Dimension Reduction Stacking for Deep Solar Wind Clustering

Daniel T. Carpenter[1], Henry Han[2], and Liang Zhao[1(✉)]

[1] Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA
{dcar,lzh}@umich.edu
[2] Department of Computer Science, School of Engineering and Computer Science, Baylor University, Waco, TX 76798, USA
henry_han@baylor.edu

Abstract. In-situ observations of solar wind plasma exhibit statistical differences according to their coronal origins. These in-situ conditions are a direct result of various processes such as ionization and acceleration occur in the inner corona. Machine learning methods have been successful in characterizing solar wind in-situ observations using unsupervised deep clustering and dimensionality reduction techniques, but it remains unclear as to how solar wind data embedding and downstream clustering could be improved while providing better interpretability in machine learning process. In this study, we explore the impact of distance metrics on solar wind in-situ data clustering. We evaluate the metric performance by applying it to dimension-reduction-stacking and deep clustering techniques and comparing it with state-of-the-art methods using solar wind in-situ measurements. Our work demonstrates the potential for customized distance metrics to improve the interpretability and performance of deep clustering approaches applied in solar wind in-situ observations.

Keywords: Solar wind · Classification · Machine Learning · Heliophysics

1 Introduction

The solar wind is the stream of supersonic ionized particles released from the Sun, and drives *space weather* at Earth's geo-space environment. Space weather impacts Earth's climate, satellite communication, power grids, and other domains important to life on the surface. The physical processes occurring in the base of the solar corona that ionize, heat, and accelerate the solar wind plasma are of central importance to space weather forecasting and the ways in which the Sun affects the Earth.

One way we observe the solar wind is through in-situ measurements (e.g., ACE and Ulysses missions), which are inextricably tied to conditions in the solar corona–the outermost part of the sun's atmosphere–where the solar wind

H. Han and E. Baker (Eds.): SDSC 2023, CCIS 2113, pp. 111–125, 2024.
https://doi.org/10.1007/978-3-031-61816-1_8

originates. The in-situ properties of solar wind can be linked back to the coronal structures where it originates [24]. These properties encapsulate the dynamic and thermal properties of the bulk solar wind plasma (protons), minor constituents (Helium and heavier ions, such as Carbon, Nitrogen, Oxygen, Neon, Magnesium, Silicon, Sulfur, Iron, etc.), and magnetic field associated with them. Properties associated with these so-called heavy ions include total abundances and charge states, either as the average charge state of a specific element, or by the ratio of densities of individual charge states (such as O^{7+}/O^{6+}, C^{6+}/C^{5+}, measured by SWICS instrument onboard ACE and Ulysses).

There are certain solar wind structures identifiable by in-situ measurements. For example, plasma originating from coronal hole (CH) regions is usually observed to have high proton speeds: the aptly named fast wind. The plasma from CHs also has lower charge state ratios, indicating cooler electron temperatures and low plasma densities in the coronal origins [4,26]. This is in contrast of the slow solar wind, whose coronal origins are more difficult to ascertain from in-situ properties. The slow solar wind could come from anywhere from the periphery of active regions (ARs) [11,13], the helmet streamers [21,22], or the pseudostreamers [5,19,23], and so forth. The wind from these source regions is more ionized, which indicates hotter electron temperatures and higher densities in their coronal sources. Solar wind can also be attributed to occasional transient events which are important to space weather, such as Interplanetary Coronal Mass Ejections (ICMEs), which are energetic eruptions originating in magnetically active regions [7,8].

The Challenges of Solar Wind Classification. The solar wind has been traditionally classified according in-situ physical properties via statistical means; however, there are at least three challenges that arise when attempting to use in-situ properties to assign different types of solar wind to specific coronal sources: 1) singular observable, such as proton speed, is of poor use as a categorization metric (slow wind can arise from CHs [2,9,17,20,22]); 2) in-situ solar wind speeds and composition are on a continuum [27], and 3) the dimensionality of the data limits how the behavior can be visualized (there are 77 different parameters related to the heavy ion composition and elemental abundances alone ACE/SWICS [6]).

Unsupervised learning methods and dimensionality reduction algorithms have already proven effective at answering these challenges as data-driven characterization schemes [1,3,15]. However, an approach has yet to be shown which minimizes instances of subjectivity in parameter selection and explains how the downstream embedding and clustering results are delivered. In this work, we detail the appropriate domain knowledge for solar wind data and introduce a novel deep clustering approach, PCA+t-SNE+DBSCAN, for characterizing the solar wind using in-situ properties. This method utilizes our dimension reduction stacking technique, PCA+t-SNE, proposed in this study, along with various distance metric probing, to effectively and more transparently identify solar wind clusters. To the best of our knowledge, this is the first explainable machine learning method proposed for the clustering of solar wind data. The proposed

dimension reduction stacking can also be extended to other AI and data science fields.

2 Dimension Reduction Stacking

The application of dimension reduction stacking is novel to the field of solar physics. Dimension reduction stacking is the technique of combining reduction methods by using the output of one method as the input of another. Formally, the original input data $X = \{x_i\}_{i=1}^{N}$, where $x_i \in \mathbb{R}^M$, is mapped to a low-dimensional representation $Y = \{y_i\}_{i=1}^{N}$, where $y_i \in \mathbb{R}^l$ and $l \ll p$. The stacking results in a composite function $f(g(X)) \rightarrow Y$. Typically, f and g belong to different types of dimension reduction methods, and their combination enables the extraction of features at a deeper level by having one method address the limitations that another has in the stack.

2.1 Principal Component Analysis

The first dimension reduction method we used in our stack is Principal Component Analysis (PCA). PCA focuses on minimizing information loss by creating new uncorrelated variables that successively maximize variance. These variables are called principal components (PCs), and they are the solutions to an eigenvalue/eigenvector problem of the covariance matrix of the input data. The quality of the reduction can be measured using the variability associated with the set of retained PCs. To measure the quality the amount of selected PC's, the cumulative explained variance percentage is calculated. PCA, being a classic holistic method, is capable of extracting the global behavior of the data; however, PCA usually cannot capture the local behavior of the data because each PC only contains some levels of global characteristics of the data.

2.2 t-Distributed Stochastic Neighboring with Embedding

The second dimension reduction technique in the stack is the t-Distributed Stochastic Neighboring with Embedding (t-SNE). This is a non-linear method, capable of capturing local data behaviors in the dimensionality reduction. The t-SNE algorithm minimizes the Kullback-Leiber (K-L) divergence between a distribution P and student t-distribution Q to achieve the low-dimensional embedding by solving a non-convex optimization problem. The Gaussian and t-distributions model the pairwise similarities between data points in the original high-dimensional input space and embedding. t-SNE aims to force a similarity of the embedding data to the original data by seeking the minimum K-L distance between P and Q. By having the PCA transformation emphasize the global structure of the data, and t-SNE generate an embedding that captures local structure in the data (in multiple scales), the data expression can be represented in two dimensions in a way that considers both global and local structures. Once the data is in this form, the embedding can be clustered in order to identify attributes of the clusters.

2.3 PCA+t-SNE Dimension Reduction Stacking

The dimension reduction stacking is done by composing PCA and t-SNE, titled PCA+t-SNE. We project the input data onto the PCA space while ensuring a minimum of 95% total explained variance. Subsequently, reduction is used to compute distance matrices. These distance matrices will then be fed into t-SNE. It is worth mentioning that PCA+t-SNE outperforms PCA+UMAP in terms of performance, as t-SNE has a better ability to capture local behavior of the data compared to UMAP [10]. This observation is further supported by our downstream DBSCAN clustering results.

Algorithm 1: Dimension Reduction Stacking (X,η,p)

Input:
  The input data: $\mathrm{X} \in \mathbb{R}^{N \times M}$.
  The explained variance ratio: η.
  The perplexity in t-SNE: p.
Output:
  The dimension reduction stacking embedding: $\mathrm{X}_{PCA+TSNE}$.
Begin. PCA for the scaled data:
  $\mathrm{X}_{PCA}, pcVariance \leftarrow$pca(X).
  Retrieve reduction:
  IF $\sum_{i=1}^{i=l} pcVariance_i \geq \eta$
    $\mathrm{X}_{embedding} \leftarrow \mathrm{X}_{PCA}[:,1{:}l]$.
  Compute pairwise distances:
    $\mathrm{D}_X \leftarrow f_{dist}(\mathrm{X}_{embedding})$
  t-SNE embedding:
    $\mathrm{X}_{PCA+TSNE} \leftarrow$tsne($\mathrm{D}_X$,p)
  Return: $\mathrm{X}_{PCA+TSNE}$.
End

3 Data and Preprocessing

We use the Advanced Composition Explorer (ACE) spacecraft as the platform for our data. ACE is positioned at the L1 point, measuring the solar wind plasma and interplanetary magnetic field since 1998. A subset of the data from 2000–2002 was chosen, as because the heightened solar activity in this time range resulted in more frequent equatorial CHs and ICMEs, providing in a more balanced inclusion of solar wind sources in this interval. From this range, we use a random subset of 2500 2-hour-binned measurements.

To create the data expression for reduction, we collect 12 variables linked to solar wind in-situ signatures: from the Solar Wind Electron, Proton, and Alpha

Monitor (SWEPAM) [14], we use the proton temperature (T_p) and proton density (n_p) to compute the proton entropy ($S_p = T_p\, n_p^{-2/3}$), and we also include alpha-to-proton ratio (denoted as α/H). From SWICS [9], we include the elemental composition (relative abundances of Magnesium, Silicon, Iron, Carbon, Neon, and Helium to Oygen, denoted as Mg/O, Si/O, Fe/O, C/O, Ne/O, and He/O respectively) and heavy ion composition signatures of Oxygen, Carbon and Iron (O^{7+}/O^{6+}, C^{6+}/C^{4+}, C^{6+}/C^{5+}, and $\langle Q_{Fe}\rangle$).

After removing samples with null values in the original data set and taking the random subset (as described previously), the data variables (described above) are scaled using a Min-Max scaling along each dimension.

4 Solar Wind Deep Clustering Under Dimension Reduction Stacking

We generate a meaningful embedding space for input solar wind data that reveals both latent global and local data characteristics through PCA+tSNE dimension stacking, before seeking meaningful similarity via density-based clustering. PCA is applied as the first dimensionality reduction technique. The left panel of Fig. 1 shows the percentage of explained variance (blue) and accumulated explained variance (red) by each PC. The subset of PCs up to PC_7 explains 95% of the original data variance. All of the data are visualized in the first two PC components frame through a Gaussian Kernel Density Estimate (KDE) in the right panel of Fig. 1.

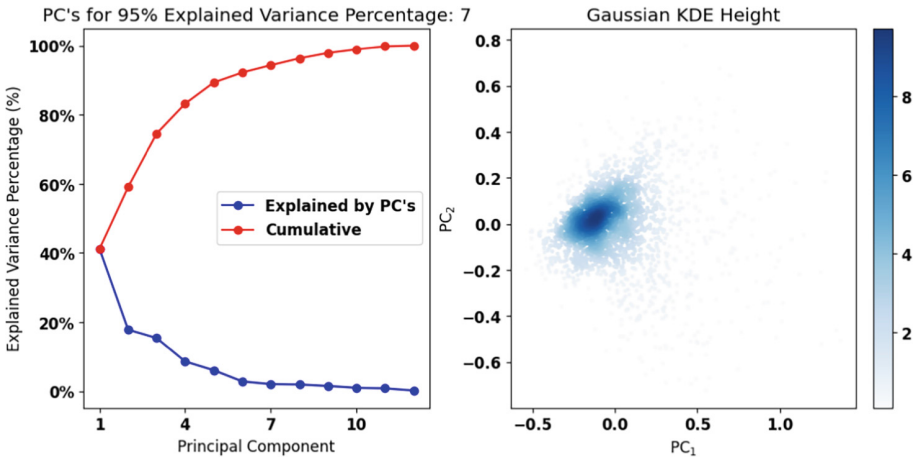

Fig. 1. Result of PCA on the in-situ solar wind data. The left panel shows the percentage of variance (blue) and cumulative sum (red) explained by each PC. The middle panel shows the eigenvalues of each PC. The right plot visualizes the Gaussian KDE height of the first two PCs. (Color figure online)

In our implementation, we utilize the projected data of the original solar wind dataset in the PCA subspace, maintaining a 95% explained variance ratio, for calculating pairwise distance matrices using the Euclidean, Cosine, and Mahalanobis distance metrics respectively. These pre-computed distance matrices are then employed as inputs for t-SNE to generate the t-SNE embedding. This PCA+tSNE dimension reduction stacking approach captures both the global and local characteristics of the data in dimension reduction, in addition to a denoising procedure. Moreover, this stacking method yields a meaningful embedding space for exploring similarities in solar wind data, which benefits subsequent density-based clustering, such as DBSCAN. DBSCAN is robust to noise and adaptable to data of any shape, making it suitable for the noisy, nonlinear solar wind data that can demonstrate any shape after dimension reduction stacking. Deep: PCA+t-SNE+DBSCAN here means an in-depth exploration of latent global and local data characteristics revealed in the latent embedding generated from PCA+t-SNE stacking, as well as the examination of various similarity metrics in PCA+t-SNE stacking.

DBSCAN (Density Based Spatial Clustering of Applications with Noise) is a density-based clustering algorithm designed to cluster data of arbitrary shape, and to account for noise. DBSCAN classifies points as either core, reachable, or outlier (noise). Core points are within a radius ε of a neighborhood–the size of the neighborhood is specified by a minimum number of points, including the point in question. A reachable point is within radius ε of a core point, but does not neighbor the required minimum number of points. An outlier is a point which is beyond ε of any core point. In our context, outliers will be solar wind samples which have physical qualities which differ enough from the main groups of solar wind. The core points form clusters because of their high densities, while the reachable points form the edge of said clusters and the outliers stand out as noise. DBSCAN will allow us to find the boundaries of clusters without imposing any model or numerical restrictions. Once these cluster labels are generated, we can then project the labels back onto the original data in order to examine the underlying physics of the solar wind clusters.

4.1 Deep Clustering Under Different Distance Metrics

We employ three distance metrics-Euclidean, cosine, and Mahalanobis-in PCA+t-SNE stacking before DBSCAN clustering. This implies that these distances are utilized to calculate the pairwise distance matrix using data projected into the PCA subspace

The corresponding PCA+t-SNE+DBSCAN clustering results are shown in Figs. 2, 3, and 4. In the DBSCAN clustering process, we set the minimum number of points for a cluster to be considered a core cluster to 75 points. We then vary epsilon for each embedding until we see some of the smaller scale clusters. The epsilon values are shown in Table 1. The clusters are projected onto physical parameters associated with each data point.

The first distance metric we examine is the Euclidean metric, resulting in 6 clusters and one outlier group (Fig. 2). The winds in Class 2 have very low

Table 1. Selected epsilon values for each distance metric

Distance Metric	Euclidean	Cosine	Mahalnobis
Epsilon	11.0	10.3	11.2

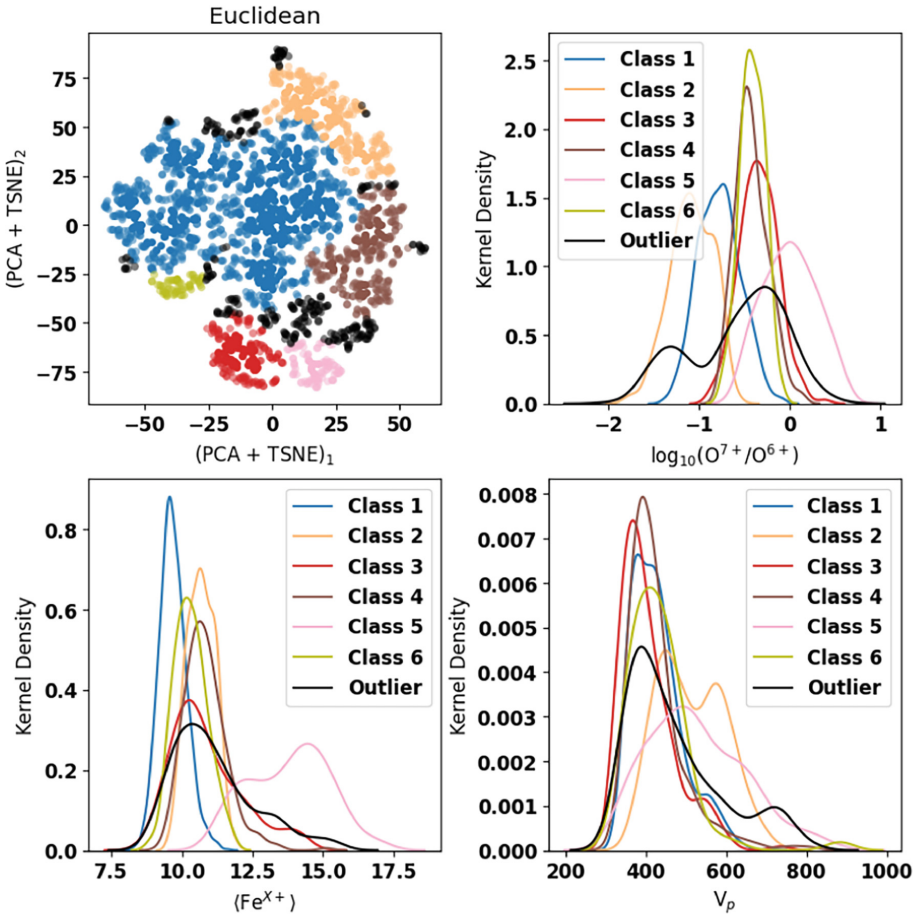

Fig. 2. Embedding and clustering using Euclidean distance as the metric in the PCA+TSNE process (top left). Class labels are assigned by DBSCAN, and are projected onto the charge state ratio of oxygen (O^{7+}/O^{6+} ratio), the average charge state of Iron, and the bulk proton speed (top right, bottom left, and bottom right, respectively).

O^{7+}/O^{6+} and relatively high V_p, indicative of equatorial coronal hole associated fast wind. Class 3, 4, and 6 have relatively high O^{7+}/O^{6+} and slow proton speed, indicative of typical streamer-associated slow wind [22]. The O^{7+}/O^{6+} of 0.145 (logarithm base 10 value is about −0.84) was found by previous work [22] that

can effectively separate the streamer-associated slow solar wind from the coronal-hole-associated fast wind; and here we see that the coronal hole (cluster 2) and slow wind (clusters 3, 4, 6) are separated by a very similar value of O^{7+}/O^{6+}. Class 5 is an interesting class; it has characteristics similar to slow solar wind; however, it has higher wind speeds and very high average iron charge states. These are consistent with the characteristics of ICMEs passing by a spacecraft. It appears that clustering the solar wind using this Euclidean distance metric can be useful to extract ICMEs, and separated the non-ICME solar wind into distinct sub-groups, such as coronal-hole associated fast and streamer-associated slow wind.

The result based on Cosine distance metric is shown in Fig. 3. There are two very similar slow wind clusters (2 and 6), possessing relatively high O^{7+}/O^{6+} ratio and slow proton speed. These two clusters are likely the streamer-associated slow solar wind. Oppositely, in cluster 5 the solar winds have low O^{7+}/O^{6+} ratios and relatively high proton speed, those winds are more likely contributed by the low latitude coronal-hole associated fast wind. Interestingly, the separation point of O^{7+}/O^{6+} between these two different coronal-originated solar winds is still located near the value of 0.145 (logarithm base 10 value is about -0.84), consistent with the result of Euclidean distance and the previous study [22]. The O^{7+}/O^{6+} ratios of the clusters 1 and 3 are in between the streamer slow wind (cluster 2 and 6) and coronal-hole fast wind (5), indicating that they may be some combination of these two types of winds. In the average charge state of Iron plot, the relatively high value of Iron charge state in the class 4 implies that some winds in this class may contain the ICME plasma, but not all of them are ICMEs.

The result based on Mahalanobis distance is shown in Fig. 4. The solar wind in class 6 possesses the lowest O^{7+}/O^{6+} ratio, and relatively high proton speed, indicating that this class is more likely to be the fast wind originated from equatorial coronal holes. Class 5 is characterized as having the highest averaged charge state of Iron, indicative of a group of ICME winds. Class 3 possesses relatively high O^{7+}/O^{6+} ratio (mostly higher than 0.145) and relatively slow proton speed, consistent with the features of typical streamer-associated slow wind. Classes 1, 2 and 4 seem to posses the majority of the data, however their O^{7+}/O^{6+} ratio, average charge state of Iron and proton speed are in the moderate ranges which implies that they are probably some combinations of the coronal-hole and streamer winds, therefore they cannot be assigned to any specific solar wind types or coronal origins.

Maximum Fusion Distance Metric. The clustering results calculated using the three different distance metrics have different strength and weakness in relating them with the solar wind features and the coronal origins. It is hard to determine which distance metrics is the best one among these three. Therefore, in the next step, we create a new distance metric which is designed to prioritize the effectiveness of all of the three metrics, in order to obtain the optimized solar wind clustering result.

The new distance metric is defined as the maximum of the three normalized metrics as calculated previously (Eq. 1). We name this new distance metric Maximum Fusion.

$$p_{max} = \max\left(\frac{p_{euc}}{\max(p_{euc})}, \frac{p_{cosine}}{\max(p_{cosine})}, \frac{p_m}{\max(p_m)}\right) \quad (1)$$

We then apply the same t-SNE parameters and clustering process, retaining MinPts as 75 and choosing epsilon to reveal multiple scales of clusters. The result of the clustering is shown in Fig. 5. It is clear that the solar wind clusters are in general separated at the threshold of $O^{7+}/O^{6+} = 0.145$ (logarithm base 10 value is -0.84), with class 2 and 5 as hotter (O^{7+}/O^{6+} ratio > 0.145, streamer-

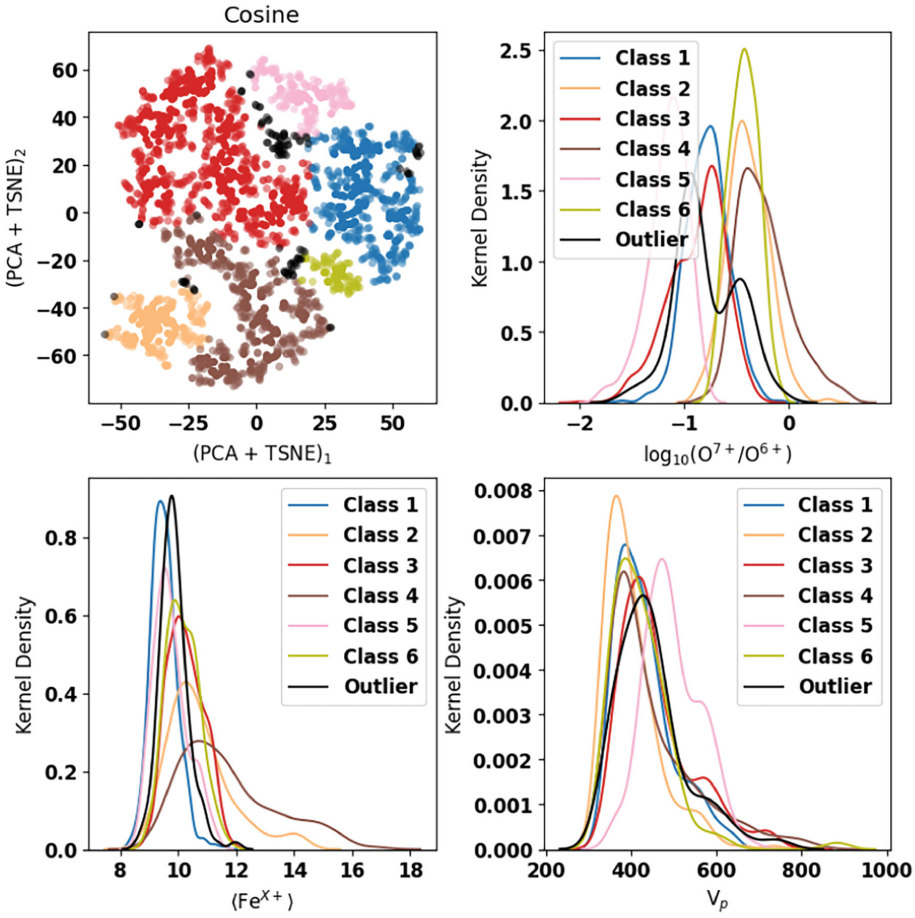

Fig. 3. Embedding and clustering using Cosine distance as the metric in the PCA + TSNE process (top left). Class labels are assigned by DBSCAN, and are projected onto the charge state ratio of oxygen (O^{7+}/O^{6+} ratio), the average charge state of Iron, and the bulk proton speed (top right, bottom left, and bottom right, respectively).

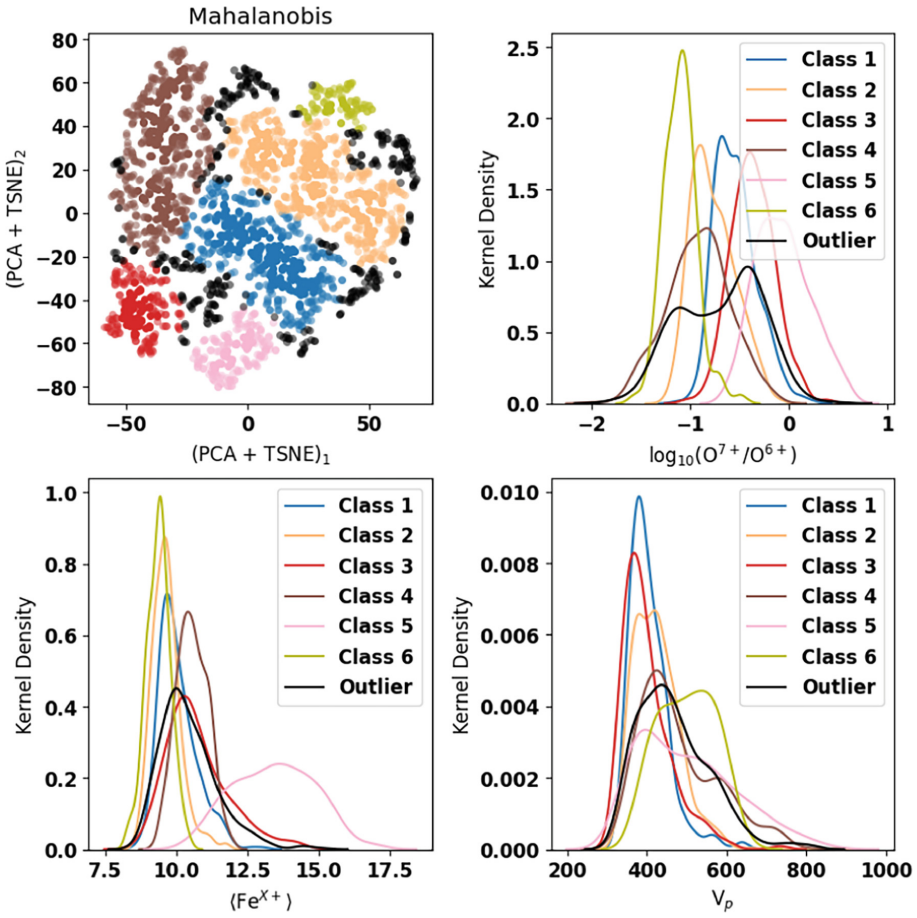

Fig. 4. Embedding and clustering using Mahalanobis distance as the metric in the PCA+TSNE process (top left). Class labels are assigned by DBSCAN, and are projected onto the charge state ratio of oxygen (O^{7+}/O^{6+} ratio), the average charge state of Iron, and the bulk proton speed (top right, bottom left, and bottom right, respectively).

associated) and slower wind and class 3 as colder (O^{7+}/O^{6+} ratio < 0.145, coronal-hole associated) and faster wind; meanwhile class 1 which possesses the majority of the data points, stays in between. Particularly, the winds in class 4 have the highest O^{7+}/O^{6+} ratios and average charge state of Iron, indicating that they are likely ICMEs. These five clusters in Fig. 5 are more explainable and better understandable in terms of the solar wind features and the coronal origins than the results from the other three distance metrics.

5 Conclusion

In this work, we find that PCA+t-SNE can characterize in-situ solar wind data in a way that reveals the hidden features in the data. We then present the solar wind clustering results using DBSCAN with three distance metrics, Euclidean, Cosine, and Mahalanobis, and a new distance metrics that we invent, Max Fusion. We describe the impact that these metrics have on the solar wind clustering, and compare the clustering results (Table 2).

Table 2. Comparison of the solar wind clustering results among different distance metrics

	slow wind (high O^{7+}/O^{6+})	fast wind (low O^{7+}/O^{6+})	CME (high O^{7+}/O^{6+} and Q_{Fe})	undefined
Euclidean	3 classes (3,4,6)	1 class (2) wide speed range	1 class (5)	1 class (1)
Cosine	2 classes (2,6)	1 class (5) wide speed range	not very clear	2 classes (1,3)
Mahalanobis	1 class (3)	1 class (6) wide speed range	1 class (5)	3 classes (1,2,4)
Max Fusion	2 classes (2,5)	1 class (3)	1 class (4)	1 class (1)

The comparison of the clustering results show that all of the four distance metrics can produce solar wind clusters that are indicative of streamer-associated slow solar wind (slow speed and high O^{7+}/O^{6+} ratio). Euclidean and Mahalanobis distance metrics can also produce a cluster of solar wind that matches the characteristics of ICMEs (high O^{7+}/O^{6+} and average charge state of Iron), but Cosine distance metrics fails to produce a cluster of solar wind that can be exclusively considered as ICMEs. All of the first three distance metrics result in a cluster of solar wind that possesses low O^{7+}/O^{6+} ratio solar wind, and has a proton speed range spanning from 300 to 700 Km/s with two peaks, indicating that this cluster may be partially contributed by coronal-hole associated winds, but also with some contamination from other slow-speed types of winds. Differently, the clusters identified using Max Fusion distance metrics show clear different types of solar wind: two clusters are slow, streamer-associate wind; one cluster is coronal-hole associate fast wind; and one cluster is like ICME wind. Therefore, we conclude that the Max Fusion distance metrics so far is the best distance metrics that can effectively classify the solar wind into different categories of physically distinct characteristics and indicative of different coronal origins. We summarize the comparison of these clustering results in Table 2.

6 Discussion

The MinPts parameter is essential in defining core points and influencing cluster density in the DBSCAN algorithm, particularly for large, noisy datasets such as

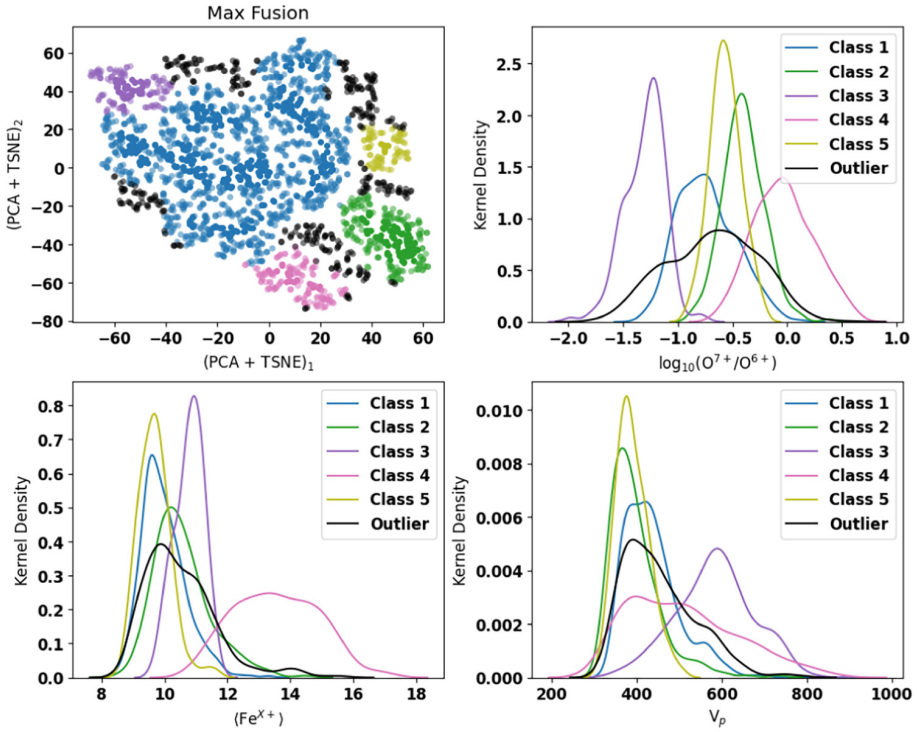

Fig. 5. Embedding and clustering using max-fusion distance as the metric in the PCA+TSNE process (top left). Class labels are assigned by DBSCAN, and are projected onto the charge state ratio of oxygen (O^{7+}/O^{6+} ratio), the average charge state of iron, and the bulk proton speed (top right, bottom left, and bottom right, respectively).

solar wind data. This becomes even more significant when applied to embedding data derived from the proposed dimension reduction stacking (PCA+tSNE). Our study, as shown in the left panel of Fig. 6, illustrates how the Eps (epsilon) value fluctuates in relation to the Euclidean distance metric in the embedding. Technically, it refers to the radius of a neighborhood around a solar wind projection point in the embedding space.

A lower MinPts setting results in a restricted epsilon range and a higher cluster count, whereas a very high MinPts leads to underfitting, as indicated by a reduced number of clusters. Interestingly, the left side of these curves tends to display a consistent cluster count across various MinPts values, though this can cause overfitting with many outliers. To counteract this, a higher MinPts value enables DBSCAN to concentrate on fewer, larger-scale clusters, a crucial approach in our study to prevent the formation of overly small clusters and ensure the applicability of our findings to a wide range of solar wind datasets. Therefore, we have selected a MinPts value of 75 for our analysis. However, it's important to note that this empirical choice may not be the best fit for generalization in larger solar wind datasets.

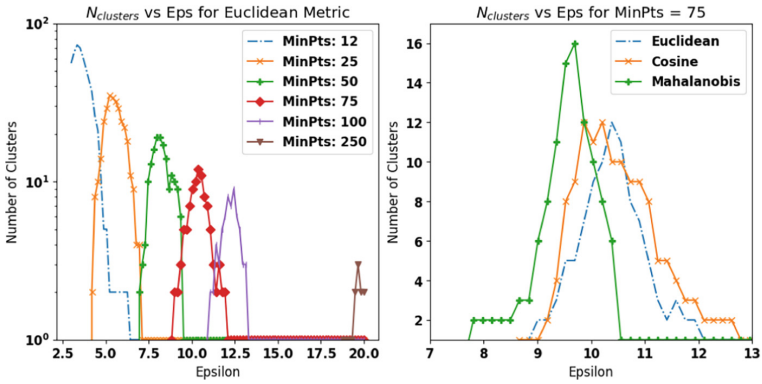

Fig. 6. The number of clusters that result from DBSCAN clustering with increasing epsilon. The left panel shows varying MinPts on the Euclidean distance metric. The right panel shows various distance metrics on MinPts = 75.

Moreover, our previous findings indicate that the epsilon values are relatively similar across different distance metrics. As depicted in the right panel of Fig. 6, there is a noticeable overlap in cluster numbers when using both Euclidean and Cosine distance metrics, in contrast to the Mahalanobis metric, which shows peak values at lower epsilon levels. This observation suggests that the choice of epsilon not only depends on the scale of the structures under study but also varies in its impact across different distance metrics within the embedding space. To address this, we are developing an optimal parameter tuning strategy for large-scale solar wind data analysis. This approach involves utilizing various distance metrics and a probing learning method that leverages a small subset of the data for initial insights. Additionally, we are incorporating distance metric learning techniques, including Siamese and Triplet Networks, to uncover more meaningful and insightful solutions in our analysis [28].

Acknowledgements. The work of D.C. is supported by NASA grant 80NSSC 22K1015. L.Z. is supported by NASA Grants 80NSSC21K0579, 80NSSC22K1015, NSF SHINE grant 2229138, and NSF Early Career grant 2237435. H.H is supported by the McCollum endowed chair startup fund of Baylor University.

References

1. Bloch, T., Watt, C., Owens, M. et al.: Data-driven classification of coronal hole and streamer belt solar wind. Sol. Phys. **295**(41) (2020) https://doi.org/10.1007/s11207-020-01609-z
2. Bravo, S., Stewart, G.A.: Fast and slow wind from solar coronal holes. Astrophys. J. **489**, 992 (1997)
3. Carpenter, D., Zhao, L., Lepri, S.T., Han, H.: Characterizing in-situ solar wind observations using clustering methods. In: Han, H., Baker, E. (eds.) SDSC 2022. CCIS, vol. 1725, pp. 125–138. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-23387-6_9

4. Cranmer, S.R.: Coronal holes and the high-speed solar wind. Space Sci. Rev. **101**, 229 (2002)
5. Crooker, N.U., Antiochos, S.K., Zhao, X., Neugebauer, M.: Global network of slow solar wind. J. Geophys. Res. **117**, A04104 (2012). https://doi.org/10.1029/2011JA017236
6. Garrard, T., Davis, A., Hammond, J., et al.: The ACE science center. Space Sci. Rev. **86**, 649–663 (1998)
7. Gibson, S.E., Fan, Y.: The partial expulsion of a magnetic flux rope. Astrophys. J. **637**(1), L65–L68 (2006)
8. Gibson, S.E., Fan, Y., Török, T., Kliem, B.: The evolving sigmoid: evidence for magnetic flux ropes in the corona before, during, and after CMEs. Space Sci. Rev. **124**(1–4), 131–144 (2006)
9. Gloeckler, G., Cain, J., Ipavich, F.M., et al.: Investigation of the composition of solar and interstellar matter using solar wind and pickup ion measurements with SWICS and SWIMS on the ACE spacecraft. Space Sci. Rev. **86**, 497 (1998)
10. Han, H., Wentian, L., Wang, J., Qin, G., Qin, X.: Enhance explainability of manifold learning. Neurocomputing **500**, 877–895 (2022)
11. Ko, Y.-K., Raymond, J.C., Zurbuchen, T.H., et al.: Abundance variation at the vicinity of an active region and the coronal origin of the slow solar wind. Astrophys. J. **646**, 1275 (2006)
12. Lepri, S.T., Zurbuchen, T.H.: Iron charge state distributions as an indicator of hot ICMEs: possible sources and temporal and spatial variations during solar maximum. J. Geophys. Res. **109**, A01112 (2004). https://doi.org/10.1029/2003JA009954
13. Liu, S., Su, J.T.: Multi-channel observations of plasma outflows and the associated small-scale magnetic field cancellations on the edges of an active region. Astrophys. Space Sci. **351**, 417 (2014)
14. McComas, D., Bame, S., Barker, P., et al.: Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. Space Sci. Rev. **86**, 563–612 (1998)
15. Roberts, D.A., et al.: Objectively determining states of the solar wind using machine learning. ApJ **889**, 153 (2020)
16. Smith, C., L'Heureux, J., Ness, N., et al.: The ACE magnetic fields experiment. Space Sci. Rev. **86**, 613–632 (1998)
17. Stakhiv, M., Landi, E., Lepri, S.T., Oran, R., Zurbuchen, T.H.: On the origin of mid-latitude fast wind: challenging the two-state solar wind paradigm. Astrophys. J. **801**, 100 (2015)
18. Wang, Y.-M., Ko, Y.-K.: Observations of slow solar wind from equatorial coronal holes. Apj **880**, 146 (2019)
19. Wang, Y.-M., Grappin, R., Robbrecht, E., et al.: On the nature of the solar wind from coronal pseudostreamers. Astrophys. J. **749**, 182 (2012)
20. Wang, Y.-M., Ko, Y.-K., Grappin, R.: Slow solar wind from open regions with strong low-coronal heating. Astrophys. J. **691**, 760 (2009)
21. Zhao, L., Landi, E., Zurbuchen, T.H., Fisk, L.A., Lepri, S.T.: The evolution of 1 AU equatorial solar wind and its association with the morphology of the heliospheric current sheet from solar cycles 23 to 24. Astrophys. J. **793**, 44, 8 pp (2014). https://doi.org/10.1088/0004-637X/793/1/44
22. Zhao, L., Zurbuchen, T.H., Fisk, L.A.: Global distribution of the solar wind during solar cycle 23: ACE observations. Geophys. Res. Lett. **36**, L14104 (2009)
23. Zhao, L., Gibson, S.E., Fisk, L.A.: Association of solar wind proton flux extremes with pseudostreamers. J. Geophys. Res. Space Phys. **118**, 2834–2841 (2013)

24. Zhao, L., et al.: On the relation between the in-situ properties and the coronal sources of the solar wind. Astrophys. J. **846**(2), 135 (2017)
25. Zhao, L., et al.: An anomalous composition in slow solar wind as a signature of magnetic reconnection in its source region. ApJS **228**, 1 (2017)
26. Zirker, J.B.: Coronal holes and high-speed wind streams. Rev. Geophys. Space Phys. **15**, 257 (1977)
27. Zurbuchen, T.H., Fisk, L.A., Gloeckler, G., von Steiger, R.: The solar wind composition throughout the solar cycle: a continuum of dynamic states. Geophys. Res. Lett. **29**(9) (2002). https://doi.org/10.1029/2001GL013946
28. Han, H., Li, D., Liu, W., Zhang, H., Wang, J.: High dimensional mislabeled learning. Neurocomputing **573**, 127218 (2024)