A Novel Approach for Data Analysis Based on Visualization of Phase Space Distribution Function in Plasma Turbulence Simulations*)

Tsubasa SADAKATA[1)], Shuta KITAZAWA[1)], Masanori NUNAMI[1,2)], Takahiro KATAGIRI[1)], Satoshi OHSHIMA[1)] and Toru NAGAI[1)]

[1)]*Nagoya University, Nagoya 464-8601, Japan*

[2)]*National Institute for Fusion Science, Toki 509-5292, Japan*

(Received 26 December 2021 / Accepted 18 April 2022)

Gyrokinetic simulations are important for analyzing magnetically confined plasmas. However, the data obtained from the gyrokinetic simulations are time-series of a five-dimensional phase space distribution function, making analyzing the transport phenomena extremely difficult because of its high dimensionality and large data size. We propose a novel method for analyzing such phase space distribution functions. First, the two-dimensional velocity space distribution function is mapped into the wavenumber space and visualized as an image. This enables us to easily capture the global features and the features of the individual velocity space distribution functions. Second, we apply similarity analysis based on the local features of images and cluster analysis based on distances between images and the velocity space distribution function. The proposed method enables us to automatically extract similar structures in the velocity space distribution function and quantify the duration of these structures.

Keywords: gyrokinetic simulation, visualization, similarity analysis, cluster analysis

DOI: 10.1585/pfr.17.2403079

1. Introduction

The gyrokinetic simulation is essential for analyzing the turbulent transport of the magnetically confined plasmas [1]. However, fully capturing the physics in the turbulent transport phenomena is difficult because the data obtained from the gyrokinetic simulations consist of a time-series of five-dimensional (5D) distribution functions. Additionally, for practical analysis, the amount of data to be handled is several terabytes, and an efficient data analysis method is strongly demanded. In the previous work, there has been little direct analysis of the structure of the distribution function on a 5D phase space. For examples, the 2D velocity space has been analyzed after fixing a single 3D space point. Many studies investigated the velocity moments, such as density and fluctuation levels in a 3D space, by integrating the 2D velocity space. However, these approaches are limited in their scope of analysis and do not capture all the physics from the entire simulation data. Therefore, we propose a comprehensive visualization and analysis method that gives a complete perspective of the entire 5D data.

In this study, to efficiently analyze the gyrokinetic simulations, we first map the velocity space $(v_\parallel, v_\perp)$ distribution function into the wavenumber space (k_x, k_y), which are the wavenumbers perpendicular to the field, with fixed z, the coordinate along the field line. This visualization method can easily capture the characteristics of each velocity space distribution function and the features of a higher-dimensional wavenumber space distribution function. Previous studies showed that zonal flow is critical for reducing the turbulent transport of plasmas [2]. Therefore, if we can specify the region in the phase-space with similar structures, such as the zonal flow region ($k_y = 0$), we can identify the important coordinates and focus on the analysis range. We employ a similarity analysis method [3] used in image processing. This method can extract image features, such as contours of velocity space distribution function plots, and compute the similarity between plots. Furthermore, we apply cluster analysis as a data-driven approach to large-scale data for gyrokinetic simulations. Cluster analysis makes capturing similar structures in the wavenumber space easy. Next, we can quantify the duration of similar structures by calculating the time for the region, which is continuously assigned to the same cluster.

The rest of this article is organized as follows: Section 2 describes the gyrokinetic simulation and the 5D phase space distribution function obtained from the simulation. A novel visualization method for the 5D distribution function, mapping the velocity space distribution function into the wavenumber space, is presented in section 3, discussing the analysis methods, similarity analysis,

author's e-mail: sadakata@hpc.itc.nagoya-u.ac.jp

*) This article is based on the presentation at the 30th International Toki Conference on Plasma and Fusion Research (ITC30).

and cluster analysis. Lastly, section 4 concludes the paper.

2. Gyrokinetic Simulations

The gyrokinetic Vlasov flux-tube code, GKV [4] can solve the electrostatic gyrokinetic equation for the perturbed ion gyrocenter distribution function $\delta f_{ak_\perp}$

$$\left(\frac{\partial}{\partial t} + v_\parallel \mathbf{b}\cdot\nabla - \frac{\mu}{m_a}\mathbf{b}\cdot\nabla B\frac{\partial}{\partial v_\parallel} + i\omega_{aD}\right)\delta f_{ak_\perp} - \frac{c}{B}\sum_{\Delta}\mathbf{b}\cdot(\mathbf{k}'_\perp\times\mathbf{k}''_\perp)J_0\left(\frac{k'_\perp v_\perp}{\Omega_a}\right)\delta\phi_{k'_\perp}\delta f_{ak''_\perp} = \frac{e_a}{T_a}F_{aM}(-v_\parallel\mathbf{b}\cdot\nabla - i\omega_{aD} + i\omega_{a*})J_0\left(\frac{k_\perp v_\perp}{\Omega_a}\right)\delta\phi_{k_\perp} + C_a, \quad (1)$$

where $\omega_{aD} = \mathbf{k}_\perp\cdot\mathbf{v}_{aD}$ and $\omega_{a*} = \mathbf{k}_\perp\cdot\mathbf{v}_{a*}$ are the magnetic and diamagnetic drift frequencies with $\mathbf{v}_{aD} = (c/e_aB)\mathbf{b}\times(\mu\nabla B + m_a v_\parallel^2\mathbf{b}\cdot\nabla\mathbf{b})$ and $\mathbf{v}_{a*} = (cT_a/e_aB)\mathbf{b}\times[\nabla\ln n_a + (m_a v^2/2T_a - 3/2)\nabla\ln T_a]$.

In this paper, we used the data for the ion temperature gradient (ITG) turbulence simulation [5] in Large Helical Device (LHD) [6], where the case of inward shifted magnetic axis is an optimized configuration to enhance the zonal flows [7].

3. Visualization and Data Analysis for Gyrokinetic Simulations

3.1 Mapping of velocity distribution functions into the wavenumber space

Our previous paper [8] developed the visualization tool for the gyrokinetic simulation data using the head-mounted display. This tool enables us to visualize the data of the 5D distribution functions from the gyrokinetic simulations in the 3D real space with 2D velocity space distribution functions. However, since the gyrokinetic simulations are based on drift wave turbulence, discussing them in the Fourier space perpendicular to the field is valuable. We can discuss the turbulence physics compared with real space in the wavenumber space, for example, the wavenumber spectra of the turbulence. Consequently, to discuss them directly, we visualized them as they are in the wavenumber space in this study. Here we visualize the 5D phase space distribution functions as image data. To capture the features in phase space, which is in the higher dimensions, we map the 2D velocity distribution plots at each simulation time into the 2D wavenumber space (k_x, k_y) with the remaining z-coordinates fixed. Figure 1 shows an example of mapping the velocity distribution function into the wavenumber space. The mapping results are continuously visualized for each simulation time, which enables us to see the global structure change on the wavenumber space. In the previous work, it was known that the fluctuation levels of the distribution function in the case of the zonal flow enhances are dominant at $k_y = 0$. However, the 5D structure of the distribution function in the region has not been clarified. In this study, we visualized these structures directly and confirmed for the first time that a certain structure of the distribution function remains for a long simulation time.

3.2 Similarity analysis

To evaluate the similarity of the distribution functions and to quantify the sustainability time of the structures in the mapped distribution functions, a method to compute the data similarity is needed. Therefore, we employ a method based on local image features, which has success-

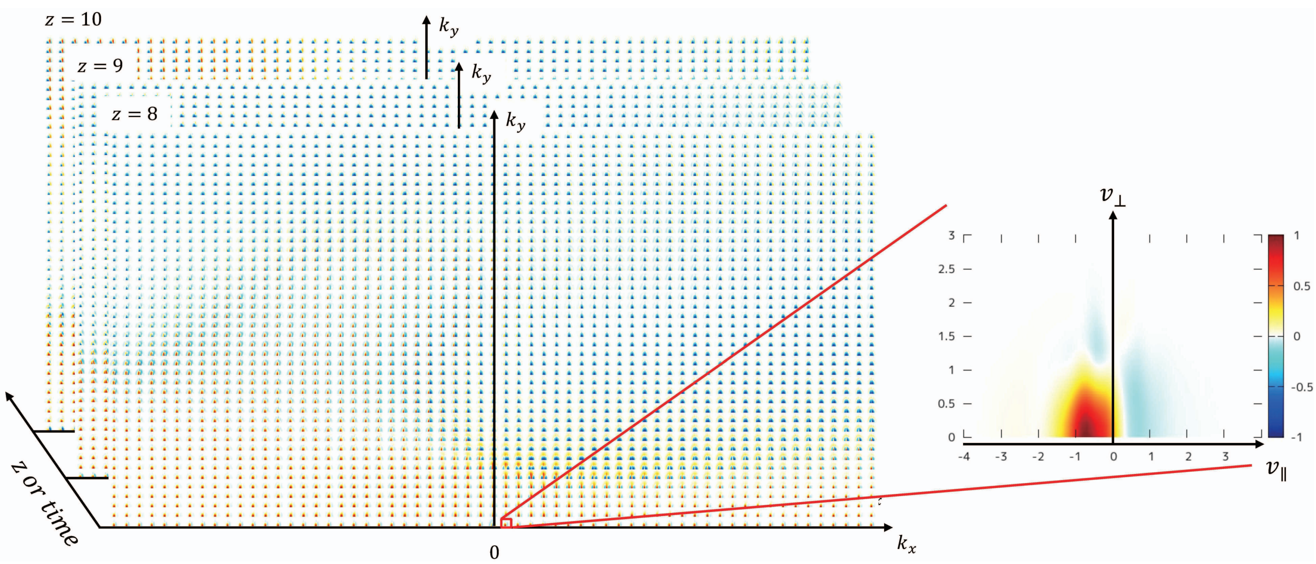
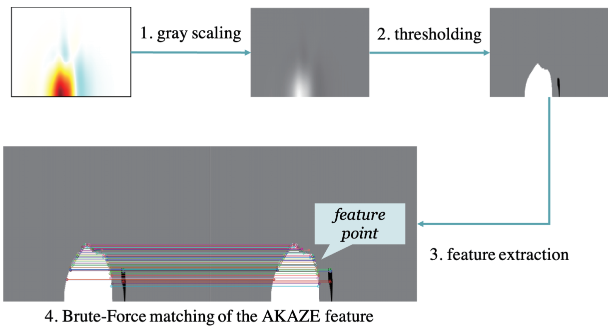
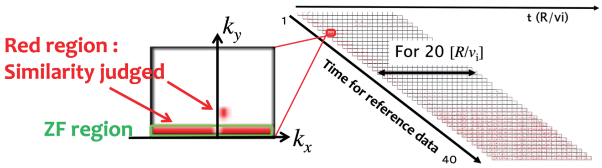


Fig. 1 An example of mapping the velocity space distribution function into the wavenumber space. Here, $v_\parallel$ and $v_\perp$ are normalized by the ion thermal velocity v_{ti}.



(a) The flow of similarity analysis



(b) Example of comparison for each time step

Fig. 2 (a) is a process flow from feature extraction to matching using OpenCV [10], a library for computer vision. In (b), the similarity is analyzed for each simulation time; red dots indicate the similarities.

fully determined image similarity. In this method, first, we extract the features from each velocity space distribution plot. Then, we compare the features to be evaluated in a brute force comparison. Here, we use the AKAZE [9] local features to detect and match key points on two images, as shown in Fig. 2 (a), where the AKAZE is a fast multiscale feature detection and description that exploits the benefits of nonlinear scale spaces. By computing the similarity of the velocity distribution function at the same wavenumber space coordinates, we can identify how long the same structure is sustained in the simulation. As shown in Fig. 2 (b), this similarity analysis indicates that many long-lasting structures are found, especially in the zonal flow region.

3.3 Cluster analysis

We employ the cluster analysis to discuss more precise characteristics of the structures in the wavenumber space. Cluster analysis groups a set of objects together in a cluster. One of the famous cluster analysis algorithms is the k-means algorithm [11] which is efficient and simple to implement. It has been successfully used in various fields such as computer vision, marketing, and bioinformatics [12]. The k-means algorithm finds cluster centers that minimize the sum of squared distances from each data object being clustered to its closest cluster center. The objective is to partition N data objects into K clusters ($K < N$), such that objects in the same cluster are as similar as possible and as dissimilar as possible from those in different clusters.

Algorithm 1 k-means algorithm

Input: number of clusters k, N data objects, *threshold*
Output: k centroids

1: Randomly select k objects as the initial cluster centroids, denoted as μ_k
2: Calculate J in 2
3: Assign each object x_n to the closest cluster
4: Calculate a new centroid μ_k for each cluster
5: Recalculate J in 2
6: Repeat step3–5 until the moving distance of $J <$ *threshold*.

Fig. 3 k-means algorithm.

The k-means algorithm is summarized in Fig. 3. Let $x_n (1 \leq n \leq N)$ be a data object and $\mu_k (1 \leq k \leq K)$ be the mean of the data objects belonging to cluster $c_k (1 \leq k \leq K)$. Here, k is just the cluster index number, not a wavenumber. The sum of squared error between μ_k and the data objects in cluster x_k is defined as

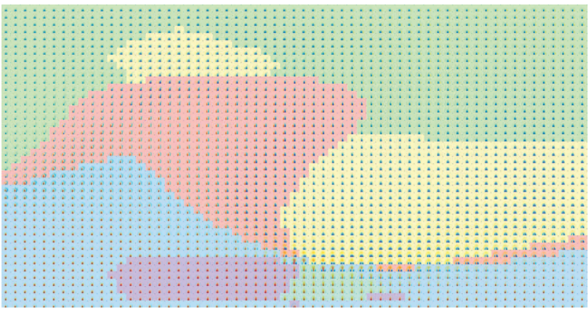
$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} \|x_n - \mu_k\|^2. \tag{2}$$

Cluster assignment and centroids recalculation is repeated until the centroid shift is less than a predetermined threshold.

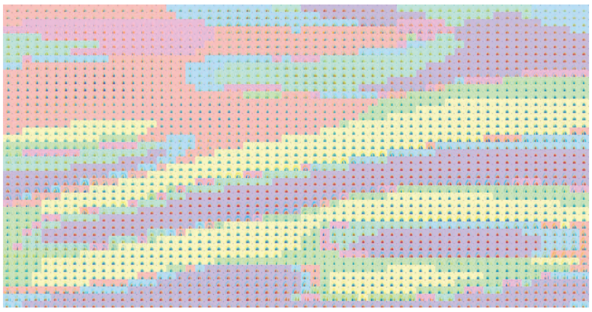
As shown in Fig. 4, we plot the time-series of the velocity space distribution plots belonging to the same cluster by highlighting them with the same color to capture the arrangement of similar structures in the wavenumber space. This cluster analysis confirms that the similarity of the 2D velocity space distribution function demonstrates a certain structure in the wavenumber space. Furthermore, to quantify the duration of the structure, we count the time assigned to the same cluster for each velocity space distribution function in the wavenumber space coordinate. Figure 5 illustrates the result of visualizing the duration. The figure shows that some structures are sustained for a long time at the coordinates near the origin in the wavenumber space. In the figure, we also indicate the numbers in the upper right area of each velocity space distribution plot, which represents the counts corresponding to the duration time. The number near the origin in the wavenumber space is larger than the other coordinates in the plots. Using the counts' information, we can identify the region where the robust structures of the distribution functions occur in the simulations.

4. Conclusion

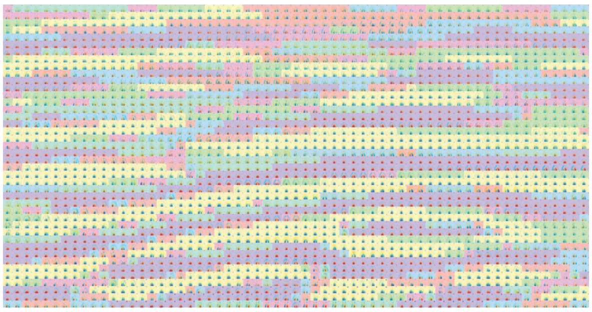
In this study, we proposed a novel scheme to evaluate and analyze the 5D phase space distribution function of time-series obtained from gyrokinetic simulations. To capture the global characteristics in the wavenumber space, we mapped the 2D velocity space distribution function into the



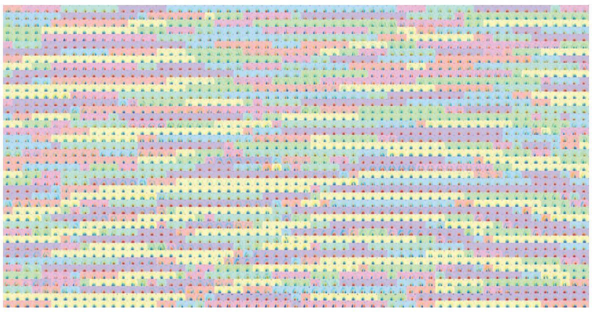
(a) simulation time = 1.0



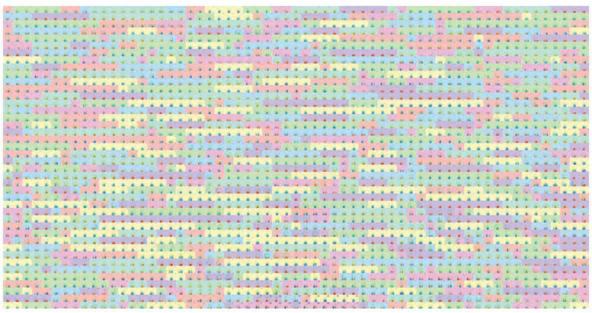
(b) simulation time = 5.0



(c) simulation time = 10.0



(d) simulation time = 15.0



(e) simulation time = 20.0

Fig. 4 Visualization of the results of the cluster analysis of a velocity space distribution function (k = 7). The simulation time is in the unit of R_0/v_{ti}, where R_0 is the major radius and v_{ti} is the ion thermal velocity.

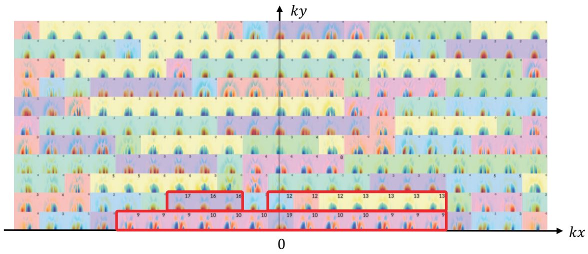


Fig. 5 Duration of the structure in each wavenumber space coordinate when simulation time = 20.0. The area enclosed by the red box sustains the same structure longer than the other coordinates.

2D wavenumber space with the z-coordinates fixed at each simulation time. Second, for detailed similarity analysis of the two dimensional velocity space distribution functions, we applied similarity analysis by image to local feature extraction and comparison. This method facilitates identifying the structure that persists in each velocity space distribution function. Finally, we applied the clustering method using the k-means algorithm on the entire data output from the gyrokinetic simulations to efficiently identify similar structures among the massive number of velocity space distribution functions and compute their durations. Similar structures of the 2D velocity distribution functions in the wavenumber space can easily be identified through cluster analysis. Furthermore, we revealed that the similarity of shapes of the 2D velocity space distribution function retains a certain structure in the wavenumber space. Additionally, we can quantify the duration of the remaining structure by counting the time assigned to the same cluster.

The method proposed in this study, which treats 5D phase space distribution functions as image data for analysis, can be widely adapted. Hence, the proposed method is general-purpose and can be applied to various simulation results. Also, it could facilitate data-driven predictions instead of numerical simulation, using machine learning methods, which are currently researched and developed in various fields, such as computer vision. The application to the gyrokinetic simulation studies and correspondence with fluid dynamics through a comparison of the velocity space distribution function and the velocity moment variables in real space will be discussed elsewhere in the future.

Acknowledgment

This work is supported in part by NIFS Grants NIFS20KNST161, NIFS21KNTS071, and JSPS Grant 16H04620, 20K03907, 17K14899, 19H05662, and by MEXT as “Program for Promoting Researches on the Supercomputer Fugaku” (Exploration of burning plasma confinement physics, JPMXP1020200103). The simulations were performed on “Plasma Simulator” (NEC SX-Aurora

TSUBASA) of NIFS with the support, and JFRS-1 at IFERC-CSC with the supports.

[1] T. Görler *et al.*, Phys. Plasmas **21**, 122307 (2014).
[2] T.-H. Watanabe, H. Sugama and S. Ferrando-Margalet, Phys. Rev. Lett. **100**, 195002 (2008).
[3] R. Szeliski, *Computer Vision: Algorithms and Applications* (2010).
[4] T.-H. Watanabe and H. Sugama, Nucl. Fusion **46**, 24 (2006).
[5] M. Nunami, T.-H. Watanebe, H. Sugama and K. Tanaka, Phys. Plasmas **19**, 042504 (2012).
[6] Y. Takeiri *et al.*, Nucl. Fusion **57**, 102023 (2017).
[7] M. Nunami, T.-H. Watanebe and H. Sugama, Phys. Plasmas **20**, 092307 (2013).
[8] S. Kitazawa, M. Nunami *et al.*, The 81st national convention of IPSJ, 2M-05, Fukuoka, Japan, 2019.
[9] P.F. Alcantarilla, J. Nuevo and A. Bartoli, Fast explicit diffusion for accelerated features in nonlinear scale spaces, British Machine Vision Conference (BMVC), 09 2013.
[10] Opencv official web site, https://opencv.org/
[11] J. MacQueen, *Some methods for classification and analysis of multivariate observations* (Berkeley: University of California Press, 1967) p.281-297.
[12] A. Jain, Data clustering: 50 years beyond k-means, Pattern Recognition Letters **31**, no. 8, 651 (June 2010).