Appendix B: Classification evolution with the cluster number

In this Section, we explore how the classification of magnetospheric regions changes with the number of K-means clusters k. In all the cases described here (as in the rest of the paper, unless otherwise specified), the SOM map is that obtained with the F1 features in Table 1 and with $q = 10 \times 12$, $\epsilon_0 = 0.25$, $\sigma_0 = 1$. The $k = 7$ case is better described in Section 4.1.

In Figure B1, we show the classification results with $k = 6$ (panel a, b) and $k = 5$ (panel c and d). The $k = 7$ case is depicted in Figure 4. In panel a, c and b, d we depict the simulated meridional and equatorial plane, with the data points colored according to their respective class. Following the changes of the class maps from $k = 7$ to $k = 5$ allows us better insight into the results of our classification procedure, and also highlights which of the magnetospheric regions are more similar in terms of plasma parameters.

In panel a and b we plot $k = 6$. Comparing them with Figure 4, we see that reducing the number of clusters of one unit merges the three magnetosheath clusters, brown, orange and blue, into two, blue and orange. This is consistent with the fact that the magnetosheath clusters, and in particular cluster 4 (brown) and 1 (red) with $k = 7$, map to quite similar plasmas. The more “internal" clusters (inner magnetosphere, boundary layers, lobes) are not affected.

Further reducing to number of clusters to $k = 5$, instead, affects the “internal" clusters. The boundary layer cluster disappears, and the points that mapped to it are, quite sensibly, assigned to the clusters mapping to inner magnetospheric (this is mostly the case of current sheet plasma), magnetosheath or lobe plasma. Some inner magnetosphere points are mis-classified as magnetosheath plasma. With both $k = 6$ and $k = 5$ the solar wind cluster, that differs the most from the others, is left unaltered.

In Figure B2, we depict the pure K-means classification with $k = 6$ (panel and b) and $k = 5$ (panel c and d) for data points in the meridional (panel a and c) and equatorial (panel b and d) planes, to be compared with the SOM classification depicted in Figure B1.

We do not notice any significant difference between the SOM and K-means classification for $k = 5$; for $k = 6$ we notice, as already with $k = 7$, that the SOM classification reduces the mis-classification of internal magnetospheric points.

In summary, decreasing k from a large to smaller produces a more coarse-grained classification. Generally speaking, every time k is decreased, the three clusters mapping to the most similar plasma reorganize and coalesce into two. This process shows which magnetospheric regions are most similar.

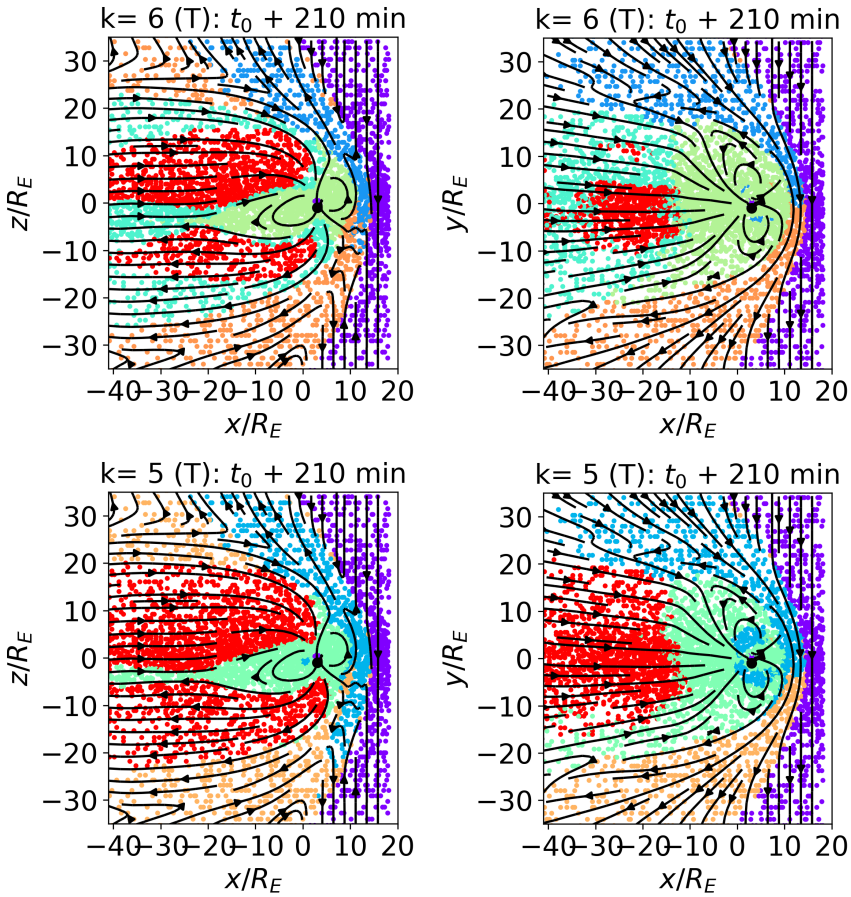

Figure B1. y/R_E= 0 (panel a and c) and z/R_E= 0 (panel b and d) cuts with $k = 6$ (panel a, b) and $k = 5$ (panel c, d).

Author contributions. M.E.I. run the simulation and performed the analysis. J.A. assisted in the analysis. J.R. and B.F. provided the OpenGGCM-CTIM-RCM code and support in using it. R.D. and G.L. supported the investigation and provided useful advice.

Competing interests. No competing interests are present

Acknowledgements. The simulation data set (“KUL_OpenGGCM") is available at Cineca AIDA-DB. In order to access the meta-information and the link to "KUL_OpenGGCM" simulation data look at the tutorial athttp://aida-space.eu/AIDAdb-iRODS.

We acknowledge funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA, Artificial Intelligence for Data Analysis, www.aida-space.eu). Work at UNH was also supported by grant AGS-1603021 from the National Science Foundation and by award FA9550-18-1-0483 from the Air Force Office of Scientific Research.

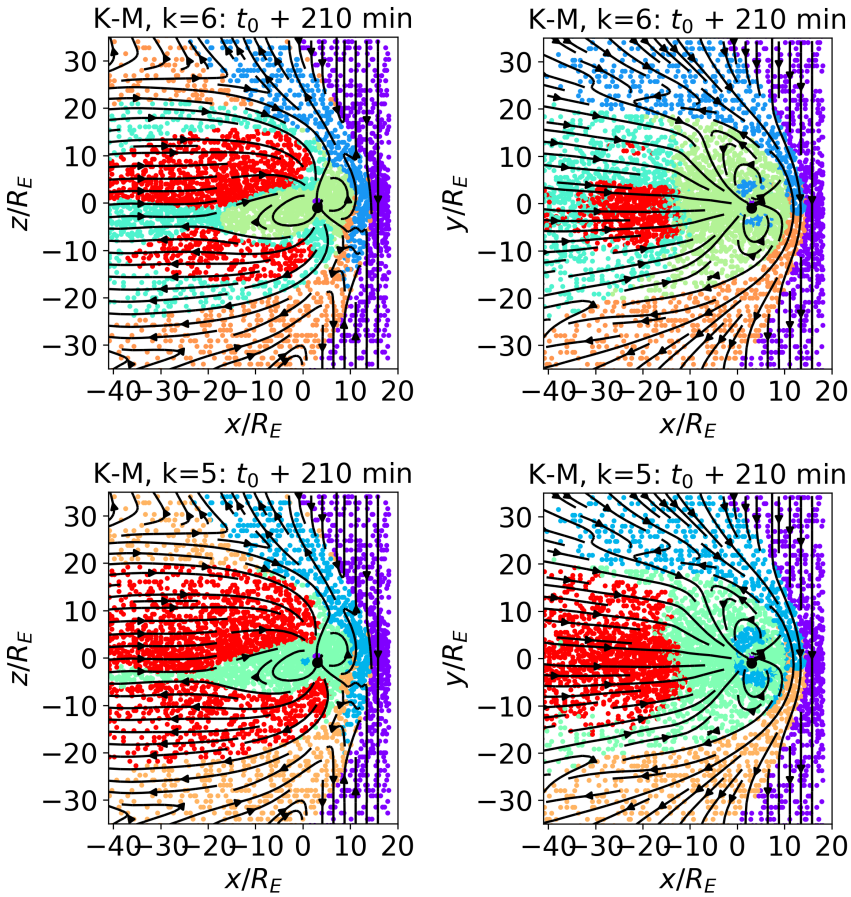

Figure B2. K-means classification with $k = 6$ (panel and b) and $k = 5$ (panel c and d) for the data points in the meridional (panel a and c) and equatorial (panel b and d) planes.

The OpenGGCM-CTIM-RCM simulations were performed on the supercomputer Marconi-Broadwell (Cineca, Italy) under a PRACE allocation.

We acknowledge the use of the MiniSom Vettigli, scikit-learn Pedregosa et al. (2011), pandas, matplotlib python packages.