

Ann. Geophys. Discuss., author comment AC1
<https://doi.org/10.5194/angeo-2021-33-AC1>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

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-AC1>, 2021

We thank the reviewer for the useful comments, that will help improving our work. We answer to the comments below, rewriting the reviewer's comment for clarity.

Major points:

L94-95: Please explicitly state that the boundary conditions vary with time. Are all boundary cells surrounding the simulation domain always at the same value, varying identically with time?

The manuscript will be edited as follows:

The OpenGGCM-CTIM-RCM boundary conditions require the specification of the three components of the solar wind velocity and magnetic field, the plasma pressure and the plasma number density at 1 AU. Boundary conditions in the sunward direction vary with time. They are interpolated to the appropriate simulated time from ACE observations (Stone et al., 1998), and applied identically to the entire sunward boundary.

L157: Please clarify the data point selection. Is it randomized? Is there any selection based on Y or Z?

We have randomly selected 1 % of the points included in the $-40 < x/RE < 18$, $-Ly < y/RE < Ly$, $-Lz < z/RE < Lz$ subdomain, seeding the random selection to be sure to be able to retrieve the same dataset if needed. No selection based on y or z has been used.

This 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 result into significantly different classification results.

Figures and analysis: Although the polar plane is very interesting in many ways and should indeed play a role in the study, the equatorial plane could be considered even more relevant, particularly considering the orbits of recent significant space missions. I strongly urge that you present also equatorial plane plots early on to show the strength of the method.

In the revised version of the manuscript, we will introduce equatorial plane results, now shown for the first time in Figure 7, in Figure 4 as well. Furthermore, we will add validation plots for the F1 feature set in the equatorial plane in Figure 6, which in the old version of the manuscript shows only the meridional plane.

The plots we will add are shown in Figure 1 (added as supplement). They are consistent with the classification in the meridional plane at the same times. Also the mis-classifications are consistent: also in the equatorial plane we see mis-classification of a few bow shock points, which are classified as inner magnetosheath (orange), and mis-classification of a few inner magnetosphere points as inner magnetosheath plasma at $t_0 + 225$ minutes (orange).

Table 2: Instead of multiplying values in the latter two feature sets, could all sets perhaps be re-normalized to a norm of 1? This would make comparison easier.

We will remove the renormalization for the first two feature sets.

Kneedle determination of optimal cluster count: did you attempt changing the number of clusters to see how robust the method is? Would setting $k=6$ merge clusters 1 and 4? How would this change the misclassified cluster 5 points at the bow shock for F1? What about setting $k=5$, would clusters 1,4 and 5 merge? I believe a written description of such a test without figures would suffice.

Short answer: $k = 6$ merges the three magnetosheath clusters into two, the other clusters are left unaltered. With $k = 5$, the boundary layer cluster disappears, and the points that mapped to it are assigned to the clusters mapping to inner magnetospheric (this is mostly the case of current sheet plasma), magnetosheath or lobe plasma. In all cases, we keep seeing few bow shock points mis-classified as inner magnetosheath plasma.

In the pdf added as supplement, we examine in some details and with plots the classification results in the meridional and equatorial plane with $k = 6$ and $k = 5$. We propose to add that as an Appendix in the revised manuscript.

Could you please add some quantitative estimate of the agreement between different feature sets? This would be particularly useful if you also included comparison against the pure K-clustering approach, to show how much improvement SOMs bring to the table. For this purpose, I think manually merging some clusters (e.g. 1,4,5) would be acceptable.

We would avoid manually merging clusters, as it would defy the purpose of unsupervised classification. In the new Appendix we propose to add a comparison of SOM + K-means vs pure K-means classification with $k = 5$ and $k = 6$ as well, as suggested. We do not observe significant differences in SOM+ K-means and pure K-means classification with $k = 5$. With $k = 6$ (as already observed with $k = 7$), the SOM+ K-means classification is more robust to the mis-classification of inner magnetospheric plasmas.

L372: Since you talk of lessons learned, I was surprised that you did not describe attempts using the logarithm of magnitude of B, instead going with the quasi-arbitrary clipping procedure. Please try out $\log(B)$ if not already attempted, and at least briefly report the results.

We will add a new feature set, F10, where $\log(|B|)$ is used as a feature, instead of clipped $|B|$. The new validation plots are reported here in Figure 2 added as supplement.

As you can see comparing F1, F9 and F10, F10 looks remarkably similar to F1 in the internal regions, including the mis-classification of some inner magnetospheric points as magnetosheath plasma at $t_0 + 225$ minutes. The three magnetosheath cluster vary at the three different times depicted with respect to F1. This behavior, and the pattern of classification of magnetosheath plasma in F9, shows that the magnetic field is a feature of

relevance in classification especially for magnetosheath regions. F10 classification results show that the blue cluster in F9 originates from the clipping procedure. This somehow artificial procedure is however beneficial for inner magnetospheric points, which are not mis-classified in that case.

Conclusions: I would like to see some more discussion about relevance and limitations. An important point to discuss, albeit outside the scope of the current paper, is the identification of small structures inside larger domains (e.g. the tail current sheet within the boundary layer) and the points of transition between domains. The latter might arise naturally from this method, the former not so much. Some mention of this should be included. Also, there has been no discussion of the drastic difference between MHD simulation descriptions of plasma parameters and the values measured by spacecraft or hybrid simulation methods, namely kinetic effects, noise, and/or instrumentation limitations. Some of these are touched upon in Amaya(2020), but they are quite relevant to the discussion here as well, if this method is indeed to become a first step in any actual classification approach.

We will add this discussion to the Conclusions:

In this paper, we have classified large scale simulated regions. However this is only one of the classification activities one may want to be able to perform on simulated, or observed, data. Other activities of interest may be the classification of meso-scale structures, such as dipolarizing flux bundles or reconnection exhausts. This seems to be within the purview of the method, assuming that an appropriate number of clusters is used, and that the simulations used to produce the data are resolved enough. To increase the chances of meaningful classification of meso-scale structures, one may consider applying a second round of unsupervised classification on the points classified in the same, large-scale cluster. Another activity of interest could be the identification of points of transition between domains. Such an activity appears challenging in the absence, among the features used for the clustering, of spatial and temporal derivatives. We purposefully refrained from using them among our training features, since we are aiming for a local classification model, that does not rely on higher resolution sampling either in space or time.

We will edit L420 as follows:

As a first step towards the application of this methodology to spacecraft data, we verify its performance on simulated magnetospheric data points obtained with the MHD code OpenGGCM-CTIM-RCM. We choose to start with simulated data since they offer several distinct advantages. First of all, we can for the moment bypass issues, such as instrument noise and instrument limitations, that are unavoidable with spacecraft data. Data analysis, de-noising, pre-processing is a fundamental component of ML activities. With simulations, we have access to data from a controlled environment that need minimal pre-processing, and allow us to focus on the ML algorithm for the time being. Furthermore, the time/

space ambiguity that characterizes spacecraft data is not present in simulations, and it is relatively easy to qualitatively verify classification performance by plotting the classified data in the simulated space. Performance validation can be an issue for magnetospheric unsupervised models working on spacecraft data. A model such as ours, trained and validated against simulated data points, could be part of an array of tests against which unsupervised classifications of magnetospheric data could be bench marked.

The code we are using to produce the simulation is MHD. This means that kinetic processes are not included in our work, and that variables available in observations, such as parallel and perpendicular temperatures and pressures, moments separated by species, are not available to us at this stage. This is certainly a limitation of our current analysis. This limitation is somehow mitigated by the fact that we are focusing on classification on large scale regions. Future work, on kinetic simulations and spacecraft, will assess the impact of including "kinetic" variables among the classification features.

Minor points:

Abstract: I would recommend you mention comparisons against the K-clustering event here, as well as the use of PCA to reduce input data.

We will add the following sentences to the abstract:

The dimensionality of the data is reduced with Principal Component Analysis before classification. ... We validate our classification results by plotting the classified data in the simulated space, and by comparing with K-means classification.

L14: Introduction to machine learning: Perhaps some more general, canonical paper could be cited?

The sentence will be edited as follows:

The growing amount of data produced by measurements and simulations of different aspects of the heliospheric environment 25 has made it fertile ground for applications

rooted in Artificial Intelligence, AI, and Machine Learning, ML (Bishop, 2006, Goodfellow et al, 2016). The use of ML in space weather nowcasting and forecasting is addressed in particular in Camporeale (2019).

L30: how has the data magnitude changed when compared with Cluster, the Van Allen probes, ISEE-1, etc?

We will add this sentence to the manuscript:

The four-spacecraft Cluster mission (Escoubet et al., 2001) has been investigating the Earth's magnetic environment and its interaction with the solar wind for over 20 years. Laakso et al. (2010), introducing a publicly accessible archive for high-resolution Cluster data, expected it to exceed 50 TB.

L90: Please clarify in better detail the simulation set-up. Is this the domain size of the whole simulation, or the portion of it shown in figure 1? In section 4 you state you use only the relevant portion of the simulation for post- processing SOM analysis. What is the spatial resolution of the simulation?

The description of the simulation set up will be updated as follows:

OpenGGCM-CTIM-RCM uses a stretched Cartesian grid, which in this work has 325x150x150 cells, sufficient for our large scale classification purposes, while running for few hours on a modest number of cores. The point density increases in the Sunwards direction and in correspondence with the magnetospheric plasma sheet, the "interesting" region of the simulation for our current purposes. The simulation extends from -3000 RE to 18 RE in the Earth-Sun direction, from -36 RE to $+36$ RE in the y and z direction. RE is the Earth's mean radius, the Geocentric Solar Equatorial (GSE) coordinate system is used in this study.

In this work, we won't be analyzing the entire simulation, but the subset with coordinate $-40 < x/RE < 18$, i.e. the magnetosphere / solar wind interaction region and the near magnetotail.

L119: The indicated magnetic field clipping value does not make much sense. Is the intention to keep the same ratio between the three components and the original signs, but re-scale the magnetic field vector to a magnitude of 100 nT? Please rephrase.

The intention of the clipping procedure was to cap the maximum magnitude of the magnetic field module to 100 nT, while retaining information on the sign of each magnetic

field component. The respective ratios of the original field components are lost with this procedure, since each of them is arbitrarily set a fixed value. Any negative effects of this procedure is limited to feature set F5 (old labelling, F7 new labelling), the only one which uses the clipped magnetic field components. Comparing the F5 (F7) validation plots with F1, it appears that the clipping procedure did not have a major influence on classification results.

L129: is the lattice really of type R^2 ? The visualizations show a hexagonal grid.

The lattice is of type R^2 . To clarify the visualization, this sentence will be added to the text:

The node location in the lattice can use a Cartesian pattern, but most applications relocate the nodes in a hexagonal arrangement in order to feature equidistant points to the six closest neighbouring nodes.

L131: I recommend you briefly mention that from available plasma variable you select n features for the SOM method to use, so that the R_n notation is meaningful.

The following text will be added to the manuscript:

n is therefore the number of plasma variables that we select, among the available ones, for our classification experiment.

L135: Could this be better described (to the layman) as altering the feature values of the code word so that the distance w_s for the winning element becomes smaller? Similarly, consolidation of terms might make it easier to read, e.g. data entry vs input data point vs input point - these are probably all the

same thing, i.e. a list of features associated with a point of measurement.

The comments will be incorporated in the manuscript as follows:

SOMs learn by moving the winning element closer to the data entry, based on their relative distance, and on a iteration number-dependent learning rate $\epsilon(\tau)$, with τ the progression of samples being presented to the map for training: the feature values of the winning element are altered as to reduce the distance between the "updated" winning element and the data entry.

Indeed, data entry, input data and input point are the same thing. We will add a clarification statement in the manuscript:

Notice that, in the rest of the manuscript, we will use terms such as "data point", "data entry", "input point" interchangeably.

L143: is the numerator of the exponent in formula (3) the integer lattice neighbor distance, up to a value of sigma(tau)?

$\sigma(\tau)$ is the value of the lattice neighbor width, which is not necessarily an integer. $\sigma(\tau)$ decreases with the iteration number (τ) to ensure that the map converges at the end of the training.

L201: Is the K-means clustering performed based on the final code words of each node?

Yes.

L282: I would suggest briefly mentioning the sunward inner magnetosphere misclassification already here.

We prefer to complete our line of thought before introducing the inner magnetosphere misclassification, which is mentioned just a few lines below

L286: This should probably be Bx, not Bz?

yes, corrected

L299: Since the cluster numbering is arbitrary, you could perhaps re-order the colors to match the earlier ordering to assist the reader.

The colors of the clusters will be re-ordered to match F1.

Figure 7: What time value is this? (for the caption, not just the text)

$t_0 + 210$ minutes. It will be made clear in text and caption.

L337: Please clarify if these are trained with t_0+210 or with mixed time data?

They are trained with $t_0 + 210$ minutes data. This will be clarified in the manuscript.

Figures 8 and 9, main text: The ordering, going from the top row of Fig8 to two rows of Fig9, back to the bottom row of Fig8, and then to the last row of Fig9 is counter-intuitive. This could surely be improved.

The labelling of the feature sets will be changed and the description of classification results with different features improved.

L367-368: Please briefly mention what F7 and F8 added

What F7 and F8 (old feature set labelling, F4 and F5 with the new labelling) add to the the feature list is mentioned in the paragraph before L367 in the old manuscript. If the comment refers instead to the rationale for adding F7 and F8 to the feature list, the idea is to verify if quantities which are meaningful for the human observer, such as Mach number, Alfven speed, plasma beta, hinder or help the unsupervised classification. As per the tests presented, the difference is not major.

L416: More comparison with Amaya(2020) of the results and potential future avenues would be good - it was only very briefly cited before.

This will be added to the conclusions:

In Amaya 2020, more advanced pre-processing techniques were experimented with, which will most probably prove useful when we will move to the more challenging environment of spacecraft observations (as opposed to simulations). Furthermore, Amaya 2020 employed windows of time in the classification, which we have not used in this work in favor of an "instantaneous" approach. In future work, we intend to verify which approach gives better results.

L447: Do you have any references to indicate what direction non-linear feature correlation analysis in deciding a dimensionality reduction could take? What about the addition of non-local features, such as spatial and temporal derivatives, curls etc?

We will try non-linear feature correlation analysis as part of future work. As such, we do not think we are in a position to speculate on results of this activity at this point in time. Regarding the addition of non local features: in this work, we preferred a local, instantaneous approach, which gave good results with identification of large-scale magnetospheric regions. It may be unavoidable to introduce spatial and temporal derivatives if we decide to use a similar approach for the identification of boundaries between magnetospheric regions, or meso- and small-scale processes.

L448: Are dynamic SOMs a reasonable approach to automatic classification, or do they require user validation after every re-learning?

If the learning parameters are the same, the net casted by the DSOM has the same rigidity for any random initial- ization of the initial code word (node) locations. The final

maps can potentially show slightly different arrangements, but as a net the SOM will cover the same area of the N-dimensional space. This has not been tested thoroughly in this paper but will be added to our future analysis.

Please also note the supplement to this comment:

<https://angeo.copernicus.org/preprints/angeo-2021-33/angeo-2021-33-AC1-supplement.zip>