minima. Clarification for this issue would be needed to the section discussing Self Organizing Maps.

To clarify the issue, the following sentences will be added to the manuscript:

It is useful to remark that, even if the same data are used to train different SOMs, the trained networks can differ (and most probably will), due e.g. to the stochastic nature of artificial neural networks and to their sensitivity to initial conditions. If the initial positions of the map nodes are randomly set (as in our case), maps will evolve differently, even if the same data are used for the training.

We have verified that SOMs trained starting from different initial node positions give comparable classification results, even if the nodes that map to the same magnetospheric points are located at different coordinates in the map. The reason for this comparable classification results is that the 'net' created by a well-converged SOM will always have a similar coverage and neighbouring nodes will always be located at similar distances with

respect to their neighbours (if the data do not change). Hence, while the final map might look different, the classes and their properties will produce very similar end results. We refer the reader to Amaya et al. (2020) for exploration of the sensitivity of the SOM method to the parameters and to initial condition, and for a study of the rate and speed of convergence of the SOM.

2. The nature of the hyper-parameters of the Self Organizing Maps are only briefly discussed in Section 3, more clarification on their influence to the algorithms performance is needed. A case study is done in appendix A, but a more theoretical description of the nature of the hyper-parameters should be added.

We refer the reader to Amaya et al (2020) for the study of the convergence of the method with different parameters.

3. In Section 3 the authors do not mention what distance metric is used for the matching rule in Equation (1) and updating rule in Equation (2) of the SOM.

We use the Euclidean norm. The manuscript will be updated accordingly.

4. Clarification should be added about how the authors initialized the neural map of the SOM.

The initial nodes are randomly distributed. We have verified that SOMs trained with the same data starting from different initial node distribution give comparable results.

5. The influence of sampling of input data during learning needs to be discussed.

This issue will be clarified in the manuscript as follows:

The selection of these points is randomized, and the seed of the random number generator is fixed to ensure that results can be reproduced. Tests with different seeds and with an higher number of training points did not give significantly different classification results.

6. The authors should describe how much confidence they have in the result of the SOM.

As we will describe in the revised Conclusions, we are at the moment quite happy with the classification results we obtain, because they map well to our knowledge of the system and appear to be quite robust to temporal variations in the simulated magnetosphere. The fact that a good subset of the features we tested gives comparably good results also points in the direction of a robust procedure on simulated data. Of course the real test for the method will be using spacecraft data, which will be extremely more challenging in terms of instrument noise, instrument limitations, presence of kinetic processes.

7. In Section 4.2 the model validation is done only by visual inspection. The colors of similar clusters differ from image to image. This fact also makes the readability of the figure very hard, as the eye automatically matches colors. A quantitative measure should be introduced for the robustness of the SOM. To discuss Figure 7 data, one could simply calculate for each result comparison pair for the same data a percentage of similarly labelled data points. Human labelling of the 7 clusters is already previously done (Figure 4 panel (d)). The same labeling system could be used for the K-means classification results displayed on Figure 7.

The colors in all plots will be matched to those of F1, to simplify visual comparisons of results.

We will add this while commenting figure 7:

To compare the two classification methods quantitatively, we calculate the number of points which are classified in the same cluster with SOMs plus K-means vs pure K-means classification. 92.15 % of the points are classified in the same cluster, 92.74 % if the two magnetosheath clusters just downstream the bow shock are considered the same. These percentage are calculated on the entire training dataset at time $t_0 + 210$ minutes, of which cuts are depicted in the panels in Figure 7.

8. Clarify what time snapshots and data is analysed on Figure 8. Do the panels on the figure correspond to the same dataset with different training features? Clusters depicted on Figure 8 follow different coloring schemes, which diminish the readability of the results. To better quantify the comparison of independent algorithm results on the same data a quantitative measure should be added. Similarly to the previous comment 7, Figure 4 panel (d) labeling system can be applied to all results. One could count the fraction of similarly labelled input data points for each compared pair of independent results for the same data.

In the submitted manuscript, Figure 8 depicted classification with different sets of training features of the training data points. In the new version, we will show results for the same validation dataset as Figure 6. In both cases, $t = t_0 + 210$ minutes. Results do not differ significantly between the two sets of pictures. The colors in the new picture will be changed to match the cluster description in Figure 4.

We will add a table with the % of data classified in the same cluster as F1, see Table 3 in supplementary material. The manuscript will be edited as follows:

In Table 3, second column ("S"), we report the percentage of data points classified in the same cluster as F1 for each of the feature sets of Table 1, for the validation dataset at $t_0 + 210$ minutes. In the third column ("M"), we consider cluster 1 and 4 as a single cluster: in the previous analysis, we remarked that cluster 1 and 4 (the two magnetosheath clusters just downstream the bow shock) map to the same kind of plasma. We keep this into account when comparing classification results with F1.

The metrics depicted in Table 3 cannot be used to assess the quality of the classification per se, since we are not comparing against ground truth, but merely against another classification experiment. However, it gives us a quantitative measure of how much different classification experiments agree.

9. Similar quantitative robustness measure should be added for the comparison of independent results for the same data in Figure 9.

Done, see table.

10. Use the same coloring scheme for all cluster analysis results (Fig 6 – Fig 9) as different colors lose information in Figures. In Figure 4 a labelling system with colors is created, which could be used for all figures. Only visual comparison of clusters over multiple images throughout the paper with different color coding is tiresome.

Done.

11. The comparison of panels between Figure 4, Figure 6, Figure 7, Figure 8 and Figure 9 would be easier if all the captions of the panel names would contain the time snapshot information. Panels on the Figure 6 have different sizes, panel

sizes on a Figure should be uniform.

The time snapshot information will be added to all captions, and labels will be made uniform. We will also add (T) and (V) to the caption, to label the training and validation datasets.

Minor suggestions and questions:

L167: Does the dimension of the input data influence the training time of the map considerably? Are there other metrics that could possibly influence it more? Can the PCA be used to initialize the neural map of the SOM? Is the main motivation for input data dimensionality reduction to have more reliable results from training of the SOM?

In our experience, it is the number of points, rather than the number of features, that influences the training time the most, so the main motivation for dimensionality reduction was in fact not training time (even if that helped) but rather, as the reviewer suggests, generating a more reliable training dataset. The correlation analysis (and previous knowledge) made clear that several of the magnetospheric variables in fact carry the same information on the state of the plasma, and we aimed at compressing that information in a lower number of features, while preserving a high percentage of the variance of the original dataset.

PCA can certainly be used to initialize the SOM map, but in this case we chose random initialization.

Fig 3: the name of panel (a) is in different size.

Fig 4: the style of panel descriptions differ in the Caption. A more uniform style would increase the readability of the figure.

Fig 7: the name of panel (b) is in different size.

We will improve figure presentation in the revised version.

Please also note the supplement to this comment:

<https://angeo.copernicus.org/preprints/angeo-2021-33/angeo-2021-33-AC3-supplement.pdf>