The Analysis of the Solar Vortex on the Solar Atmosphere

YUYANG YUAN

SUPERVISIORS:

Dr. GARY VERTH & Prof. VIKTOR FEDUN

University of Sheffield

School of Mathematics and Statistics

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

January 2024

I would like to dedicate this thesis to my parent, grandma, grandmother, grandpa, X.Y. Zhou, and other family members, my supervisors at the University of Sheffield, and my teacher at Shenzhen University.

Acknowledgements

First, I greatly thank my two supervisors, Dr. Gary Verth and Prof. Viktor Fedun, my parents, Dr. Suzana S. A. Silva. Without their help, I could not study at this wonderful university, the University of Sheffield, and had an unforgettable experience there. It is my great honor to be a student here, and thank you for the precious opportunity offered by my supervisors. Dr. Gary Verth and Prof. Viktor Fedun always exhibit professionalism. They are wise supervisors, and their creative guidance always inspires me. I very much appreciate all the support they have offered. Moreover, they always encouraged me and inspired me. Again, I want to express that it is my great honor to study from them.

I want to express great thanks to Dr. Suzana S. A. Silva. She is professional and creative, especially in the research field of solar vortex analysis. Besides that, thanks for the encouragement from Dr. Suzana S. A. Silva.

I want to express great thanks to my parents, my grandmother, grandma and other family members. Thanks for their support of my study here, and stand always with me. I miss them so much.

I want to thank the whole Plasma Dynamics Group, the School of Mathematics and Statistics, and the University of Sheffield. Thanks for the funding offered by the university.

I want to thank the Prof. J.M. Zhong. Prof. L.C. Fang. Prof. H. Chen. They were my teacher when I studied at Shenzhen University.

I want to thank the United Kingdom. The UK is a wonderful country! Here have rule and law.

Once again, thank you, Dr. Gary Verth, Prof. Viktor Fedun, and Dr. Suzana S. A. Silva.

Declaration of Authorship

I hereby declare that, except where clear reference is made to the work of others, the contents of this dissertation are original and have not, in whole or in part, been submitted to this or any other university for consideration for any other degree or qualification. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

YUYANG YUAN

January 2024

Abstract

Ubiquitous vortical structures are considered to act as a natural source of various solar plasma phenomena, e.g., a wide range of magnetohydrodynamic (MHD) waves and jet excitation. This work aims to develop an advanced vortex detection algorithm based on the Γ method and using a separable convolution kernel technique. This method is applied to detect and analyze the photospheric vortices in 3D realistic magnetoconvection numerical and observational data. We present the advanced Γ Method (AGM), and our results indicate that the AGM performs with better accuracy in comparison with the original Γ method. The AGM allows us to identify small and large-scale vortices with no vortex interposition without requiring changing the threshold. Thereby, the nondetection issue is mostly prevented. It was found that the Γ method failed to identify the large and longer-lived vortices, which were detected by the AGM. The size of the detected vortical structures tends to vary over time, with most vortices shrinking towards their end. The vorticity at the center is also not constant, presenting a sharp decay as the vortex ceases to exist. Due to its capability of identifying vortices with minimum nondetection, the vortices' properties–such as lifetime, geometry, and dynamics–are better captured by the AGM than the Γ method. In this era of new high-resolution observation, the AGM can be used as a precise technique for identifying and performing statistical analysis of solar atmospheric vortices.

Next, this Thesis introduces a novel vortex analysis method, that is, the discrete Fréchet distance vortex visualization method (DFDVVM). The DFDVVM is developed to analyze the time-dependent behavior of vortices while quantifying each detected vortex's vorticity evolution in a high-dimensional space (the DFD space). We also developed a vortex clustering algorithm for vortices and clustered them into different clusters. As a result, each cluster's average vorticity pattern is calculated, and each cluster's statistical analysis is presented.

List of Publications

The thesis leads to the following publications:

- Advanced Γ method for small-scale vortex detection in the solar atmosphere, Yuyang Yuan, Suzana de Souza e Almeida Silva, Viktor Fedun, Irina N. Kitiashvili, Gary Verth (published in the Astrophysical Journal and the Astrophysical Journal Supplement Series, DOI: 10.3847/1538-4365/acc835).
- Vortex categorizing for small-scale vortex in the solar atmosphere, Yuyang Yuan, Viktor Fedun, Gary Verth, Suzana de Souza e Almeida Silva (In preparation).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Sun

In the universe, the Sun is a common star with an absolute stellar magnitude of 4.8; its spectral type is G2 V. The Sun is unique to our solar system, and 99.86 % of the solar system's mass comes from the Sun. The Sun plays a critical role in the Earth. Thanks to the Sun and the heat and energy this star provides, the Earth has become an energetic planet with different inhabiting lifeforms. Therefore, investigating the Sun is crucial to revealing the solar system's origin and the foundation for humans to explore outer space in the future.

The Sun was born around 4.6 billion years ago. Initially, a cloud of interstellar material performs rotation motion and keeps contracting, accompanied by collapse. Then, the gravity and internal pressure of the protostar gradually reach balance, and the contraction slows down. Heat and luminosity are emitted during these processes. The internal structure and core of the star become settled down. The high internal temperature then facilitates hydrogen fusion to helium, providing the Sun's heat, energy, and luminosity. After the forming stage, the Sun entered its main stable stage until now. It is estimated that the Sun will continue this stage for another 5 billion years until it runs out of its internal hydrogen. After that, the Sun will expand and become a red giant (the radius would expand about 200 times the current radius of the Sun), which would swallow other planets of the solar system; it would collapse and end as a white dwarf with a size similar to the Earth. Table 1.1 lists some general physical properties of the Sun.

The Sun consists of two primary parts: the interior and the atmosphere. Figure 1.1 is a simplified diagram of the Sun's structure. In the following

Property	Value
Age	4.6 billion years
Mass ($M_\odot$)	$1.99 \times 10^{30}\ kg$
Radius ($R_\odot$)	695.5 Mm
Mean density	$1.4 \times 10^3\ kg/m^3$
Mean distance from Earth	$1\ AU = 1.496 \times 10^{11}\ m = 215\, R_\odot$
Surface gravity ($g_\odot$)	274 m/s^2
Escape velocity at surface	618 km/s
Radiation (luminosity $L_\odot$)	$3.86 \times 10^{26}\ W (= 3.86 \times 10^{33}\ erg/s)$
Equatorial (synodic) rotation period	26.24 $days$
Mass loss rate	$10^9\ kg/s$
Angular momentum	$1.7 \times 10^{41}\ kg \cdot m^2/s$
Effective temperature	5785 K
1 arcsec ($\equiv 1''$)	$\approx 726\ km$

Table 1.1: Physicial properties of the Sun. Courtesy of Priest (2014).

Sections, we will review the structure of the Sun.

1.1.1 The Solar Interior

The Sun's interior, being unobservable directly, poses a challenge for studying its properties. However, helioseismology offers a valuable method to infer internal characteristics. The solar interior is generally divided into three regions: the core, the radiative zone, and the convection zone, each characterized by distinct physical processes and properties (see Figure 1.1). This categorisation allows for a comprehensive understanding of the Sun's complex internal dynamics and structure.

The core, situated at the Sun's central region, is characterized by extreme conditions, featuring a temperature of approximately 16 million degrees Kelvin and a density reaching 150 g/cm^3. Conventionally, the region from the Sun's center to 0.25 $R_\odot$ is defined as the core (see Figure 1.1). This core region contains about 34% of the Sun's total mass, resulting in a tremendous inner pressure of around 265 billion bars. Combining this immense pressure with high temperature facilitates the ongoing fusion of hydrogen into helium—a remarkable process expected to persist for another 4 billion years.

The next region outside the core is the radiative zone, extending from $0.25R_\odot$ to $0.7R_\odot$ of the Sun (see Figure 1.1). In contrast to the core, the density and temperature in the radiative zone decrease significantly as the

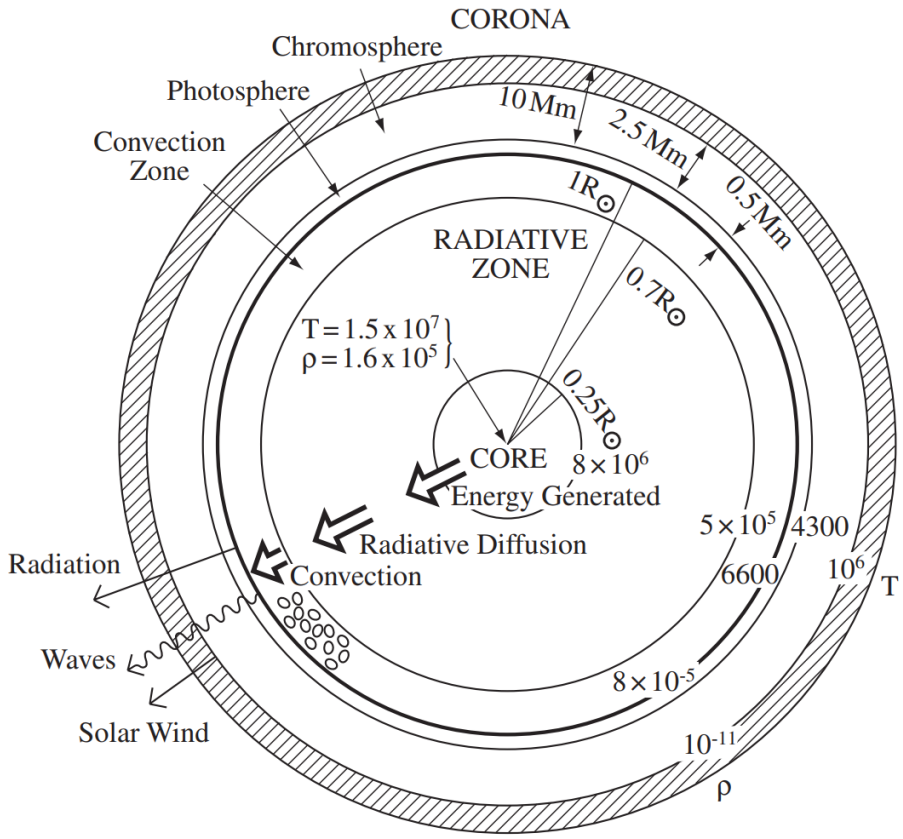


Figure 1.1: *The diagram of the structure of the Sun. Courtesy of Priest (2014).*

solar radius increases. Typically, the density of the radiative zone ranges between 20 g/cm^3 and 0.2 g/cm^3, with a temperature difference that could reach 5 million K (see Figure 1.2). Photons generated in the core traverse this region through radiation and conduction. Countless collisions occur between photons and particles in the zone, leading to energy loss and changes in the direction of photon travel after each collision. It is estimated that photons may require up to a million years to escape from this zone.

The convection zone is above the top of the radiative zone, where heat and light are transferred through the convective process. This zone's depth is nearly 200,000 km (see Figure 1.1). In this region, energy is transported as hot plasma moves upward from below toward the Sun's surface, releasing heat in the process. After cooling, the cooler plasma returns to the bottom of the zone to repeat the cycle.

In addition to the three classic zones, the latest model of the Sun introduces a thin layer known as the tachocline. The tachocline is characterized by a strong radial differential rotation of the plasma and serves to distinguish between the radiative and convection zones. This layer is believed to play a significant role in generating the solar magnetic field. The investigation of the tachocline contributing to our understanding of the mechanisms behind the creation of the solar magnetic field through processes such as dynamo mechanisms.

1.1.2 The Solar Atmosphere

Compared to the internal part of the Sun, especially in the radiative zone where photons encounter countless collisions and might need 1 million years to escape, photons can escape into outer space directly once they reach the solar atmosphere. Therefore, the solar atmosphere is defined as the solar region where photons can easily escape into space. Conventionally, the solar atmosphere consists of four parts, i.e., the Photosphere, Chromosphere, transition region, and Corona. Each part's physical properties diverge from the other. Such as each region's plasma pressure and magnetic field are different. In plasma physics, plasma-β is defined as the ratio between plasma pressure and magnetic pressure, and it is usually used to determine which pressure effect is dominant. The mathematical expression of the plasma-β is below:

$$\beta = \frac{nk_BT}{B^2/\mu_0}, \tag{1.1}$$

where n is the total number of density; $k_B \approx 1.38 \times 10^{-23} JK^{-1}$ denotes the Boltzmann constant; T denotes the temperature; B represents the density of

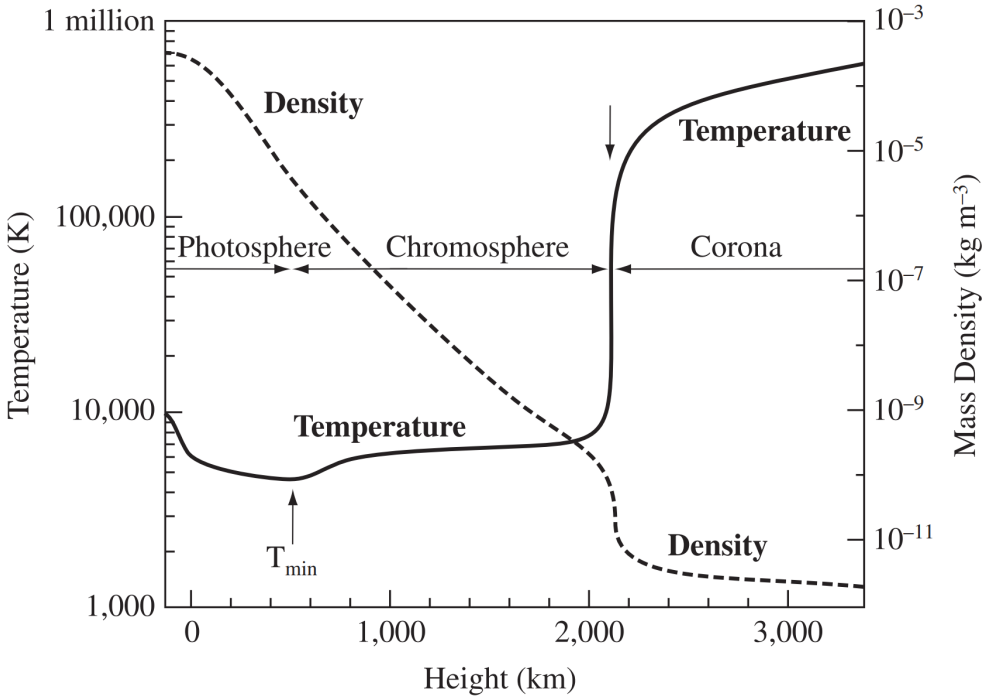


Figure 1.2: *A diagram of the mean variation of temperature and density as a function of height in the solar atmosphere. Courtesy of Avrett and Loeser (2008).*

the magnetic flux; $\mu_0 = 4\pi \times 10^7 N/A^2$ denotes the magnetic permeability of free space.

Figure 1.3 depicted the variation of plasma-β in a different region of the solar atmosphere. In the Photosphere and the lower Chromosphere, the corresponding plasma-β is greater than 1. In contrast, the corresponding plasma-β in the upper Chromosphere and the Corona is smaller than 1.

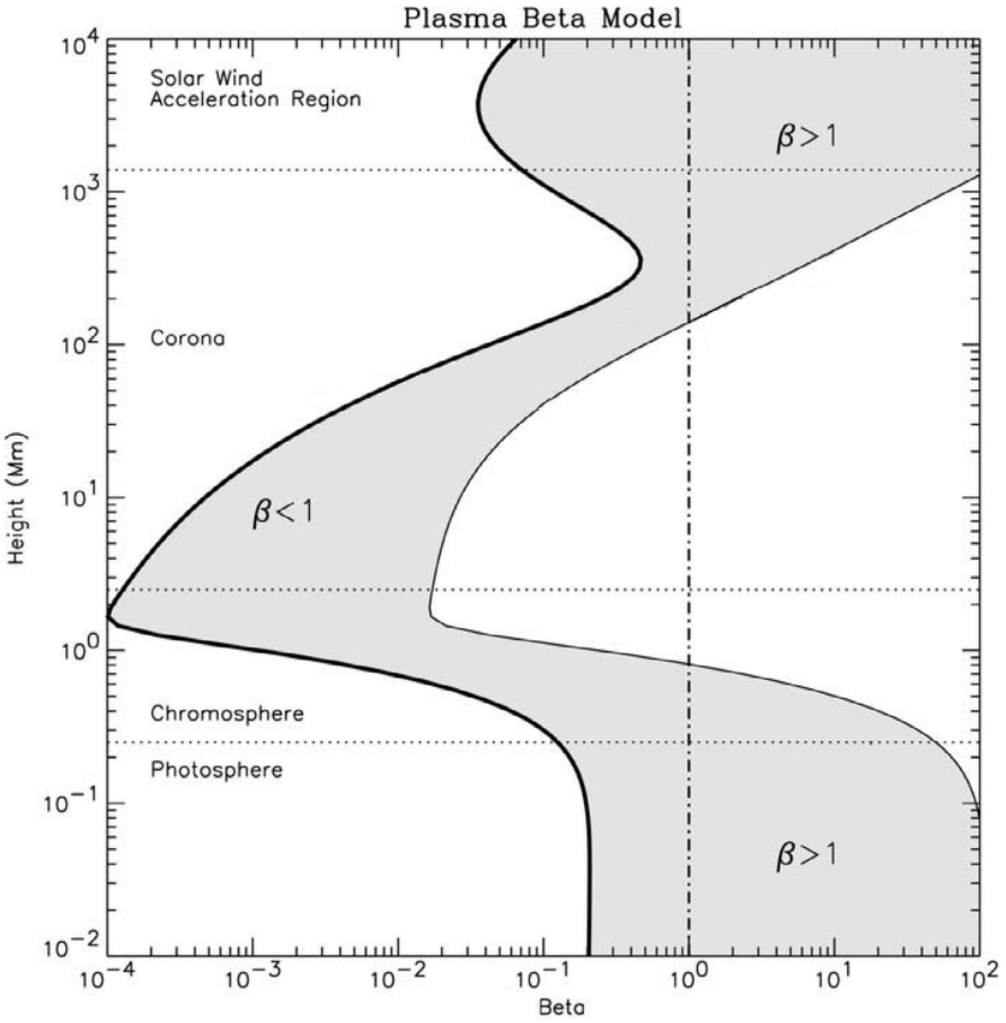


Figure 1.3: *Plasma-β model depending on the height in the solar atmosphere. Courtesy of Gary et al. (2001).*

The Photosphere is the lowest part of the solar atmosphere and is a thin visible layer (around several hundred kilometers thick) of plasma. The meaning of its name came from 'light' in Greek. Most solar radiation in solar atmospheres

comes from the Photosphere, which is relatively dense and opaque compared to other upper parts of solar atmospheres. The Photosphere's temperature is decreasing along with the solar radius (see Figure 1.2). At the bottom, the temperature is about 6400 K, and 4400 K at the top. The state of the Photosphere is not stationary. Various convection motions of plasma of different spatial and temporal scales are found in the Photosphere. The granules are found everywhere in the Photosphere, and they appear when the internal hot plasma bubbles up. Their central area contains hot rising plasma and its boundary region is relatively cool, and a part of that cool plasma follows the granule downflow near the edge and returns to the bottom part of the Photosphere. The granules exhibit an irregular appearance with a width range from 0.3 to 2 Mm, and their life varies from 5 to 10 minutes. photospheric vortices are also found in the Photosphere. They are rotating plasma structures characterized by strong vorticity. They have been discovered in different scales and periods and are considered a building block of other solar phenomena (see Section 1.2 for more details).

The Chromosphere is the second stratum within the solar atmosphere directly above the Photosphere. The thickness of it surpasses that of the Photosphere, averaging approximately 2000 km (refer to Figure 1.1). It conveys the heat from the Sun's interior to the transition region and Corona. In the Chromosphere, the atmospheric density decreases, and temperature increases with altitude. This layer exhibits a non-uniform temperature profile, typically falling within the temperature range of 4500 K to 20000 K, with a conceivable minimum temperature as low as 3800 K. Unlike the Photosphere, magnetic pressure takes precedence over plasma pressure (i.e., plasma-$\beta < 1$), resulting in the dominance of magnetic fields in driving plasma motion. Notably, chromospheric vortices are influenced by twisted magnetic fields, with their footpoints manipulated by the photospheric vortices below (see Section 1.2 for more detail). Solar spicules are plasma jet features widely observed in the Chromosphere when analysing chromospheric spectral lines (such as H-α); they span from the Photosphere to the Chromosphere (Bray et al., 1974; Zirin, 1998; Tavabi et al., 2012; Porfir'eva and Yakunina, 2016). It is estimated that at least 100,000 spicules coexist in the Chromosphere at any time. There are two types of spicule, i.e., type I and type II. The type I spicules have a lifetime ranging from 3 to 10 minutes, with a diameter varying from 120 km to 700 km, and their jet velocity is around 10 to 50 km/s. Type I spicules might reach 3 to 4 Mm height in active regions and even 5 Mm in the quiet Sun before fading away (Beckers, 1972; de Pontieu et al., 2007). Compared to type I, type II spicules exhibit more behaviors (the jet velocity is about 30 to 150 km/s), and they might reach 10 Mm height and fade much quicker (lifetime of

about 10 to 180 s). The chromospheric plasma is partially ionized, with weak ionization at the bottom and strong ionization at the top of this layer.

The transition region is another thin layer (about 100 km) next to the Chromosphere but below the Corona. The temperature changes dramatically within the transition region, and a typical transition region model shows its temperature can increase from 3×10^4 K to 3×10^5 K within 30 km (Priest, 2014).

The Corona is the last and outermost part of the solar atmosphere, extending millions of kilometers toward outer space (e.g. Burton et al., 1971; Boehm-Vitense, 1984; Kerr, 2012; Witze, 2017). The Corona has the highest average temperature compared to other inner parts of the solar atmosphere (see Figure 1.2). In contrast, the density in the Corona is 10 million times less than the surface of the Sun. Therefore, the brightness of the Corona is much lower than that of the solar surface. During eclipses, the Corona can be observed as a faint halo. As noted before, the magnetic pressure is more significant than the plasma pressure in the Corona (see Figure 1.3), and the magnetic field significantly influences phenomena observed in the Corona. A coronal loop is a remarkable feature observed in the Corona. This giant feature is a magnetic arc originating from the Photosphere, across the Corona, and ending in the Photosphere. The scale of the coronal loop may reach thousands of kilometers, and its temperature is exceptionally high. Besides the coronal loop, other dynamic phenomena exist, such as coronal mass ejection. Most of them have a strong connection with the magnetic field. To date, revealing the Corona's mysterious high temperatures is still a challenging problem in solar physics (see e.g., Bingert et al., 2008; Rappazzo and Velli, 2010; Winebarger et al., 2012; Bourdin et al., 2014).

1.2 Solar Vortex

Solar vortices play a crucial role in the dynamic processes of the solar atmosphere. They potentially contribute to the excitation of magnetohydrodynamic (MHD) waves (Schüessler, 1984; Attie et al., 2009; Fedun et al., 2011; Kitiashvili et al., 2011; Yadav et al., 2022) and the generation of plasma jets (Kitiashvili et al., 2013; Iijima and Yokoyama, 2017; Snow et al., 2018; Skirvin et al., 2022) . Additionally, these vortices may serve as conduits for the transfer of energy to the solar corona (Wedemeyer-Böhm et al., 2012; Shelyag et al., 2012; Yadav et al., 2021). Therefore, a comprehensive understanding of solar vortices is crucial for gaining insight into the dynamics of the solar atmosphere. This section first reviews solar vortices observed in the Photosphere (Section

1.2.1) and the Chromosphere (Section 1.2.2). Subsequently, the formation mechanism of solar vortices is discussed in Section 1.2.3. Finally, the role of solar vortices in the broader context of the solar atmosphere is presented in Section 1.2.4.

1.2.1 Vortices Observed in the Photosphere

At the Photosphere, the swirling motion of the turbulent flow has been widely reported at different locations and observed on various scales (e.g., Brandt et al., 1988; Wang et al., 1995; Bonet et al., 2008; Attie et al., 2009; Bonet et al., 2010; Vargas Domínguez et al., 2011; Wedemeyer-Böhm et al., 2012).

In intergranular lanes, vortex flow was identified as rotation motions of magnetic bright points (BPs; Riethmüller et al., 2014). They are also known as magnetic field concentrations. By analysing G-band images obtained from the Swedish Solar Telescope (SST; Scharmer et al., 2003a), Bonet et al. (2008) reported numerous small-scale (0.5 to 2 Mm), short lifetime (5 to 15 min) vortices when tracing the motions of magnetic BPs. These vortices were found concentrated on the downdrafts of intergranular lanes. Therefore, they related the granulation downdraft to the formation of these swirls. With the help of the SUNRISE (Solanki et al., 2010; Barthol et al., 2011; Gandorfer et al., 2011; Berkefeld et al., 2011) that equipped with Imaging Magnetograph eXperiment (IMaX, Martínez Pillet et al., 2011), Steiner et al. (2010) have identified vortices in the intergranular lanes with much smaller scale (the mean radius is around 150 km). According to their statement, detected vortices might have a smaller scale while there is a limitation in the resolution of the equipment. Steiner et al. (2010) found various moving horizontal lanes characterized by bright and dark edges. After comparing the observational data with the solar surface convection simulation, Steiner et al. (2010) concluded that these structures were horizontal vortex tubes. Via visually inspecting the magnetograms of the observational data, vortices with a mean lifetime of about 8 min were detected by Bonet et al. (2010). They concluded that the angular momentum transported to the intergranular lanes drives the vortex. Balmaceda et al. (2010) identified solar vortices with a radius of around 0.25 to 1 Mm in the Photosphere. Vargas Domínguez et al. (2011) also identified a similar scale of vortex flow in the quiet-Sun region when analysing the G-band image from SST.

Swirling motions are also found in granular and supergranular junctions with much larger scales. Attie et al. (2009) have detected vortical motions in supergranular junctions of the quiet-Sun region when analysing G-band images obtained from Solar Optical Telescope/Filtergraph (Ichimoto et al.,

2008) on the Hinode (Kosugi et al., 2007). These vortical structures were reported on a much larger spatial (15 to 20 Mm) and temporal scale (1 to 2 h) than those vortices near the intergranular lane. Requerey et al. (2018) have detected long-lasting (24 h) supergranular vortices by analysing data acquired by the Narrowband Filter Imager/Hinode. These vortices are interpretations of converging flows near the supergranular junctions.

1.2.2 Vortices Observed in the Chromosphere

In addition to vortex flow detected in the Photosphere, their imprint is also found in the Chromosphere. Analysing chromospheric spectral lines obtained from the CRISP Imaging Spectro-Polarimeter (Scharmer et al., 2008a) at the SST, Wedemeyer-Böhm and Rouppe van der Voort (2009) detected apparent swirling motion near the center of the Ca II 854.2 nm. They name them chromospheric swirls. These swirls had an average size of about 1.5 Mm, and a speed of up-flow of around 2 to 7 km/s. Moreover, they might relate to the photospheric magnetic BPs found in intergranular lanes. Shetye et al. (2019) found that these chromospheric swirls exhibit various appearances on different scales, and these structures seem to have a tight connection with magnetic concentrations at downdraft centers. Chromospheric swirls were also identified by Wedemeyer-Böhm et al. (2012) when analysing Ca II 854.2 nm spectral processed from CRISP. They postulated that the detected swirls and the photospheric vortices below are the different parts of a vortex tube across several solar atmospheres. And the vortex tube might further extend to the Corona, and its cross-sectional area increases with height.

1.2.3 The Formation of Vortex Flow

Formation mechanism of intergranular vortex flow and atmospheric vortex flow

Based on a solar convection simulation, Nordlund (1985) associated the 'bath-tub effect' as a possible formation mechanism of photospheric vortex flow. The effect refers to the granulation downdraft motion when plasma diverges from the neighboring granular center and converges on an intergranular lane. During the process, a considerable amount of angular momentum is transported via the motion of plasma, and fluid elements might perform rotation at various scales when approaching the downdraft center. Therefore, photospheric vortex flow might be initiated. This flow type is also known as intergranular vortex flow (IVF), which extends from the Photosphere to the upper part of the

convection zone and was introduced by Wedemeyer and Steiner (2014). However, downdraft centers do not always guarantee the vortex dynamic. Vortical motion is possible only when there is sufficient vorticity in the sink region.

Besides the IVF, Wedemeyer and Steiner (2014) also suggested another type of vortex flow system, that is, atmospheric vortex flow (AVF), when analysing the connection between the observed chromospheric swirls and photospheric BPs below. They proposed that the AVF and the IVF spatial correlate with each other with the help of a coincided rotating magnetic field. Figure 1.4 demonstrates the whole picture of these two types of flow. In the convection zone where the plasma pressure dominates the magnetic pressure, i.e., plasma-$\beta > 1$, the gas undergoes substantial ionization, resulting in the magnetic field essentially entrapped and transported along with the convective motion of plasma flow. Thus, the plasma flow inside an IVF would drag the surrounding magnetic field and force them to co-rotate, and, as a result, a twisted magnetic field structure that coincides with the IVF is established (Wedemeyer and Steiner, 2014). If the twisted magnetic field structure persist with the vortical motion of the corresponding IVF long enough, the upper part of the magnetic field structure might induce an AVF (see Wedemeyer and Steiner (2014) for further details).

Vortex tube formation in the near-surface layers

Besides the suggested formation mechanism for IVF and AVF mentioned above, Kitiashvili et al. (2012b) proposed two other basic formation mechanisms for vortex flow while analysing vortex obtained from the SolarBox–a 3D radiative MHD code developed by Jacoutot et al. (2008). These two mechanisms are (1) the instability of the convective granule and (2) the Kelvin Helmholtz instability within intergranular lanes.

The instability of convective granule: Analysing the numerical simulation described above, small-scale upflowing plumes inside the granules might trigger small-scale vortex flow. Firstly, vortex sheets may be produced by these upflowing plumes. Then, neighboring turbulence flow would overturn these vortex sheets, and vertically oriented vortex flows may form during the process.

The Kelvin Helmholtz instability within intergranular lanes: The interaction between strong downflows and horizontal sheering flow along granules' peripherals is complicated in intergranular lanes. Therefore, the Kelvin-Helmholtz instability of shearing flow might induce vortex flow. In general, the region's Richardson number distribution can be used to characterize its Kelvin–Helmholtz instability and the small distribution indicates the region

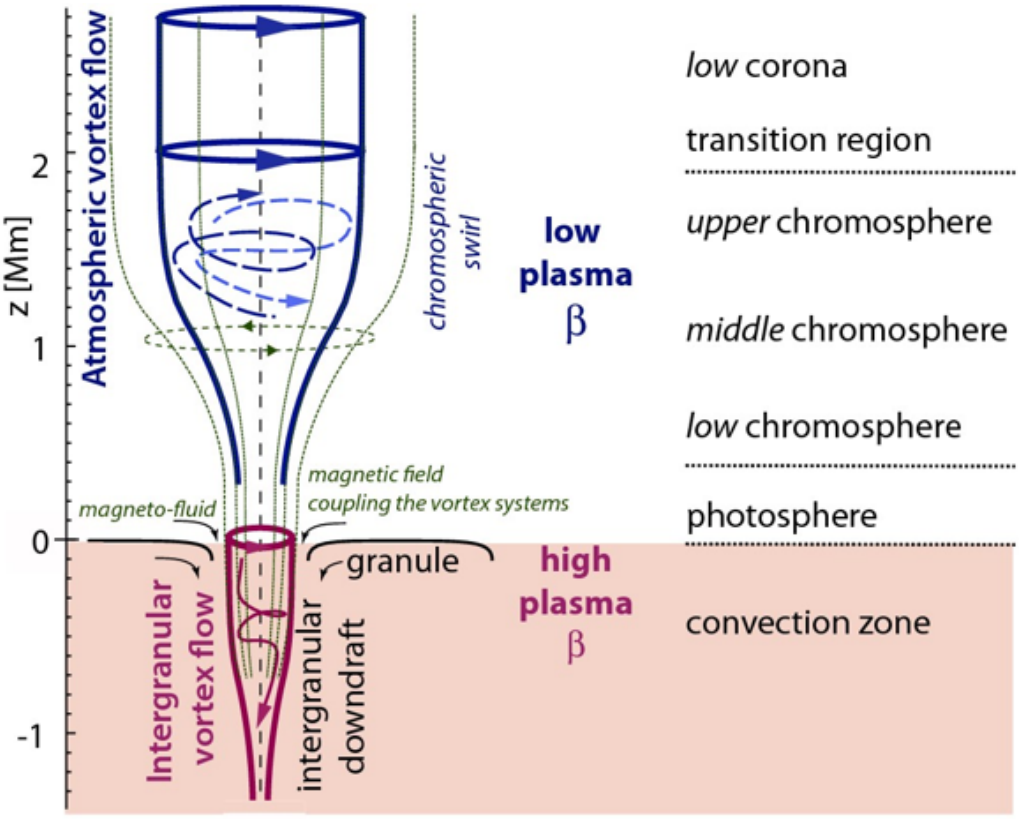


Figure 1.4: *A schematic representation of the double-nature of vortex flows. An intergranular vortex flow (IVF) usually extends from the Photosphere to the above part of the convection zone. It might initiate a corresponding atmospheric vortex flow (AVF), also introduced as a 'magnetic tornado' by Wedemeyer-Böhm et al. (2012). An AVF can exist on top of an IVF. The imaginary borders of both vortex flows are denoted with thick solid lines. Plasma within AVF can swirl up and down (thin lines with arrows), while plasma within IVF can only swirl down to the downdraft center. The environment to establish an AVF requires low plasma-β conditions, whereas high plasma-β conditions are required for an IVF. In general, the core line of AVF and IVF might not perfectly align; that depends on where the twisted magnetic field is. Courtesy of Wedemeyer and Steiner (2014).*

would have vortex flow.

1.2.4 The Role of Solar Sortices

By analysing observational and simulation data, solar vortices have been associated with the heating mechanism of the solar atmosphere and have been inferred to play a significant role in the dynamic of the solar atmosphere. In this Section, we overview vortex analysis, whether based on observational data or realistic simulation, and the corresponding findings of the role of the solar vortices.

The intricate interaction between solar vortices and surrounding magnetic fields in the solar atmosphere is a complex and dynamic process that significantly influences the behavior of solar phenomena. Based on the numerical simulation generated by the SolarBox, Kitiashvili et al. (2012c) highlights the thermodynamic properties of magnetized vortex tubes in the quiet Sun. Solar vortices were found to play a crucial role in twisting nearby magnetic fields. The mutual influence between them is crucial; the vortical motion of plasma intensifies the twisted magnetic field's strength; subsequently, the magnetic field stabilizes the kinetic motion of the vortices and mitigates the impact of turbulent flows in the periphery. Furthermore, utilizing CO^{5}BOLD simulations (a 3D MHD simulation code), Wedemeyer-Böhm et al. (2012) and Wedemeyer and Steiner (2014) show that chromospheric swirls are excited by the twisted magnetic field and the scale of these swirls depending on the magnetic field strength. Analysing observational data obtained from Hinode, Requerey et al. (2018) found magnetic flux preferentially dragged in the vortex region and twisted by the vortical motion, supporting the idea that a magnetic flux tube gains stability when surrounded by vortex flows, influencing various magneto-hydrodynamic wave modes in the solar atmosphere. Additionally, Rappazzo et al. (2019) also found that the magnetic field coexisted with a kinematic vortex and was twisted by the vortical plasma flow based on the simulation.

Solar vortices are found to correlate with plasma jet excitation (see, e.g., Kitiashvili et al., 2013; Iijima and Yokoyama, 2017; Snow et al., 2018). Through analysing observational data acquired from the Swedish Solar Telescope (SST), De Pontieu et al. (2012) underscores the connection between solar vortices in the Photosphere and the formation of spicules. Simulation studies, such as those conducted by Kitiashvili et al. (2013) using SolarBox, reveal that small-scale jet-like ejections in the Chromosphere can be associated with spontaneous upflows within vortex tubes. Additionally, Iijima and Yokoyama (2017) concluded vortex flow as a possible formation mechanism for chromospheric jets and spicules based on the simulation employing the numerical code RAMENS^3.

They pinpoint the crucial role of the Lorentz force from twisted magnetic field lines in producing chromospheric jets. These collective findings suggest that despite originating within downflowing intergranular lanes, solar vortices play a significant role in supporting intense plasma upflows, interacting with magnetic fields, and forming chromospheric jet-like features.

Apart from interacting with magnetic fields and being responsible for exciting chromospheric jet-like features, vortex tubes, potentially extending into the Corona, serve as conduits for plasma and energy transfer within the solar atmosphere. This transportation process results in the heating of the upper solar atmosphere. Simulation studies, exemplified by Kitiashvili et al. (2012c), illustrate that vortices featuring strong upflows in their centers can propel material toward higher atmospheric layers. Additionally, a considerable amount of Poynting flux transported upward through the vortex flow contributes significantly to the heating, as indicated by Wedemeyer-Böhm et al. (2012). In addition to simulation analysis, Park et al. (2016) link chromospheric heating with solar vortices activity while investigating observational data obtained from CRISP/SST. Furthermore, solar vortices have the potential to induce various magnetohydrodynamic (MHD) waves, including torsional Alfvén, kink, or sausage modes (see Fedun et al., 2011; Shelyag et al., 2013). And a certain amount of energy transported by these MHD waves (see e.g., Mumford and Erdélyi, 2015; Mumford et al., 2015).

1.3 The Outline of Thesis

The Thesis aims to contribute practical and efficient methods for identifying and analysing solar vortices, with a primary focus on the accurate detection of solar vortices and developing novel analysis strategies. The Thesis is structured as follows:

Chapter 2 provides the theoretical foundation for establishing existing vortex identification methods and introduces fundamental knowledge in clustering analysis. The primary objective of the Thesis is to develop a practical and efficient vortex identification method. A thorough understanding of vortex definition and identification is necessary to achieve this. This Chapter begins by reviewing critical theoretical knowledge related to vortex identification. Subsequently, it explores essential and representative vortex identification techniques currently being applied. The second goal of the Thesis involves devising a creative and suitable analysis strategy to enhance our comprehension of solar vortices. Inspired by clustering analysis, new findings may be discovered by comparing solar vortices; that is, solar vortices with similar characteristics

would be naturally grouped through clustering analysis. Therefore, the last part of this Chapter is dedicated to clustering analysis. Specifically, representative and fundamental clustering techniques will be reviewed and discussed. Additionally, the last Section of this Chapter focuses on evaluating clustering performance for different clustering techniques in different scenarios.

Chapter 3 focuses on accomplishing the first goal of the Thesis, that is, developing a practical vortex identification method. This Chapter introduces the advanced Γ method (AGM), an enhancement of the Γ method (Graftieaux et al. (2001)), drawing inspiration from the fast convolution Γ_1 (Zigunov et al. (2020)). The AGM shows robustness in the identification of vortex structure through the use of various sizes of the convolution kernel. Additionally, two novel algorithms, the circular sector algorithm (CSA) and the whole-circular sector algorithm (WCSA), are introduced for extracting an accurate vortex boundary. Our results show that the AGM identified a better vortex boundary, and the extracted vortex boundary aligns well with the velocity field. In order to examine the capability of the AGM, the method is applied to identify solar vortices simulated by StellarBox, a 3D radiative MHD code. A comparative analysis is conducted with the detection results obtained from the Γ method. The AGM and the Γ method are also applied to observational data, demonstrating that the AGM outperforms the Γ method in both cases.

Chapter 4 corresponds to the second goal of the Thesis, introducing the discrete Fréchet distance vortex visualization method (DFDVVM) as a novel vortex analysis technique. This innovative method is applied to analyse the time-dependent behavior of simulated solar vortices identified by the AGM in Chapter 3, with a specific focus on vorticity evolution. Utilizing the discrete Fréchet distance metric, the DFDVVM systematically quantifies each vorticity pattern in a high-dimensional space, namely the DFD space. The method facilitates the visualization of differences between vorticity patterns and provides an intuitive means of understanding the variance among detected vortices. Furthermore, this Chapter introduces a practical vortex clustering algorithm based on the K-means algorithm. This innovative clustering method is applied to analyse each identified vortex processed by the DFDVVM. Consequently, solar vortices with similar vorticity evolution are grouped into the same cluster, allowing for a further investigation of the underlying vorticity patterns. Additionally, the introduced clustering analysis method calculates the average vorticity pattern of each cluster and presents corresponding statistical analyses. These methods are expected to significantly contribute to a comprehensive exploration of solar vortex dynamics, offering valuable insights into the intricate behavior of solar vortices.

Chapter 5 serves as the conclusion to this Thesis, providing a comprehen-

sive summary of the research and its outcomes. It offers an overview of the entire Thesis, encapsulating the key achievements from each Chapter. Furthermore, this Chapter outlines potential avenues for future research and exploration in the field.

Chapter 2

Overview of Vortex Analysis Techniques and Clutsering Analysis

Vortices typically refer to a phenomenon in which particles perform rotational motion around a shared central axis. Proper observation of this phenomenon often requires consideration of an appropriate reference frame (Lugt, 1979; Robinson, 1991). Nature provides numerous examples of vortices, including solar vortices, ocean eddies, and tornadoes. Establishing a universally applicable definition for the vortex region is still challenging due to variations in fluid properties (e.g., viscosity) across different scenarios. Additionally, there is no clear boundary between the vortical structure and its surrounding environment.

The foundational step in analysing solar vortices is defining the vortex region and delineating its boundary. This precise identification is crucial for gaining profound insights into the intricate dynamics of solar vortices. The first goal of this Thesis is to develop an efficient and practical method for vortex identification. To achieve this, the following Sections will discuss the foundational concepts underpinning vortex identification. Section 2.1 will lay the groundwork by investigating the vector field, vector calculus, and Jacobian of the vector field. Subsequently, Section 2.2 will introduce the topology of a vector field. Section 2.3 concerns the critical aspect of reference frame invariance since its strong influence on vortex observation and extraction. Section 2.4 will comprehensively review representative and popular vortex extraction techniques. The second goal of this Thesis is to develop a novel analysis method for detecting vortices, which is expected to facilitate the investigation of solar vortices. In pursuit of the second goal, Section 2.5 will discuss various

clustering techniques widely used in data analysis.

2.1 Theory of Vector Field

2.1.1 Vector Field

Steady and unsteady vector fields are distinguished based on their temporal variability. A steady vector field remains constant over time, exhibiting time independence. On the other hand, an unsteady vector field changes over time, displaying time-dependent characteristics. The mathematical representations of steady and unsteady fields in a 3D space are provided below:

$$Steady : \boldsymbol{v}(x, y, z) = \begin{bmatrix} u(x, y, z) \\ v(x, y, z) \\ w(x, y, z) \end{bmatrix}, \tag{2.1}$$

$$Unsteady : \boldsymbol{v}(x, y, z, t) = \begin{bmatrix} u(x, y, z, t) \\ v(x, y, z, t) \\ w(x, y, z, t) \end{bmatrix}, \tag{2.2}$$

where x, y, and z denote the coordinates of space; t denotes the time; u, v, and w represent components of vector field.

2.1.2 Vector Calculus

In this Section, we will review operators associated with vector calculus and explore their properties. We begin with the definition of the gradient operator, denoted as ∇, which is expressed as follows:

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)^{\intercal}. \tag{2.3}$$

The operator ∇ is commonly employed to simplify partial differential equations (PDEs) involving specific partial derivative quantities. When applied to a scalar field, the gradient yields a vector field. For example, taking the gradient of a temperature field results in a vector field that signifies the rate of change of the temperature field and the direction of that change at each point in the field.

The divergence of a vector field is a scalar field, with each scalar quantity representing the volume change of a virtual finite-sized sphere associated with

the vector field. The mathematical expression for the divergence of $\boldsymbol{v}$ is given by:

$$\nabla \cdot \boldsymbol{v} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = u_x + v_y + w_z. \tag{2.4}$$

When the vector field represents a flow field, the divergence of the vector field is associated with the volume change at each point in the flow field. The sign of $\nabla \cdot \boldsymbol{v}$ indicates the nature of the change: positive signifies volume increase, while negative signifies volume decrease. In particular, if $\nabla \cdot \boldsymbol{v}(x, y.z) = 0$ holds for the entire domain of the flow field, then the flow is considered incompressible or divergence-free.

The curl of a vector field describes the rotation or spin of a vector field at each point. For a 3D vector field, the curl, denoted as $\nabla \times \boldsymbol{v}$, is given by:

$$\nabla \times \boldsymbol{v} = \begin{pmatrix} \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \\ \frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \\ \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \end{pmatrix} = \begin{pmatrix} w_y - v_z \\ u_z - w_x \\ v_x - u_y \end{pmatrix}. \tag{2.5}$$

Geometrically, the curl represents the tendency of the vector field to rotate around a point. The magnitude of the curl at a specific point is proportional to the local rotation intensity, and the curl vector's direction indicates the rotation axis. In physics and fluid dynamics, the curl is used to characterize the vorticity of a fluid flow. Vorticity measures the local rotation of fluid particles within the flow. a vector field is considered curl-free or irrotational if its curl is the zero vector everywhere in the domain.

2.1.3 Jacobian and Jacobian Invariants

For a 3D steady flow in a Cartesian system, the mathematical format of the spatial Jacobian $\boldsymbol{J}$ is defined as follows:

$$\boldsymbol{J} = \nabla \boldsymbol{v} = (\boldsymbol{v}_x, \boldsymbol{v}_y, \boldsymbol{v}_z) = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{pmatrix}. \tag{2.6}$$

The eigenvalue and eigenvector of the Jacobian are usually related to analysing the local flow pattern at a given location. The Jacobian matrix can be decomposed into the symmetric matrix ($\boldsymbol{S}$) and the corresponding anti-symmetric matrix ($\boldsymbol{\Omega}$). They are also known as the strain rate tensor and the vorticity tensor, respectively.

$$\boldsymbol{J} = \boldsymbol{S} + \boldsymbol{\Omega}, \tag{2.7}$$

where $\boldsymbol{S}$ and $\boldsymbol{\Omega}$ have the following forms:

$$\boldsymbol{S} = \frac{\boldsymbol{J} + \boldsymbol{J}^{\intercal}}{2} \tag{2.8}$$

and

$$\boldsymbol{\Omega} = \frac{\boldsymbol{J} - \boldsymbol{J}^{\intercal}}{2}. \tag{2.9}$$

In addition to capturing the local flow pattern, computing the invariants of the Jacobian matrix helps provide various definitions of vortex regions. It facilitates the design of relevant measures for vortex extraction. In the case of a 3D steady flow, the eigenvalues (λ) of the Jacobian matrix of the vector field satisfy the following condition:

$$\lambda^3 - P\lambda^2 + Q\lambda - R = 0, \tag{2.10}$$

where P, Q, and R are Jacobian's invariants quantites. And they are defined as follows:

$$P = tr(\boldsymbol{J}) = \nabla \cdot \boldsymbol{v}, \tag{2.11}$$

$$Q = \frac{1}{2}(P^2 - tr(\boldsymbol{J}^2)) = \frac{1}{2}P^2 + \frac{1}{2}(\parallel \boldsymbol{\Omega} \parallel^2 - \parallel \boldsymbol{S} \parallel^2), \tag{2.12}$$

$$R = det(\boldsymbol{J}). \tag{2.13}$$

The $\parallel \cdot \parallel$ notion represents Euclidean norm. When analysing a 2D steady flow, the corresponding Jacobian matrix only has two invariants, i.e., P and Q, where $Q = R = det(\boldsymbol{J})$.

2.2 The Topology of Vector Field

The essential part of understanding the vector field topology is calculating the first-order critical points. In vector field topology, critical points usually refer to local minima, maximum, or saddle of a specific type of scalar field. According to Helman and Hesselink (1989), critical points of a velocity field are typically characterized by zero velocity magnitude. Thus, if $\boldsymbol{x}_c$ is such a critical point, then:

$$\boldsymbol{v}(\boldsymbol{x}_c) = \boldsymbol{0}. \tag{2.14}$$

In addition, Helman and Hesselink (1989) point out that these critical points are ideally isolated, i.e., their surrounding velocity field has a nonzero velocity magnitude. Moreover, analysing their properties and categorising them will further require the corresponding field to be differentiable, and the

Jacobian of critical point is non-singular. Once those requirements are met, the corresponding eigenvalues and eigenvectors will indicate the neighboring flow pattern of the critical point. For this thesis, we are only focusing on 2D vortical behavior.

Six specific critical points for a 2D steady field are summarized and displayed in figure 2.1. They can be classified according to the eigenvalues (Helman and Hesselink, 1989) of the Jacobian matrix of the velocity field at the position of the critical point. These patterns can be categorised into two groups regarding the imaginary part of the eigenvalues. Usually, the nonzero imaginary part relates to a local spirling pattern. The sign of the corresponding real parts would further indicate whether the detected vortical center is attracting focus, repelling focus, or center. The concept of critical points could also be extended to a 3D steady field. In that scenario, an additional eigenvector and corresponding eigenvalue are required to compute the types of critical points (Helman and Hesselink, 1991). Moreover, these points may perform a specific movement (e.g., parabolic, hyperbolic, and elliptic motion) in the nearby 3D field (Scheuermann et al., 1998). However, this technique is for finding critical points in a steady field. For an unsteady flow field, motions of fluid particles are time-dependent and, thus, do not contain any critical points. In addition, attracting and repelling nodes do not exist when the flow is incompressible.

2.3 The Invariance of Reference Frame

Choosing the proper reference frame is crucial to observing the vortex properly, as the relative motion between the vortex flow and the observer significantly influences the detection results. Günther and Theisel (2018) have summarized existing vortex extraction methods according to the reference frame invariance into four major categories, i.e., No invariance, Galilean invariance, Rotation invariance, and Objectivity.

No invariance: Identification techniques of no invariance imply that it is only suitable for the steady flow, or the correct reference frame was selected when the flow behaves approximately steady. Except for the first category, techniques of the remaining three categories are invariant when reference frames perform selected motion types.

Galilean invariance: A vortex extraction technique is considered Galilean invariant when the chosen reference frame undergoes equal-speed translation motion and the observed flow appears steady. In other words, Galilean invariant techniques can accurately measure vortices during equal-speed translation motion. However, if the reference frame or vortex undergoes different types of

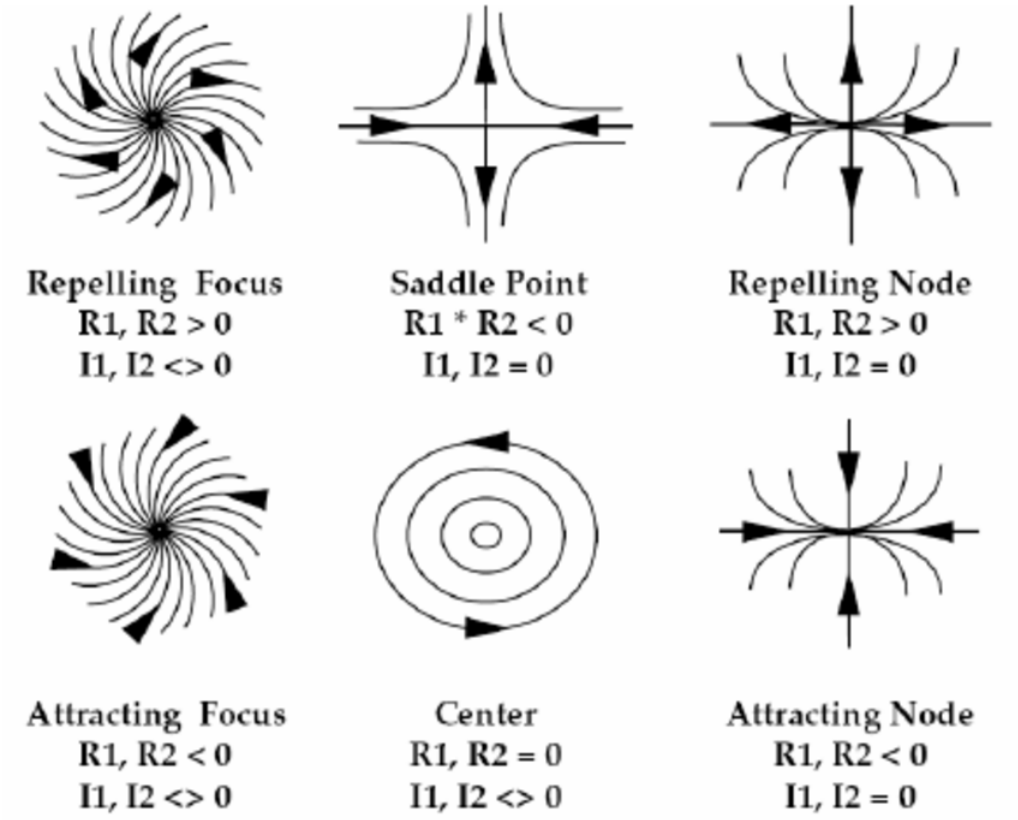


Figure 2.1: *Six typical critical points of a 2D steady vector field. R1 and R2 represent the real parts of the eigenvalues of the Jacobian at critical points, and I1 and I2 denote the corresponding imaginary parts, respectively. For instance, if R1 and R2 are positive and I1 and I2 are not equal to 0, it corresponds to a repelling focus. On the other hand, R1 and R2 are negative, while I1 and I2 are not equal to zero, corresponding to an attracting focus. Image adapted from Menelas et al. (2009).*

motion, the methods in this category may incorrectly measure vortices.

Rotation invariance: In contrast to Galilean invariance, Rotation invariance refers to the reference frame performing an equal angular speed of rotational motion around a known center, i.e., the rotational invariant method can identify vortices that perform an equal angular speed of rotations. However, without a known rotation center in advance, the methods of this category will fail to measure vortices.

Objectivity: In a real scene, vortices may perform any motion that combines translation and rotation. Thus, vortex extraction techniques desire to be objective when analysing them (Haller, 2015; Haller et al., 2016; Günther et al., 2017). Objectivity, in this context, implies that measurements remain consistent when the reference frame undergoes a combination of smooth rotation and translation. Objective extraction methods can accurately identify vortices during smooth rotations and translations. A formalized definition of objectivity can be found in Truesdell and Noll (1965).

2.4 Overview of Vortex Identification Methods

This Section provides an overview of primary vortex identification techniques. In nature, a vortex region is characterized by fundamental properties such as pressure and vorticity. Consequently, vortex identification typically involves extracting the vortex region based on thresholds for these characteristics. Basic methods include the pressure threshold method and the vorticity threshold method. More sophisticated approaches incorporate computations of the velocity gradient (Jacobian) and the corresponding Jacobian invariant. Notable examples are the Q-criterion (Hunt, 1987), λ_2-criterion (Jeong and Hussain, 1995), and Δ-criterion (Chong et al., 1990). Additionally, popular methods include the Γ method (Graftieaux et al., 2001), the Instantaneous Vorticity Deviation (IVD, Haller et al. (2016)), and its Lagrangian version, the Lagrangian-Averaged Vorticity Deviation (LAVD) (Haller et al., 2016). Table 2.1 summarises and categorises these vortex identification techniques based on their reference frame invariance.

2.4.1 Pressure Method

If the viscosity of a 2D steady flow is neglectable, a vortex region and a low-pressure region usually coincide. In this case, the vortex center can be found at the pressure minimum of the region. Via estimating a reasonable pressure

Reference frame invariance	Vortex identification technique
Galilean invariance	vorticity magnitude Q-criterion (Hunt, 1987) Δ-criterion (Chong et al., 1990) λ_2-criterion (Jeong and Hussain, 1995) Swirling strength (Zhou et al., 1999) Γ method (Graftieaux et al., 2001)
Objectivity	Pressure method (Hunt et al., 1988) IVD (Haller et al., 2016) LAVD (Haller et al., 2016)

Table 2.1: Popular vortex identification techniques.

criterion (p_t), a region with pressure below the threshold (i.e., p_t) in 2D flow is extracted and regarded as a vortex region by Hunt et al. (1988):

$$p \leq p_t. \tag{2.15}$$

However, the criterion might fail in the case of unsteady viscous flow or 3D flow. The reason is that the pressure minimum may not exist in those cases (Cucitore et al., 1999). In addition, the criterion might fail to distinguish each vortex when the flow field has muti-vortices that coexist and are adjacent.

2.4.2 Vorticity Method

The vorticity, denoted by $\boldsymbol{\omega}$, of a vector field is equivalent to the curl of the vector field and can be directly employed to identify vortices:

$$\boldsymbol{\omega} = \nabla \times \boldsymbol{v}. \tag{2.16}$$

As previously mentioned, strong vorticity is a crucial characteristic of vortex flow. In the vorticity method, a region is considered a vortex flow if its vorticity is above a certain reasonable threshold ($|\boldsymbol{\omega}_t|$, typically determined by the specific scenario):

$$|\boldsymbol{\omega}| \geq \boldsymbol{\omega}_t. \tag{2.17}$$

In this approach, the vorticity maximum is regarded as the vortex center.

In a 2D flow field, the vorticity vector is perpendicular to the horizontal flow field, with its sign indicating the flow's direction (positive for counterclockwise, negative for clockwise). However, a notable drawback of this thresholding method is its susceptibility to the false positive detection problem of shear rotation (Lugt, 1979; Robinson, 1991). It struggles to differentiate between a

swirling pattern and shear flow, as both are characterized by strong vorticity. The method merely captures the isocontour above the pre-defined threshold, making it insufficient to delineate the vortex's proper boundary accurately. Additionally, setting a vorticity threshold may result in non-detection during vortices' early or ending stages, especially when characterized by weak vorticity.

2.4.3 Q-criterion and Okubo–Weiss Criterion

When dealing with a 3D steady flow field that is divergence-free, expressed as $\nabla \cdot \boldsymbol{v} = 0$, the Q-invariant (see Equation (2.12)) of the spatial Jacobian matrix of the velocity field can be written in the following form:

$$Q = \frac{1}{2}(\| \boldsymbol{\Omega} \|^2 - \| \boldsymbol{S} \|^2). \tag{2.18}$$

Hunt (1987) introduced Q-criterion when comparing the magnitude between vorticity tensor $\boldsymbol{\Omega}$ and strain rate tensor $\boldsymbol{S}$. If the Q value exceeds 0, i.e., $\boldsymbol{\Omega}$ is greater than $\boldsymbol{S}$, the region will be extracted as a vortex region. Moreover, the pressure minimum requirement (i.e., the vortex region has lower pressure than the ambient pressure) is auxiliary to defining a vortex region properly.

Okubo (1970) and Weiss (1991) introduced the Okubo–Weiss criterion independently when identifying a vortex region within a 2D divergence-free steady flow. The mathematical definition of the criterion is defined as follows, and it is equal to the negative of the determinant of the corresponding 2D spatial Jacobian $\boldsymbol{J}$:

$$W = \frac{\partial u}{\partial y}\frac{\partial v}{\partial x} + \frac{\partial v}{\partial y}^2 = -det(\boldsymbol{J}) < 0. \tag{2.19}$$

Considering the flow in 2D and divergence-free, Equation (2.12) can be further simplified in the following form:

$$Q = \frac{1}{2}(P^2 - tr(\boldsymbol{J}^2)) = det(\boldsymbol{J}). \tag{2.20}$$

Obviously, the Okubo–Weiss criterion is equivalent to a 2D version of the Q-criterion.

2.4.4 λ_2-criterion

Jeong and Hussain (1995) introduced the λ_2-criterion to identify the vortex region when searching for a pressure minimum. When dealing with the incompressible flow, the reduced strain rate transport equation can be expressed in

the following form after ignoring the viscosity effect and the unsteady irrotational straining:

$$\boldsymbol{S}^2 + \boldsymbol{\Omega}^2 = -\frac{1}{\rho}\nabla(\nabla p), \tag{2.21}$$

where ρ denotes the flow's density, and p denotes the pressure. Jeong and Hussain (1995) consider a region as vortex region when $\boldsymbol{S}^2 + \boldsymbol{\Omega}^2$ has two negative eigenvalues. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ represent the eigenvalues of $\boldsymbol{S}^2 + \boldsymbol{\Omega}^2$. The λ_2-criterion refers to $\lambda_2 < 0$. In addition, Jeong and Hussain (1995) noticed that the Q-criterion and the λ_2-criterion have the following relationship:

$$Q = -\frac{1}{2}tr(\boldsymbol{S}^2 + \boldsymbol{\Omega}^2) = -\frac{1}{2}(\lambda_1 + \lambda_2 + \lambda_3). \tag{2.22}$$

2.4.5 Δ-criterion and the Swirling Strength Criterion

The Δ criterion introduced by Chong et al. (1990) is related to computing Jacobian's characteristic Equation (see Equation 2.10) and defines a vortex region if it has complex eigenvalues. The characteristic Equation described above has the following discriminant, and it is called the Δ-criterion:

$$\Delta = \left(\frac{\tilde{Q}}{3}\right)^3 + \left(\frac{\tilde{R}}{2}\right)^2, \tag{2.23}$$

where

$$\tilde{Q} = Q - \frac{P^2}{3} \tag{2.24}$$

and

$$\tilde{R} = -R - \frac{2P^3}{27} + \frac{PQ}{3}. \tag{2.25}$$

In the case of divergence-free flow, i.e., $P = \nabla \cdot \boldsymbol{v} = 0$, thus, $\tilde{Q} = Q$ and $\tilde{R} = -R$. Equation (2.23) become

$$\Delta = \left(\frac{Q}{3}\right)^3 + \left(\frac{R}{2}\right)^2. \tag{2.26}$$

A vortex region is extracted if the corresponding $\Delta > 0$ implies the corresponding characteristic Equation has complex solutions (i.e., one real eigenvalue and the remaining two eigenvalues are conjugate complexes). Through analysis of the compressibility effect on the vortex extraction, Kołár (2009) derived a compressible version of the Δ-criterion, which can be used to identify the vortex region when the flow is compressible.

Stem from the Δ-criterion, Zhou et al. (1999) introduced a similar vortex extraction criterion, i.e., the swirling strength criterion, to detect the region of a vortex. The criterion would extract a vortical structure when a pair of complex conjugate eigenvalues of the corresponding velocity gradient exists. The magnitude of imaginary parts of the eigenvalues indicates the detected region's swirling strength. For 2D flow, the real part of the complex conjugate solution indicates whether the vortex region is diverging or converging (see Figure 2.1). Like the Δ-criterion, the swirling strength criterion is extendable and can be applied to compressible flows (Kołár, 2009). In addition, the Q-criterion, Δ-criterion, and λ_2-criterion yield similar detection results when detecting vortex within incompressible 2D flows (Jeong and Hussain, 1995).

The swirling strength method has been applied to identify solar vortices simulated on the solar atmosphere by Moll et al. (2011). Besides, the method was also adopted by Kato and Wedemeyer (2017), Yadav et al. (2020), and Canivete Cuissa and Steiner (2020) when analysing the properties of solar vortices.

2.4.6 Γ Method

Different from the above methods based on the velocity derivatives, Graftieaux et al. (2001) introduced Γ_1 and Γ_2 to detect the center and boundary of a vortex, respectively. Γ methods can identify vortical structures when proper thresholds are chosen. The Γ_1 method has the following form:

$$\Gamma_1(P) = \frac{1}{N} \sum_S \frac{(\boldsymbol{PM} \times \boldsymbol{U}_M) \cdot \boldsymbol{n}}{\| \boldsymbol{PM} \| \cdot \| \boldsymbol{U}_M \|} = \frac{1}{N} \sin(\theta_M), \tag{2.27}$$

where N denotes the number of discretely sampled velocity points of a 2D rectangular region S surrounding P. M is a point or a set of points lies within S, and $\boldsymbol{PM}$ denotes the displacement vector from P to M. $\boldsymbol{U}_M$ denotes the velocity vectors at location of M, and $\boldsymbol{n}$ is the unit vector normal to the flow field. The angle between $\boldsymbol{U}_M$ and $\boldsymbol{PM}$ denoted in θ_M.The $||\cdot||$ represents the Euclidean norm.

The Γ_1 function derives a signed scalar field in the region S, and each field value measures the topology of the flow surrounding P. Notably, a $|\Gamma_1|$ maximum appears to coincide with the vortex center. Moreover, the sign of Γ_1 at the center indicates the rotation direction of the vortex, i.e., negative indicates clockwise rotation and positive for anti-clockwise. The $|\Gamma_1|$ values have been normalized between 0 and 1. The vortex center will have $|\Gamma_1|$ values equal to 1 in the ideal case when the vortex is axisymmetric. However, the

appearance of the vortex contains a broad range of irregular shapes due to their dynamics and interaction with the surrounding environment. Thus, the threshold of Γ_1 function for the vortex is less than 1.

In contrast to Γ_1, the Γ_2 function considers the local velocity field neighboring the vortex center. The definition of the discrete version of Γ_2 is:

$$\Gamma_2(P) = \frac{1}{N} \sum_S \frac{[\boldsymbol{PM} \times (\boldsymbol{U}_M - \overline{\boldsymbol{U}}_P)] \cdot \boldsymbol{n}}{\parallel \boldsymbol{PM} \parallel \cdot \parallel \boldsymbol{U}_M - \overline{\boldsymbol{U}}_P \parallel}. \tag{2.28}$$

Here, $\overline{\boldsymbol{U}}_P$ refers to the local average velocity around the point P, i.e.,

$$\overline{\boldsymbol{U}}_P = \frac{1}{N} \sum_S \boldsymbol{U}. \tag{2.29}$$

In the case of incompressible flow, for a tiny 2D region ($S \to 0$), $\Gamma_2(P)$ function has a strong relationship with the rotation rate $\boldsymbol{\Omega}$, and the eigenvalue of the symmetrical matrix (μ). This relationship shows that if $|\Gamma_2| > 2/\pi$, the current flow is dominant by rotation, and $|\boldsymbol{\Omega}/\mu| > 1$. The exact relationship is not defined when varying the size of the rectangular region S (Graftieaux et al., 2001). The definition of Γ_2 ensures it is Galilean invariance.

2.4.7 Fast Computation of Γ_1

To accelerate the computation efficiency when processing a huge volume of data, Zigunov et al. (2020) introduces the fast computation of Γ_1, which is based on the Graftieaux et al. (2001)'s Γ_1 method (see Equation (2.27)). When analysing the 2D flow and assuming the field lies in the $x-$ and $y-$ horizontal plane of a Cartesian coordinate system, the unit normal vector $\boldsymbol{n}$ is parallel to the $z-$ axis and denoted as unit vector $\boldsymbol{z}$. Thus, by separating the x and y components of $\boldsymbol{PM}$ (the displacement vector) and $\boldsymbol{U}_M$ (the velocity vector), respectively, Equation (2.27) can be rewritten into the following form:

$$\Gamma_1(P) = \frac{1}{N} \sum_S \frac{(PM_x U_{My} - PM_y U_{Mx})}{\parallel \boldsymbol{PM} \parallel \cdot \parallel \boldsymbol{U}_M \parallel}, \tag{2.30}$$

where U_{Mx} and U_{My} correspond to the x and y components of the velocity vector at the point M. PM_x and PM_y are the corresponding components of displacement vectors from P to M. Further, separating the numerator parts allows the 2D version of Γ_1 expressed as the operation of two convolutions:

$$\Gamma_1(P) = \frac{1}{N} \Big[\sum_S \frac{PM_x}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{My}}{\parallel \boldsymbol{U}_M \parallel} - \sum_S \frac{PM_y}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{Mx}}{\parallel \boldsymbol{U}_M \parallel} \Big], \tag{2.31}$$

where '⊛' denotes the convolution operator. The trigonometric representation of two convolution kernels is shown below:

$$\frac{PM_x}{\parallel \boldsymbol{PM} \parallel} = \cos\left(\theta_{PM}\right), \quad \frac{PM_y}{\parallel \boldsymbol{PM} \parallel} = \sin\left(\theta_{PM}\right). \tag{2.32}$$

Analogously, velocity vectors have the following geometric representation forms:

$$\frac{U_{Mx}}{\parallel \boldsymbol{U}_M \parallel} = \cos\left(\theta_{U_M}\right), \quad \frac{U_{My}}{\parallel \boldsymbol{U}_M \parallel} = \sin\left(\theta_{U_M}\right). \tag{2.33}$$

Therefore, the 2D trigonometric representation of Γ_1 is:

$$\Gamma_1(P) = \frac{1}{N}\left[\sum_S \cos\left(\theta_{PM}\right)\sin\left(\theta_{U_M}\right) - \sum_S \sin\left(\theta_{PM}\right)\cos\left(\theta_{U_M}\right)\right]. \tag{2.34}$$

Thus, the form given in Equation (2.31) is still an analysis of the geometry of the velocity field in the region S. This form of representation is straightforward, and current algorithms/libraries of mathematical tools have optimized their convolution operation efficiently. Figure 2.2 is a graphical diagram of Γ_1.

2.4.8 Instantaneous Vorticity Deviation (IVD)

Any subtraction of two vorticity values that sample at the same instance will erase the constant rotation motion of the coordinate of reference (Haller et al., 2016). In other words, subtraction between two vorticity values yields objective quantities, and their spatial derivatives are also objective. Therefore, Haller et al. (2016) introduces the instantaneous vorticity deviation (IVD) method to define a vortex region objectively from the vorticity. This method relates to performing a subtraction between each sampled vorticity ($\boldsymbol{\omega}$) and their spatial mean of vorticity ($\boldsymbol{\omega}_{avg}$) within a nearby region. The mathematical form of IVD is below:

$$\boldsymbol{IVD}(\boldsymbol{x}, t) = |\boldsymbol{\omega}(\boldsymbol{x}, t) - \boldsymbol{\omega}_{avg}(t)|. \tag{2.35}$$

In a 2D flow field, the vortex's center is located at a local IVD maximum, and the vortex boundary is extracted when the center is enclosed by the outer closed convex contour of the IVD field (Haller et al., 2016). This closed contour also relates to particle trajectories, while those flow particles have nearly the same intrinsic rotational rate (Haller et al., 2016). Therefore, IVD defines the vortex boundary such that the particles inside the boundary rotate coherently

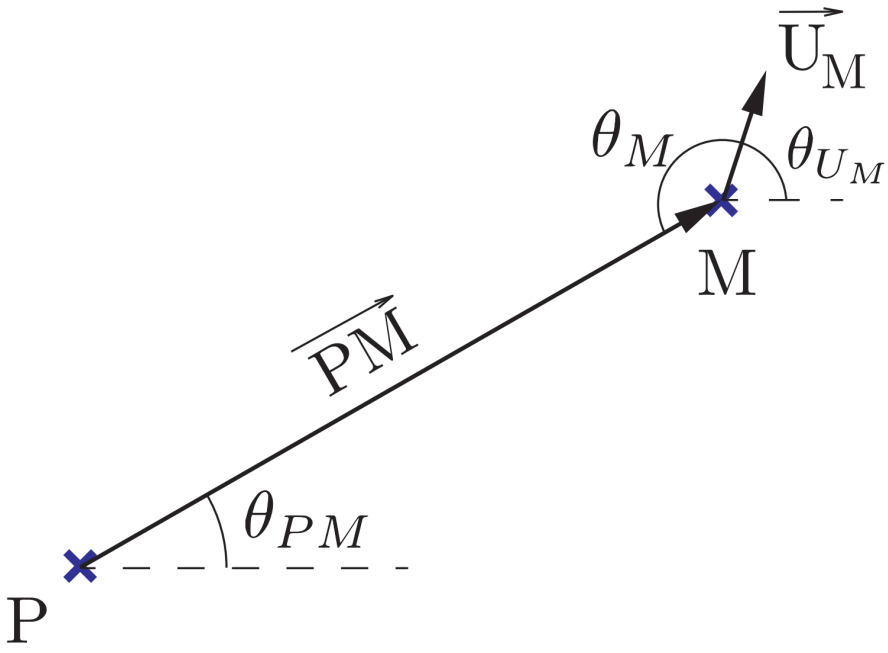


Figure 2.2: *Graphical interpretation of the integrand of the* Γ_1 *computation as defined by Graftieaux et al. (2001). Courtesy of Zigunov et al. (2020).*

in an elliptical trajectory. The IVD is convenient and is only needed to determine the convexity deficiency before the calculation. The convexity deficiency is defined as the maximum deviation from convexity that a trajectory of those particles can have. The convexity deficiency is defined in the following form:

$$\varepsilon = \frac{A_c - A_{ch}}{A_c}. \tag{2.36}$$

Here A_c indicates the area enclosed by a closed curve, and the area enclosed by the corresponding convex hull is given by A_{ch}.

2.4.9 Lagrangian-averaged Vorticity Deviation (LAVD)

Besides the IVD, Haller et al. (2016) also introduced the Lagrangian version of the IVD, i.e., Lagrangian-averaged vorticity deviation (LAVD), which takes into account the evolution of the coherent structure by integrating the IVD. The mathematical expression of the LAVD is below:

$$\boldsymbol{LAVD}_{t0}^{t}(\boldsymbol{x}) = \int_{t0}^{t} |\boldsymbol{\omega}(\boldsymbol{x}, t) - \boldsymbol{\omega}_{avg}(t)| dt. \tag{2.37}$$

Like the IVD, the LAVD is an objective measurement, and Haller et al. (2016) has provided further details about it. The detection result depends on the size of the local region that is used to calculate the $\boldsymbol{\omega}_{avg}$. However, the IVD and the LAVD may yield false detection in shear flow rotation, while the strong shear flow region is also characterized by high vorticity. Based on the streamlines of the displacement vector, Silva et al. (2018) introduced a novel parameter d to remove the false detection result from the IVD and the LAVD. Notices, the analysis of Haller et al. (2016) shows that the LAVD requires the flow to be incompressible.

2.5 Overview of Clustering Algorithms

Advances in observation and storage technology have already assisted in solar research. Huge-volume data sets have been created and stored on the way of exploration. With such a massive scale of data, there is an urgent need to develop a suitable, automatic, efficient, and practical data analysis strategy for the given data. In machine learning, data analysis deals with predictive modeling; given the training data generated by a system, the goal is to predict the behavior of the unseen data. This process is also called learning.

There are two major categories of learning problems: supervised and unsupervised. In supervised learning, the training data set has been labeled. In other words, examples of a system's input (training data) and output (labels) are provided. The goal is to build a model that approximates this system and is used to deal with unseen data so that the model's output is similar to those example labels. Conversely, unsupervised learning tasks lack labeled data, requiring a principled approach to discern patterns and clusters within the data. Data clustering, fundamental to unsupervised learning, involves partitioning data based on similarities, revealing natural groupings, and enhancing our understanding of underlying structures. This Section will overview several fundamental clustering techniques and evaluate them in different scenarios.

2.5.1 The K-means Clustering Algorithm

The K-means algorithm, developed by Lloyd (1982), is one of the most fundamental clustering techniques that aims to partition a group of data into K different clusters, and it is required to determine K before the calculation.

Let $X = \{x_i | i = 1, ..., N\}$, and x_i is n-dimensional data; let $C = \{C_j | j = 1, ..., K\}$ and C_j stand for the j^{th} cluster; for all clusters and each data in C_j, the K-means algorithm aims to calculate the minimum value of within-cluster sums of squares (WCSS) defined below

$$J(C) = min \sum_{j=1}^{K} \sum_{x_i \in C_j} \parallel x_i - \mu_j \parallel^2, \tag{2.38}$$

where μ_j stands for the centroid of C_j. The K-means algorithm is an iterative calculation. The calculation steps of this algorithm are as follows

1. Initialize cluster centroid of K cluster.
2. Calculate the distance between each data and each centroid.
3. Assign each data to the cluster with the closest centroid.
4. Calculate the average of the data of each cluster and update all cluster centroid.
5. Iterate steps 2 to 4 until the sum of WCSS does not change significantly or exceed the maximum number of iterations.

Figure.2.3 shows an example of the K-means partitioning. In panel (a), green points indicate the data set before partitioning, and two black crosses denote

two initial cluster centroids. Panel (b) displays the partition result after the initial clustering. Panel (c) shows an intermediate partition result. The final result is shown in panel (d).

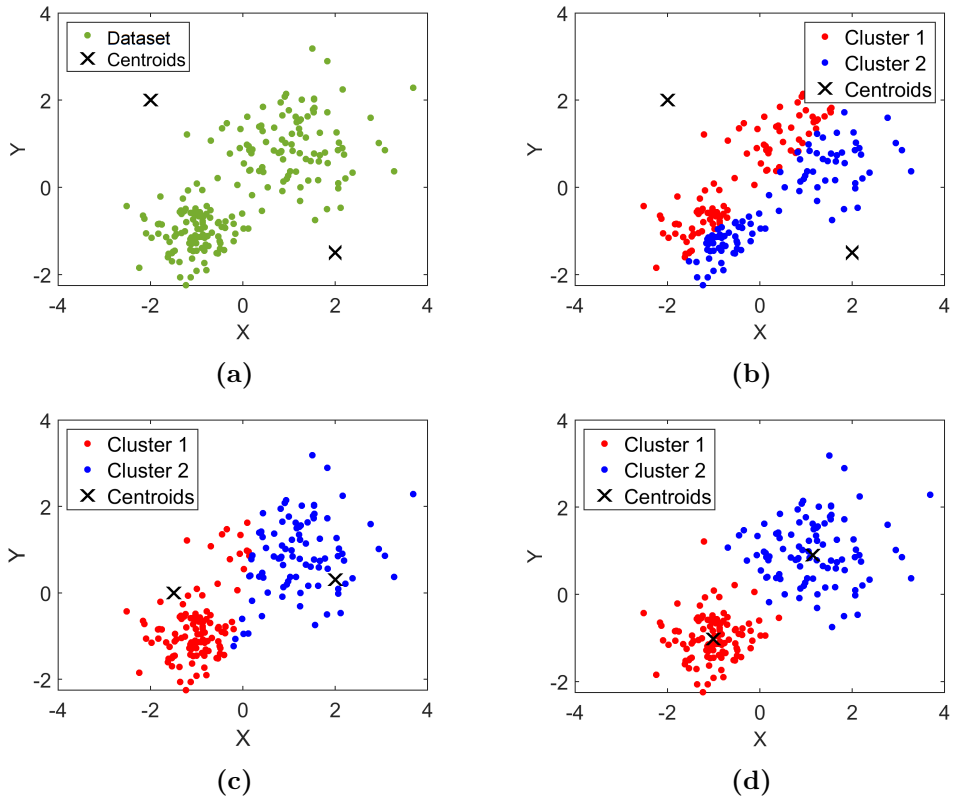


Figure 2.3: *Example of the K-means clustering for a 2D dataset (green) into two clusters (red and blue). The Black cross indicates cluster centroids. (a) Datas before clustering (b) Initial clustering result. (c) Intermediate clustering result. (d) Final clustering result.*

The optimal number of clusters can be determined using the elbow point method. This involves calculating the overall within-cluster sum of squares (WCSS) for different numbers of clusters (k) and plotting it against k. The resulting plot typically resembles the shape of an elbow. The optimal k corresponds to the point where the slope of the WCSS plot changes significantly, known as the elbow point. WCSS is a measure used to quantify the variation of the data within their clusters. A low WCSS value indicates that data within clusters are compact, while a high value suggests significant variability. In Chapter 4, we introduce a novel vortex categorizing strategy based on the

K-means algorithm and apply it to the clustering analysis of solar vortices behaviour.

2.5.2 Gaussian Mixture Models

Mixture models are a powerful tool for unsupervised clustering. It assumes that the data results from a combination of different probability distributions. Let x_i represent a data point in dataset X. The mixture model defines the probability density function (PDF) of x_i (that is, $f(x_i, \mathbf{\Theta})$) as a linear combination of a series of unknown distributions, and corresponding coefficients are weighting these distributions:

$$f(x_i, \mathbf{\Theta}) = \sum_{p=1}^{k} \alpha_p f_p(x_i, \mathbf{\Theta_p}), \tag{2.39}$$

where p is the index of the current probability distribution of x_i; $f_p(\cdot)$ is the p^{th} individual PDF of the x_i; k is the total number of all individual probability distributions being used to assume in the calculation; each $f_p(\cdot)$ has its parameter $\mathbf{\Theta_p}$; α_p is a weight coefficient corresponding to the $f_p(\cdot)$, so that the overall PDF of x_i (characterized by parameter $\mathbf{\Theta}$), that is, $f(x_i, \mathbf{\Theta})$, less than or equal to 1.

Before clustering, the mixture model is required to make an assumption about the total number of distributions (k) for the given data and the form of these distributions ($f_p(\cdot)$). In particular, when the Gaussian distribution model is being applied as a form of every $f_p(\cdot)$, the mixture model becomes the Gaussian mixture model (GMM). The form of the GMM is shown below

$$f(x_i, \mathbf{\Theta}) = \sum_{p=1}^{k} \alpha_p N_p(x_i, \mu_p, \sigma_p), \tag{2.40}$$

where μ_p and σ_p refer to the mean and the standard deviation of the Gaussian distribution $N_p(\cdot)$, respectively.

After determining the form of $f_p(\cdot)$, the next step is determining the corresponding α_p and $\mathbf{\Theta_p}$ of each distribution in order to fit the given data best. In the case of the GMM, the $\mathbf{\Theta_p}$ for each $f_p(\cdot)$ refer to the μ_p and the σ_p. The mixture models use the maximum-likelihood estimation (MLE) algorithm to estimate the $\mathbf{\Theta_p}$ of the given data. The MLE algorithm calculates the value of $\mathbf{\Theta}$ when taking the derivative of the log-likelihood function for $\mathbf{\Theta}$ and setting it equal to zero, that is,

$$\frac{\partial L(\mathbf{\Theta})}{\partial \mathbf{\Theta}} = 0, \tag{2.41}$$

where the log-likelihood function defined below

$$L(\mathbf{\Theta}) = \sum_{i=1}^{n} \log f(x_i, \mathbf{\Theta}). \tag{2.42}$$

And n is the total number of observations. The log operation shown here would simplify the calculation when canceling out the exponential operation inside the PDF. The optimal solution for the mixture model of the given data can be computed through an iterative optimization process, that is, the Expectation-Maximization (EM) algorithm.

Given an initial estimate of the parameters (k, $\mathbf{\Theta}$ and α_p), the EM algorithm begins the iterative calculation with two main procedures, that is, the E-step and the M-step.

1. E-step: for each data, the algorithm calculates the corresponding posterior probability of component membership of x_i in the p^{th} distribution with the current estimate of $\mathbf{\Theta}$, that is, $\tau_p(x_i, \mathbf{\Theta})$ (see Equation (2.43)).

$$\tau_p(x_i, \mathbf{\Theta}) = \frac{\alpha_p f_p(x_i, \mathbf{\Theta_p})}{f(x_i, \mathbf{\Theta})}. \tag{2.43}$$

   In the case of the GMM, with the initial α_p, μ_p, and σ_p, Equation (2.43) becomes

$$\tau_p^{(k)}(x_i) = \frac{\alpha_p^{(k)} N_p(x_i, \mu_p^{(k)}, \sigma_p^{(k)})}{f(x_i, \mathbf{\Theta^{(k)}})}. \tag{2.44}$$

2. M-step: with the calculated $\tau_p^{(k)}(x_i)$ in the E-step, the algorithm updates the weights and parameters, correspondingly (see Equation (2.45), (2.46), and (2.47)).

$$\alpha_p^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \tau_p^{(k)}(x_i), \tag{2.45}$$

$$\mu_p^{(k+1)} = \frac{\sum_{i=1}^{n} x_i \tau_p^{(k)}(x_i)}{\sum_{i=1}^{n} \tau_p^{(k)}(x_i)}, \tag{2.46}$$

$$\Sigma_p^{(k+1)} = \frac{\sum_{i=1}^{n} \tau_p^{(k)}(x_i)(x_i - \mu_p^{(k+1)})(x_i - \mu_p^{(k+1)})^T}{\sum_{i=1}^{n} \tau_p^{(k)}(x_i)}, \tag{2.47}$$

   where $\Sigma_p^{(k+1)}$ refer to the covariance matrix that contains the corresponding variance of each of the Gaussian models.

The GMM algorithm categorising data by iterating over these two steps until convergence.

2.5.3 Density-based Spatial Clustering of Applications with Noise

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a clustering technique introduced by Ester et al. (1996). The most distinctive feature of this method is its clustering of data according to the local density of each data point. Therefore, both data clustering in arbitrary shapes with different densities and noise data can be identified and separated by the DBSCAN algorithm. Different from the K means algorithm and the GMMs algorithm, the DBSCAN algorithm would specify the optimal number of clusters automatically during clustering.

Besides clustering data, data are also identified into three categories of points, that is, core points, border points, and noise points, in the DBSCAN. These three kinds of points are related to two hyperparameters, that is, epsilon and minpts. Both are numeric parameters where epsilon is used to specify the radius of the search region around a point, and minpts is used to specify the minimum number of neighboring data required for a core point. The definitions of the three kinds of points are below

1. Core point: a core point in a cluster is surrounded by at least minpts points, and all these neighboring points are within epsilon of this point.

2. Border point: a border point in a cluster has neighboring points fewer than minpts within the epsilon of this point.

3. Noise point: A point that does not satisfy the condition of the core point and border point is a noise point. Such an outlier is not associated with any cluster.

The clustering procedure of the DBSCAN algorithm is below:

1. Pick the first data (x_1) from the whole dataset as the current point and initialize the first cluster with label C_1.

2. Calculate the relationship between the current point and its neighboring points within epsilon.

   (a) If the number of those neighboring points is more than minpts, the current point is one of the core points of the current cluster.

   (b) Otherwise, the current point is regarded as a noise point, and skip to step 4.

3. Update the current point by iterating over each neighboring point and repeat the examination procedure in step 2. Go to the next step once no new neighboring points are found that can be assigned to the current cluster.

4. Pick the subsequent fresh data as the current point and create a new cluster.

5. Iterate steps 2 to 4 until all data have been examined.

Caution: If a noise point later satisfies the condition of the border point of a cluster. The DBSCAN algorithm would relocate that noise point as a border point for that cluster.

As described above, the DBSCAN algorithm is required to determine the value of minpts and epsilon before the calculation. Ester et al. (1996) suggests an appropriate value of minpts should always be larger than the number of dimensions of the input data. As for epsilon, Ester et al. (1996) uses the k-distance graph to estimate the proper value of it. For each data within the dataset, calculate the average distance between it and the k points near it; sort the data point according to the average distance; plot the k nearest average distance as a function of points sorted with that nearest distance. The graph typically exhibits a 'knee,' and the corresponding location indicates the optimal epsilon value.

2.5.4 Silhouette Value Criterion

Rousseeuw (1987) develops the Silhouette method to evaluate the clustering result when calculating the corresponding Silhouette value of each data of different clusters. The Silhouette value ranges from -1 to 1, quantifying the wellness of clustering of the current data; that is, a high value implies data is well assigned to the current cluster, while a low value implies ill-assignment, and this data might have a higher silhouette value in other clusters. The Silhouette value of data i is defined below

$$S_i = \frac{b_i - a_i}{max(a_i, b_i)}, \tag{2.48}$$

where a_i is defined as the mean distance between data i and the other data belonging to the same cluster as data i; b_i is the minimum value of the mean distance between data i and other data belonging to different clusters. Besides measuring the clustering result of each data, the Silhouette value as a function of the cluster number can be used to determine the optimal cluster number for other clustering methods.

2.5.5 Calinski-Harabasz Criterion

Dedicated to evaluating the number of clusters for the given data, Caliński and Harabasz (1974) introduces the Calinski-Harabasz criterion values to estimate the optimal number of clusters. The Calinski-Harabasz criterion is also called the variance ratio criterion (VRC) and is defined as the following form

$$VRC_k = \frac{SS_b}{SS_w} \times \frac{(N-k)}{(k-1)}, \tag{2.49}$$

where SS_b is the between-cluster sum of squares; SS_w is the within-cluster sum of squares; N is the total number of the data; k is the cluster number.

The between-cluster sum of squares is defined as

$$SS_b = \sum_{i=1}^{k} n_i \parallel m_i - m \parallel^2, \tag{2.50}$$

where n_i is the number of data belonging to i^{th} cluster; the centroid of the i^{th} cluster is denoted as m_i; m is the total mean of the data; and $\| \cdot \|$ denote the $L2$ norm.

The within-cluster sum of squares is defined as

$$SS_w = \sum_{i=1}^{k} \sum_{x \in C_i} n_i \parallel x - m_i \parallel^2, \tag{2.51}$$

where x refers to a data point; C_i denote the i^{th} cluster.

A well-clustering result is characterized by a large SS_b and a small SS_w. Therefore, the highest Calinski-Harabasz value relates to the optimal number of clusters.

2.5.6 Comparison of Clustering Techniques

This Section provides three examples to illustrate how to select the optimal clustering technique for the given data. In particular, the three clustering techniques described above (the K-means algorithm, the GMM algorithm, and the DBSCAN algorithm) are being applied and compared. Moreover, we use the Silhouette value criterion and the Calinski-Harabasz criterion to evaluate the corresponding clustering results.

Example 1

In Figure 2.4, the example data (blue) is synthesized from four bivariate Gaussian distributions with different means vectors and the same covariance matrix. The corresponding covariance matrix is $\Sigma = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.3 \end{bmatrix}$. Each distribution generates 200 data points.

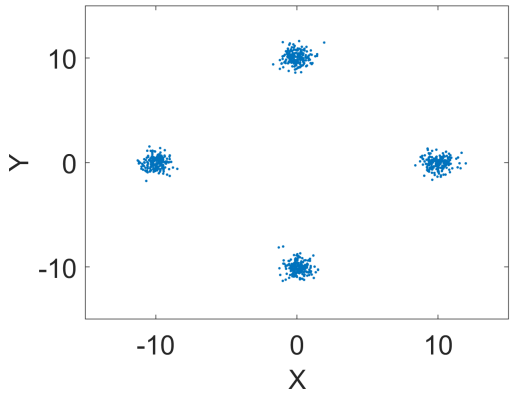


Figure 2.4: *Data set 1: synthesis data (blue) generated from four bivariate Gaussian distributions.*

Before applying the three clustering techniques, it is required to determine some pre-request parameters for different methods. In this example, it is intuitive to cluster data into four groups without knowing the distributions. Therefore, we assume the total number of distributions is 4 for the GMM algorithm; we set the epsilon equal to 1 and the minpts equal to 5 for the DBSCAN algorithm; for the K-means algorithm, the optimal cluster number is set equal to 4. With these parameters, the corresponding clustering results are displayed in Figure 2.5. The left column panels ((a), (c), and (e)) are the clustering results corresponding to the GMM algorithm, the DBSCAN algorithm, and the K-means algorithm, respectively. The right columns panels ((b), (d), and (f)) are the corresponding Silhouette values of each clustered data of the four clusters classified by the three methods. All methods have successfully clustered data into four groups denoted by four different colors (see panel (a), (c), and (e)). Most of the clustered data of the four clusters have high positive Silhouette values, which implies that those data are well categorized by the corresponding clustering methods (see panel (b), (d), and (f)). However, the DBSCAN algorithm incorrectly assigns three data points to noise clusters (purple points in panel (c)). In panel (d), their high negative

Silhouette values that are near -1 (cluster "-1" refers to the noise cluster) also imply they do not belong to the noise cluster. This is because the DBSCAN algorithm is susceptible to the choice of parameters (epsilon and minpts). In this example, an inappropriate epsilon is used to check the neighboring relationship of each data, which makes these three purple points neither satisfy the condition of the core point nor the border point defined by the DBSCAN algorithm. Therefore, they clustered into the noise cluster. This issue will be addressed if a proper epsilon is used.

A comprehensive evaluation is conducted through multiple independent tests to assess the performance of the three clustering methods on the given dataset. Figure 2.6 illustrates the mean Silhouette value obtained from 100 independent tests on dataset 1. The blue, red, and yellow data correspond to the GMM, DBSCAN, and K-means algorithms, respectively.

A lower mean Silhouette value may indicate that some data points are inaccurately assigned to the wrong cluster. In this example, the DBSCAN algorithm exhibits the most stable clustering membership, consistently yielding the same mean Silhouette value of 0.986 in all 100 tests. It is crucial to note that this mean value, although slightly below 1, is influenced by the fact that the DBSCAN algorithm consistently assigns the same data points as outliers due to inappropriate parameters, as seen in Figure 2.5 (c) and (d).

The clustering results obtained from the K-means algorithm indicate that the majority of the mean Silhouette values are high across 100 tests, with only one instance showing a low value. The GMM algorithm produces a slightly inferior result compared to the K-means algorithm, with three tests showing a low mean Silhouette value. These occurrences of low mean Silhouette values can be attributed to the tendency of the K-means and GMM algorithms to converge to a local optimum solution. In other words, the initial parameter estimates (initial centroids for the K-means algorithm; initial $\mathbf{\Theta}$ and α_p for the GMM algorithm) play a crucial role in influencing the solution result. Excluding these instances, the mean Silhouette values of the clustering results obtained from both the K-means and GMM algorithms are greater than 0.99, surpassing that of the DBSCAN algorithm.

Example 2: Clustering Dense Data

In Figure 2.7, the blue data points represent the example data set 2, which is generated from four bivariate Gaussian distributions (each of them synthesises 200 data points). The covariance of these distributions is identical to that of data set 1. Notably, in contrast to data set 1, the means of these distributions are designed to be closer, making it less intuitive to distinguish them into four

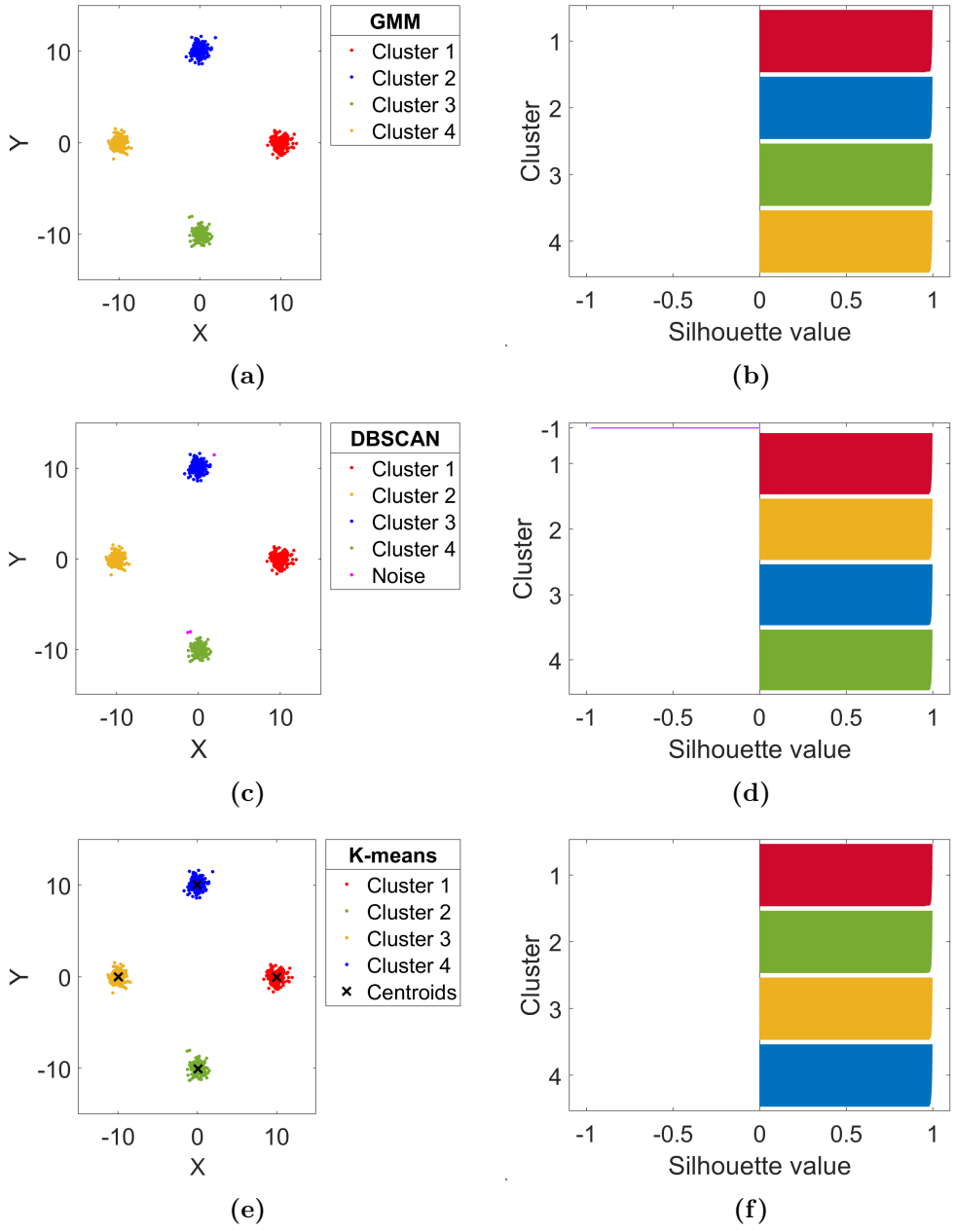


Figure 2.5: *Three clustering results of data set 1 from the GMM algorithm, the DBSCAN algorithm, and the K-means algorithm. (a) the clustering result of the GMM algorithm; (b) the Silhouette plot of the clustered data corresponding to panel (a); (c) the clustering result of the DBSCAN algorithm; (d) the Silhouette plot of the clustered data corresponding to panel (c); (e) the clustering result of the K-means algorithm; (f) the Silhouette plot of the clustered data corresponding to panel (e);*

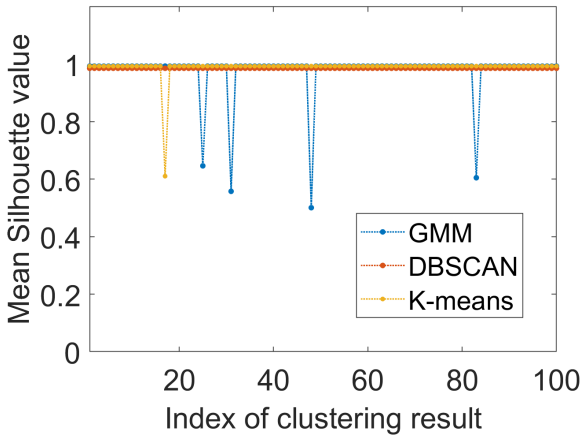


Figure 2.6: *The mean Silhouette value of the GMM algorithm (blue), the DBSCAN algorithm (red), and the K-means algorithm (yellow) in 100 independent tests in data set 1.*

clusters. To address this challenging scenario, we applied the three clustering techniques, i.e., GMM, DBSCAN, and K-means to data set 2 and evaluated the clustering results.

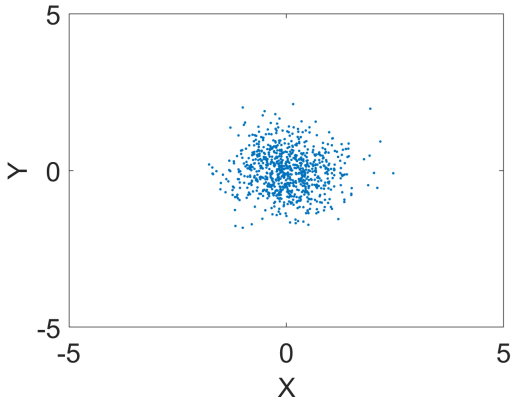


Figure 2.7: *Data set 2: synthesis data (blue) generated from four bivariate Gaussian distributions.*

As noted earlier, the only difference between data sets 1 and 2 lies in the means of each Gaussian distribution used to generate the data. Given this similarity, we utilize the same parameters as in example 1 when applying the clustering techniques to data set 2. The clustering results for the three techniques (GMM, DBSCAN, and K-means) are presented in Figure 2.8, with the

left column panels illustrating the clustering outcomes and the right column panels showing the corresponding Silhouette plots. The first row of panels corresponds to the application of GMM algorithm. Panel (a) shows data grouped into 4 clusters as intended, as we specified the initial estimate of the number of mixture models (k) to be 4. However, the GMM algorithm does not yield a satisfactory result, as a substantial portion of data in the 4^{th} cluster (yellow) exhibits a high negative Silhouette value (see panel (b)), indicating that they are incorrectly assigned.

Moving to the second row of panels (DBSCAN algorithm), it becomes evident that the DBSCAN produces an unsatisfactory clustering result, almost failing to cluster data with the specified parameter settings. Although the corresponding Silhouette plot suggests that most data belong to cluster 1 and have a positive value, this result contradicts the actual distribution. This outcome highlights the unsuitability of the DBSCAN algorithm for dense data, as it struggles to identify clusters in dense regions.

Clustering results of the K-means algorithm are presented in the third row of panels, yielding the most favorable clustering result, effectively dividing the data into 4 clusters denoted by distinct colors (see panel (e)). The locations of the four centroids provided by K-means (black crosses) are only slightly different from the means of the Gaussian distributions used to generate data set 2. In panel (f), the distribution of the 4 clusters is uniform, exhibiting similar Silhouette value behaviors. Low positive Silhouette values imply that these data points are near the border of their respective clusters. Compared to panels (b) and (d), the Silhouette plot in panel (f) aligns more closely with the actual distribution.

Similarly to example 1, we apply the three clustering techniques to data set 2 and conduct 100 independent tests to evaluate their performance. Figure 2.9 presents the mean Silhouette values from these 100 independent tests for the three methods, with blue, red, and yellow denoting the GMM, DBSCAN, and K-means algorithms, respectively. The DBSCAN algorithm consistently achieves the same Silhouette value (0.74) across the 100 tests, employing the same parameter settings. While this value is the highest among all clustering results by different methods, it is worth noting that the DBSCAN algorithm performs poorly on dense data, and the highest mean Silhouette values may not accurately reflect its effectiveness. This result suggests that the Silhouette value criterion might not be suitable for evaluating the DBSCAN algorithm when applied to dense data.

For the GMM algorithm, the mean Silhouette value (blue) fluctuates with a relatively large magnitude compared to other methods. This variability indicates that the GMM algorithm may not be a suitable clustering solution

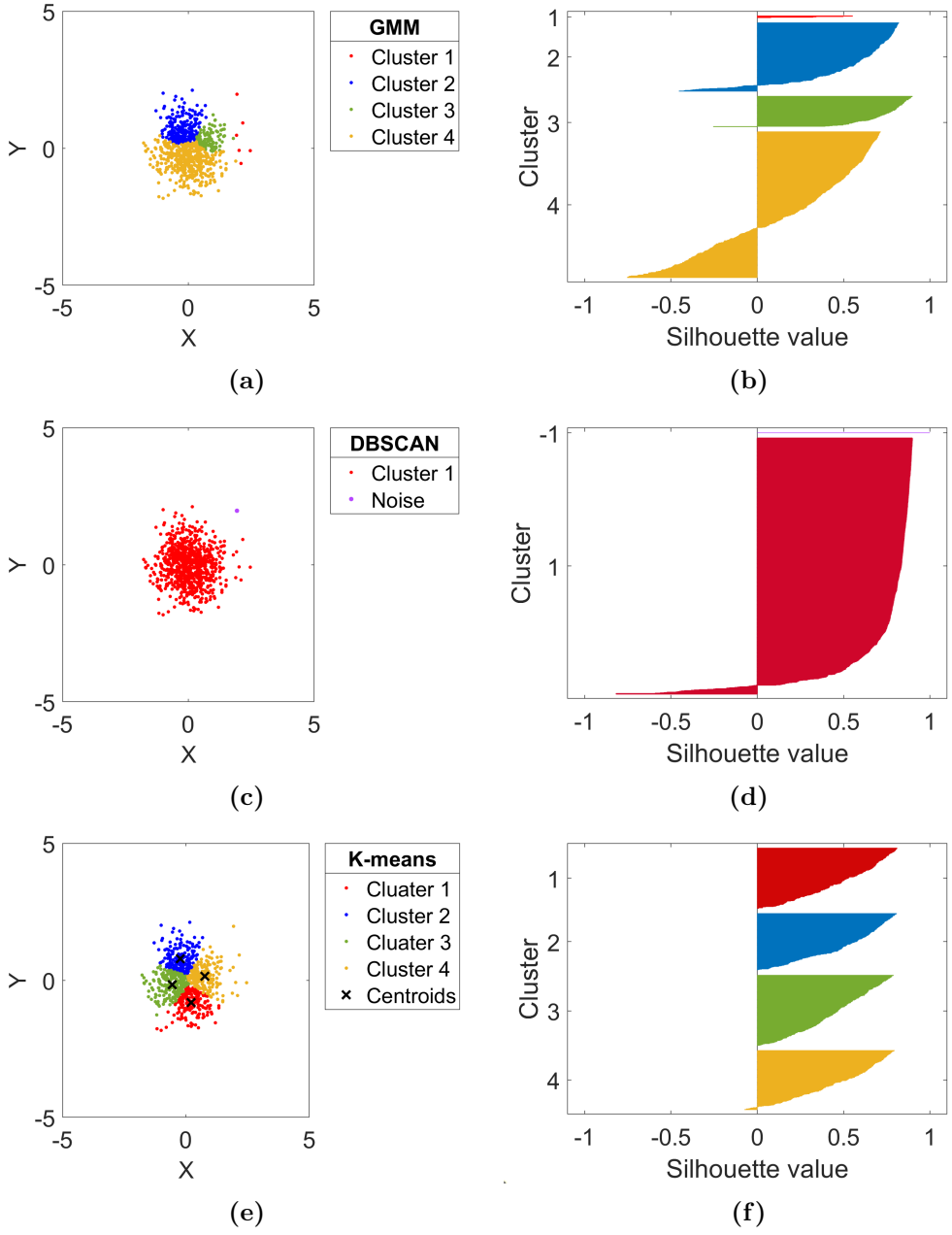


Figure 2.8: *Three clustering results of data set 2 from the GMM algorithm, the DBSCAN algorithm, and the K-means algorithm. (a) the clustering result of the GMM algorithm; (b) the Silhouette plot of the clustered data corresponding to panel (a); (c) the clustering result of the DBSCAN algorithm; (d) the Silhouette plot of the clustered data corresponding to panel (c); (e) the clustering result of the K-means algorithm; (f) the Silhouette plot of the clustered data corresponding to panel (e);*

for dense data. Although we expect that the GMM algorithm yields the best clustering result as the data set 2 generated from four bivariate Gaussian distributions, the result reflects that the GMM algorithm may fail in clustering dense data even if they are Gaussian distribution data.

In contrast, the K-means algorithm produces the most favorable clustering result, with only two tests exhibiting mean Silhouette values below 0.4. This outcome suggests that the clustering membership provided by the K-means algorithm remains stable even in dense data sets.

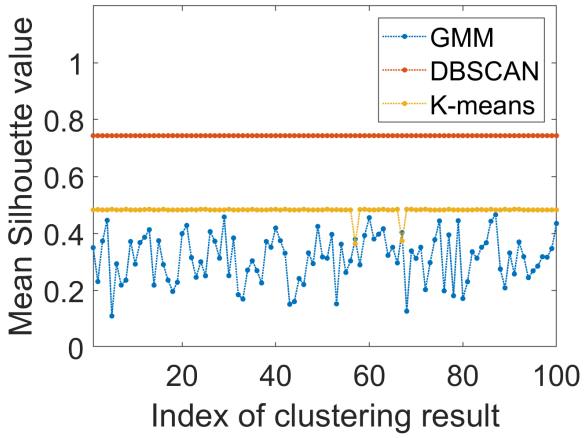


Figure 2.9: *The mean Silhouette value of the GMM algorithm (blue), the DBSCAN algorithm (red), and the K-means algorithm (yellow) in 100 independent tests in data set 2.*

In some scenarios, minimal information might be available about the given data, making it challenging to estimate the optimal number of clusters for the data intuitively. Unlike the GMM algorithm and the K-means algorithm, which require providing initial estimates for the number of mixture models and clusters, respectively, the DBSCAN algorithm autonomously determines the number of clusters based on the corresponding parameters. When the data generation process for data set 2 is unknown, and some unknown information may be revealed through data categorization into similar groups, a common question arises: how can unknown data be classified effectively? In other words, what is the optimal number of clusters for the given data?

Continuing with the dense data set 2, we apply the K-means algorithm with different cluster numbers and evaluate the corresponding clustering results. In Figure 2.10 (a), the blue data indicates the mean Silhouette values for clustering results as a function of different cluster numbers. The highest mean Silhouette values occur when the cluster numbers are equal to 3. If we solely

evaluate the K-means with the Silhouette values criterion, the optimal cluster number for the algorithm in this task would be 3, clustering the data set into three groups. However, this contradicts that data set 2 is generated from four distributions.

In Figure 2.10 (b), the blue data represents the within-cluster sum of squares (WCSS) as a function of cluster numbers, and the red data indicates the Calinski-Harabasz value as a function of cluster numbers. According to the elbow points method, the location of the elbow points of the WCSS indicates the optimal choice of the cluster number for the K-means algorithm. In this case, the blue data decline smoothly along the number of clusters (k) with no clear elbow points. However, we can still estimate that the elbow point corresponds to $k = 4$ or 5. The Calinski-Harabasz criterion considers both the within-cluster sum of squares (SS_w) and the between-cluster sum of squares (SS_b) to evaluate the clustering result. A high Calinski-Harabasz value suggests that the data are under better clustering than those with low Calinski-Harabasz values. The highest Calinski-Harabasz value corresponds to $k = 4$. Therefore, both the elbow point method and the Calinski-Harabasz criterion indicate that the optimal number of clusters is 4, which aligns with the actual data distribution.

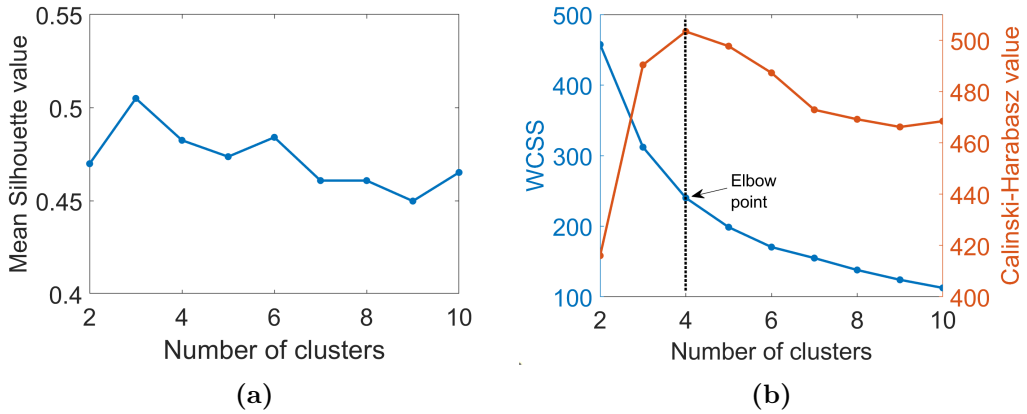


Figure 2.10: *(a) Mean Silhouette values as a function of cluster numbers of the K-means algorithm on data set 2; (b) the blue data denote the Within-cluster sum of square (WCSS) as a function of cluster numbers of the K-means algorithm on data set 2. The red data is the corresponding Calinski-Harabasz value.*

Example 3: Clustering Data in Arbitrary Shapes

In Figure 2.11 (a), data set 3 (depicted in blue) comprises two groups of data with varying densities arranged in two arbitrary shapes: one group congregates around the origin of the 2D coordinate, and the other circular data group has a radius of approximately 5. We apply three clustering algorithms—GMM, DBSCAN, and K-means—to data set 3. For the GMM and K-means algorithms, we assume a Gaussian mixture with 2 components and set the optimal cluster number to 2, respectively. For the DBSCAN algorithm, we use the same parameter setting as in examples 1 and 2. The clustering results corresponding to the three methods are shown in the remaining panels of Figure 2.11: panel (b) for the GMM algorithm, panel (c) for the DBSCAN algorithm, and panel (d) for the K-means algorithm. Only the DBSCAN algorithm successfully identifies the data clusters correctly. Both the GMM and K-means algorithms fail in the clustering task. For the GMM algorithm, it assumes the data is related to a range of distribution models. When the data is gathered in different arbitrary shapes, it is challenging to use a mixture of Gaussian distribution models to fit it. Therefore, the GMM algorithm is unsuitable for these scenarios. The K-means algorithm works well on convex data sets, assuming clusters have spherical shapes. In this case, the circular cluster is not a convex data set; therefore, the K-means algorithm fails to provide the expected result.

Summary

In clustering analysis, the K-means algorithm demonstrates efficiency and simplicity, making it suitable for large datasets with well-defined spherical clusters. However, it assumes equally sized and convexed clusters, and its results are sensitive to initialization. The DBSCAN offers flexibility in identifying clusters with arbitrary shapes and adapts to varying densities, handling noise robustly. Nevertheless, the clustering result is sensitive to parameter choices. In other words, data points may be incorrectly assigned to the noise cluster with inappropriate parameter choice. The mixture model allows the capture of clusters with various probability distributions but is sensitive to initialization and assumes data follows specific probability distributions. For instance, GMM only assumes clusters within data follow Gaussian distribution. In practical applications, K-means with optimal cluster numbers performed well on datasets with convexed clusters, DBSCAN with suitable parameters excelled in handling noise and arbitrary shapes, and GMM faced challenges on data sets that are densely collected or non-Gaussian data. In summary, the selection of

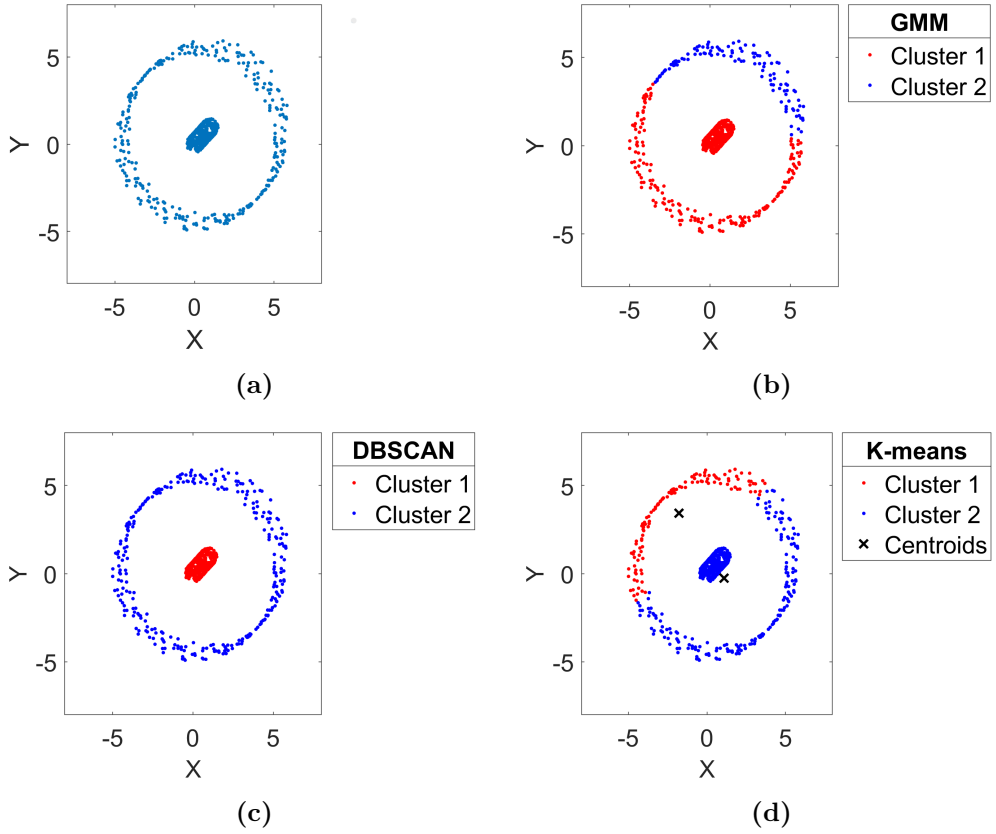


Figure 2.11: *(a) Data set 3 (blue); (b) the clustering result of the GMM algorithm; (c) the clustering result of the DBSCAN algorithm; (d) the clustering result of the K-means algorithm;*

the clustering method depends on the characteristics of the data set.

Chapter 3

Advanced Γ-method for small-scale vortex detection in the solar atmosphere

3.1 Introduction

In the solar atmosphere, vortical plasma motions are an important part of its dynamics. They may be responsible for magnetohydrodynamic (MHD) wave excitation (Schüessler, 1984; Attie et al., 2009; Fedun et al., 2011; Kitiashvili et al., 2011; Yadav et al., 2022) and the formation of plasma jets (Kitiashvili et al., 2013; Iijima and Yokoyama, 2017; Snow et al., 2018; Skirvin et al., 2022) and act as a channel for energy transfer to the solar corona (Wedemeyer-Böhm et al., 2012; Shelyag et al., 2012; Yadav et al., 2021). To understand the mechanism of vortex formation and its role in the plasma processes mentioned above, the precise automated identification of a vortex structure is essential (Tziotziou et al., 2023). However, the identification of vortices in solar numerical and observational data is still a challenge, as there is no universal definition for vortex motion (e.g., Günther and Theisel, 2018). Numerous identification approaches have been developed based on distinct definitions of what constitutes a vortex. Initially, investigations of vortical structures in the solar atmosphere were based on visual inspection of the velocity field (Brandt et al., 1988; Nordlund, 1985; Attie et al., 2009; Bonet et al., 2010) or detected motion of magnetic elements (Bonet et al., 2008; Balmaceda et al., 2010). Automated methodologies for vortex identification in solar physics were only introduced in the last decade by Moll et al. (2011) when they detected photospheric vortices using the swirling strength (Zhou et al., 1999), which was also used in posterior

studies by Yadav et al. (2020) and Canivete Cuissa and Steiner (2020). For a velocity field $\boldsymbol{U}$, the swirling strength is based on the velocity field tensor, $D = \nabla \boldsymbol{U}$, and it determines the vortex as a region where the eigenvalues of D are complex. Kato and Wedemeyer (2017) introduced a similar criterion, the vorticity strength, where a vortex region is identified as a local area where the eigenvalues of D are imaginary. The problem with such methods is that they provide a swirling region of the velocity field without giving proper information on the vortex boundary and centers. Thereby, the analysis of plasma properties and dynamics across the vortex using such techniques is not precise.

The first study on identification of a well-defined boundary and center for vortical structures in the solar photosphere was presented by Giagkiozis et al. (2018), who analysed a quiet Sun region by applying the Γ method (Graftieaux et al., 2001), which is based on local analysis of velocity vector field, to CRisp Imaging SpectroPolarimeter (CRISP, Scharmer, 2006; Scharmer et al., 2008b). Other identification methods that provide the vortex center and boundary are the Lagrangian averaged vorticity deviation (LAVD) and its Eulerian version, Instantaneous Vorticity Deviation (IVD), both defined by Haller et al. (2016). Those methods are based on vorticity and have been used to analyse vortices from observations (Silva et al., 2018; Chian et al., 2019, 2020) and simulations (Silva et al., 2020, 2021; Aljohani et al., 2022). Silva et al. (2018) compared the performance of three vortex identification methods. It was found that LAVD and vorticity strength are prone to false identifications without an additional requirement, whereas the Γ method may not detect or provide false detection of some vortex structures. Silva et al. (2018) addressed these issues by introducing a novel criterion 'd' when applying the LAVD and vorticity strength. This criterion can also be combined with other detection methods to address the false detection in shear flow.

All the vortex methods briefly described above have their advantages, depending on the goal of the analysis to be performed. Günther and Theisel (2018) have summarised existing methods according to the reference-frame invariance. Noninvariant detection techniques, such as visual inspection of the velocity field, refer to those methods that are suitable to the situation only when investigating the steady flow or the flow appears to be steady when choosing the right reference frame. Besides those noninvariant techniques, the remaining vortex detection methods are invariant under specific types of reference-frame motion. The vortex identification technique vorticity/swirling strength and Γ method are Galilean-invariant. Objective methods like LAVD or IVD are invariant when the reference frame is undergoing constant rotation and translation motion (see e.g., Günther and Theisel, 2018). Another approach, that is, the morphological method, which is applicable to the detec-

tion of chromospheric vortical motions was recently proposed by (Dakanalis et al., 2021, 2022).

Applying different vortex detection methods to solar data, both observational and simulation, has provided important information on vortex properties. By visual inspection of magnetic bright points from G-band images obtained from the Swedish Solar Telescope, Bonet et al. (2008) found several plasma vortices with sizes ranging from 0.5 to 2 Mm and lifetimes between 5 and 15 minutes in regions near the downdrafts of intergranular lanes. However, using magnetic bright points to identify vortices is not precise, as it can also overestimate the size of vortical structures and may interpret two close-by consecutive vortices as the same structure. The study by Giagkiozis et al. (2018) found an average lifetime of 17 s for intensity vortices detected by Γ method, and their statistical analysis also indicated the vortices have a diameter of around 568 km. In MURaM simulations, vortex identification with IVD indicated a longer average lifetime, around 84.7 s, and a diameter approximately seven times smaller. At supergranular scales, vortices have been detected by visual inspection (Attie et al., 2009), cork tracking (Requerey et al., 2018) and LAVD (Chian et al., 2019, 2020). While visual inspection provided sizes around 15 to 20 Mm, the LAVD identified vortices with a smaller range, 5 to 10 Mm. In all the supergranular analyses, there are persistent vortical structures that last for a couple of hours.

In this Chapter, we introduce the advanced Γ method (AGM), which is based on the original Γ method proposed by Graftieaux et al. (2001). By comparing the Γ method and the AGM, we show that our technique reduces the influence of threshold choices. The AGM is applied to detect solar vortex in a simulated solar surface obtained by 'StellarBox' (3D radiative MHD code). We investigate the lifetimes, sizes and dynamics of the velocity field components of the detected vortices. The new detection method is also applied to CRISP observational data and compared with the numerical data results.

This Chapter is structured as follows. First, in Section 3.2, we describe the AGM technique and the algorithm construction. Then, two novel algorithms dedicated to extracting the proper boundary of vortex structure are introduced in Section 3.3, that is, the circular sector algorithm (CSA) and the whole-circular sector algorithm (WCSA). In Section 3.4, we provide detection results and statistical analysis and compare the performance of the AGM with the Γ method. Last, the discussion and conclusions are given in Section 3.5.

3.2 Vortex Identification

3.2.1 Description of Simulation Data

For the numerical modelling we use a 3D radiative MHD code, StellarBox (see, e.g., Kitiashvili et al., 2012a; Wray et al., 2015), developed for realistic simulations of the upper convective zone and lower atmosphere. The size of the simulation area is 6.4 × 6.4 Mm, and it initially had an imposed uniform vertical magnetic field around 10 G. The simulation has a cadence of 5 s and the analysed data have a total duration of 625 s. For our investigations, we used the horizontal velocity field from the simulated solar surface with a resolution of 12.5 km in x- and y-directions. An example of a simulated horizontal velocity field snapshot is shown in Figure. 3.1.

3.2.2 Vortex Identification by Γ method

For vortex identification, that is, its center and boundary, Graftieaux et al. (2001) introduced two dimensionless scalar functions, Γ_1 and Γ_2, respectively. It was shown that these two functions are able to detect large-scale vortices in turbulent flows.

The discrete version of Γ_1 at a given point P (within the numerical domain) can be represented as

$$\Gamma_1(P) = \frac{1}{N} \sum_S \frac{(\boldsymbol{PM} \times \boldsymbol{U}_M) \cdot \boldsymbol{z}}{\| \boldsymbol{PM} \| \cdot \| \boldsymbol{U}_M \|} = \frac{1}{N} \sin(\theta_M), \tag{3.1}$$

where N is the number of discretely sampled velocity field points of a 2D area S centered on P. Point M lies within S, and $\boldsymbol{PM}$ denotes the displacement vector from point P to M. The variable $\boldsymbol{U}_M$ denotes the velocity vectors at point M, and $\boldsymbol{z}$ is a unit vector normal to the horizontal plane. θ_M represents the angle between $\boldsymbol{U}_M$ and $\boldsymbol{PM}$. The $|| \cdot ||$ notation represents magnitude.

The Γ_1 function defined in Equation (3.1) provides a scalar field that quantifies the topology of the flow in the neighboring region of P. At the vortex center, $|\Gamma_1|$, the cross-product value $\boldsymbol{PM} \times \boldsymbol{U}_M$ has a maximum, leading to the highest value of $|\Gamma_1|$. The sign of Γ_1 value at the vortex center indicates the direction of the vortex rotation; that is, negative indicates that the rotation is clockwise rotation and positive means it is counterclockwise. The vortex center will have $|\Gamma_1|$ values > 0.9 and $= 1.0$ in the ideal case when vortex is axisymmetric. In real turbulent flows, the vortex's appearance contains a broad range of irregular shapes. Thereby, we set a threshold for a vortex cen-

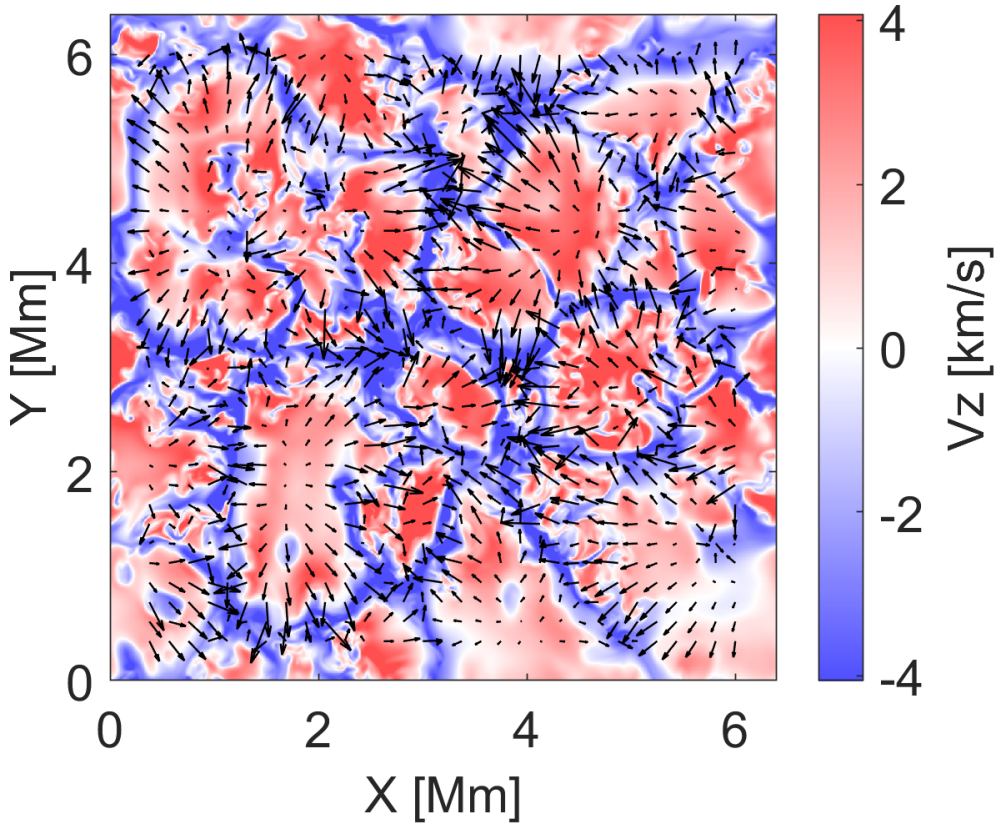


Figure 3.1: *A snapshot of the magnetoconvection simulation at $t = 5$ s, displaying the simulated solar surface coloured by the vertical velocity field. The black arrows show the direction of the horizontal velocity field.*

ter $|\Gamma_1| \geq 0.75$. More precisely, a point P will be considered as vortex center if the $|\Gamma_1|$ value is above 0.75 and if it is a local maxima within the region S.

The function Γ_2 considers the contribution of the average local velocity field around the vortex center. The definition of the discrete version of Γ_2 is

$$\Gamma_2(P) = \frac{1}{N} \sum_S \frac{[\boldsymbol{PM} \times (\boldsymbol{U}_M - \overline{\boldsymbol{U}}_P)] \cdot \boldsymbol{z}}{\| \boldsymbol{PM} \| \cdot \| \boldsymbol{U}_M - \overline{\boldsymbol{U}}_P \|}. \tag{3.2}$$

Here, $\overline{\boldsymbol{U}}_P$ is the local average velocity around the point P, that is,

$$\overline{\boldsymbol{U}}_P = \frac{1}{N} \sum_S \boldsymbol{U}. \tag{3.3}$$

It can be shown that in the case of incompressible flow, for a very small 2D vortex area ($S \to 0$), function Γ_2 depends on the antisymmetrical and symmetrical parts of the velocity gradient $\nabla \mathbf{U}$ tensor at point P. Locally, the flow is dominated by rotation if $|\Omega/\mu| > 1$, which implies $|\Gamma_2| > 2/\pi$. Here, Ω (the rotation rate) and μ (the eigenvalue) correspond to the antisymmetrical part and symmetrical part of the $\nabla \mathbf{U}$, respectively. It is important to note that such classification has not yet been defined in the case of a finite region S (Graftieaux et al., 2001).

3.2.3 Vortex Identification with the Convolution Γ Method

The discrete Γ_1 function for the 2D convolution version of the Γ_1 method (hereafter convolution version of Γ_1, CGM) is proposed by Zigunov et al. (2020):

$$\begin{aligned} \Gamma_1(P) &= \frac{1}{N} \Big[\sum_S \frac{PM_x}{\| \boldsymbol{PM} \|} \frac{U_{My}}{\| \boldsymbol{U}_M \|} - \sum_S \frac{PM_y}{\| \boldsymbol{PM} \|} \frac{U_{Mx}}{\| \boldsymbol{U}_M \|} \Big] \\ &= \frac{1}{N} \Big[\frac{PM_x}{\| \boldsymbol{PM} \|} \circledast \frac{U_{My}}{\| \boldsymbol{U}_M \|} - \frac{PM_y}{\| \boldsymbol{PM} \|} \circledast \frac{U_{Mx}}{\| \boldsymbol{U}_M \|} \Big], \end{aligned} \tag{3.4}$$

where U_{Mx} and U_{My} correspond to the x- and y- components of the velocity vector starting from the point M within the rectangular region S. Analogously, PM_x and PM_y are the corresponding components of displacement vectors from the point P to a point M. The potential vortex center P is regarded as the origin of this coordinate system defined inside S. The variable N is the number of discrete grid points around the point P within the S region. The '$\circledast$' is the convolution operator. We can write the first terms in the convolutions in a

trigonometric representation:

$$\frac{PM_x}{\parallel \boldsymbol{PM} \parallel} = \cos\left(\theta_{PM}\right), \quad \frac{PM_y}{\parallel \boldsymbol{PM} \parallel} = \sin\left(\theta_{PM}\right). \tag{3.5}$$

Thus, the two convolutions in the Equation (3.4) can be represented as multiplications of convolution kernels by the sines and cosines of the angles between the vectors $\boldsymbol{U}_M$ and the x axis, i.e., $sin(\theta_{U_M})$ and $cos(\theta_{U_M})$, respectively (see Zigunov et al., 2020, for more details). Therefore, this equation can be rewritten as

$$\Gamma_1(P) = \frac{1}{N}\left[\sum_S \cos\left(\theta_{PM}\right)\sin\left(\theta_{U_M}\right) - \sum_S \sin\left(\theta_{PM}\right)\cos\left(\theta_{U_M}\right)\right]. \tag{3.6}$$

In other words, the form presented in Equation (3.4) is still an analysis on the geometry of the velocity field in the region S. In fact, Equation (3.6) can be directly recovered from Equation (3.1) by substituting $\theta_M = \pi - \theta_{U_M} + \theta_{PM}$. The practical advantage of the Equation (3.4) is the use of convolution operations, which enable the use of small convolution kernels within the region S. It is also important to mention that a small convolution kernel helps to identify the vortices that are close to each other, which is difficult in the case of direct use of Γ_1 function (see Equation (3.1)).

The full displacement matrix for the x- components of the displacement vectors $\boldsymbol{PM}$ for a small area—for example, 7×7 points M—within the region S can be represented as

$$PM_x = \begin{bmatrix} -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix}. \tag{3.7}$$

Each element of this matrix corresponds to the normalized distance between two discrete points P and each M inside the convolution kernel. The displacement matrix PM_y can be represented as $PM_x{}^{\intercal}$. The general form of PM_x, PM_y, and $\parallel \boldsymbol{PM} \parallel$ are shown in Appendix A. The size of these matrices is $(2n+1) \times (2n+1)$, where $n \in \mathbb{Z}^+$.

Based on the Equation (3.2), and similarly to the Equation (3.4), the convolution version of Γ_2 function can be introduced as the following:

$$\Gamma_2(P) = \frac{1}{N}\left[\frac{PM_x}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{My} - \overline{U}_{py}}{\parallel \boldsymbol{U}_M - \overline{\boldsymbol{U}}_p \parallel} - \frac{PM_y}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{Mx} - \overline{U}_{px}}{\parallel \boldsymbol{U}_M - \overline{\boldsymbol{U}}_p \parallel}\right]. \tag{3.8}$$

Here, $\overline{U}_{px}$ and $\overline{U}_{py}$ are the x- and y- components of the local average velocity ($\overline{\boldsymbol{U}}_P$) around point P, respectively.

3.2.4 Advanced Γ Method

Based on the convolution Γ method, we derived the advanced Γ method (AGM) in this Section. In Appendix A, we present the general form of the full displacement matrices for the x- and y-components of the displacement vectors $\boldsymbol{PM}$ (PM_x and PM_y). From their structure, we notice that they are rank 1 matrices. This implies that these matrices, for example, PM_x and PM_y, are separable. Therefore, they can be represented as

$$PM_x = \boldsymbol{w}_1 \otimes \boldsymbol{w}_2 \tag{3.9}$$

and

$$PM_y = (\boldsymbol{w}_1 \otimes \boldsymbol{w}_2)^\intercal = \boldsymbol{w}_2^\intercal \otimes \boldsymbol{w}_1^\intercal, \tag{3.10}$$

where $\boldsymbol{w}_1$ is a column unit vector and $\boldsymbol{w}_2$ is a row vector. The general form of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ of size $2n+1$ can be represented as

$$\boldsymbol{w}_1 = \begin{bmatrix} 1 & \cdots & 1 & 1 & 1 & 1 & 1 & \cdots & 1 \end{bmatrix}^\intercal, \tag{3.11}$$

$$\boldsymbol{w}_2 = \begin{bmatrix} -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \end{bmatrix}. \tag{3.12}$$

By substituting Equations (3.9) and (3.10) into Equation (3.4), $\Gamma_1(P)$ can be presented as

$$\Gamma_1(P) = \frac{1}{N}\left[\frac{\boldsymbol{w}_1 \otimes \boldsymbol{w}_2}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{My}}{\parallel \boldsymbol{U}_M \parallel} - \frac{\boldsymbol{w}_2^\intercal \otimes \boldsymbol{w}_1^\intercal}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{Mx}}{\parallel \boldsymbol{U}_M \parallel}\right]. \tag{3.13}$$

Then, by taking into account that the outer product $\boldsymbol{w}_1 \otimes \boldsymbol{w}_2$ (for the 2D case) can be replaced by the convolution of these two vectors (see Appendix B), we arrive at

$$\Gamma_1(P) = \frac{1}{N}\left[\frac{\boldsymbol{w}_1 \circledast \boldsymbol{w}_2}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{My}}{\parallel \boldsymbol{U}_M \parallel} - \frac{\boldsymbol{w}_2^\intercal \circledast \boldsymbol{w}_1^\intercal}{\parallel \boldsymbol{PM} \parallel} \circledast \frac{U_{Mx}}{\parallel \boldsymbol{U}_M \parallel}\right]. \tag{3.14}$$

Similarly to the original Γ method, here we consider a small region S, such that the values of $\parallel \boldsymbol{PM} \parallel$ (at all the points in S) present only minor variations compared with its mean value. Therefore, $\parallel \boldsymbol{PM} \parallel$ at every point can be approximated by a given constant value Z_1; that is, Z_1 replaces the components

of the matrix $\| \boldsymbol{PM} \|$ in the region S. This approximation simplifies the calculations, which is needed for the better performance of computation on larger data sets. According to the Equation (A.3), one may consider that there are some difficulties with satisfying this condition, but later on, in Section 3.2.5, it is shown that this approximation provides both valid and accurate results.

By applying the commutative and associative properties of the convolution operation, Equation (3.14) can be represented as

$$\Gamma_1(P, S_1) = \frac{1}{Z_1 N_1} \left[\left(\boldsymbol{w}_1 \circledast \frac{U_{My}}{\| \boldsymbol{U}_M \|} \right) \circledast \boldsymbol{w}_2 - \left(\boldsymbol{w}_2^\intercal \circledast \frac{U_{Mx}}{\| \boldsymbol{U}_M \|} \right) \circledast \boldsymbol{w}_1^\intercal \right]. \tag{3.15}$$

Here, we define S_1 as a subset of S (e.g., $S_1 \subset S$) and S_1 is a small enough region to satisfy $\| \boldsymbol{PM} \| \approx Z_1$; N_1 is the number of grid points within the region S_1. The use of Equation (3.15) has several advantages. First, as demonstrated in Section.3.2.5, there is no need to compute $\| \boldsymbol{PM} \|$. Second, due to the separation of the convolution kernel, the AGM for the calculation of Γ_1 is computationally faster than CGM (see Equation (3.4)) by $ks/2$ times, where ks is the kernel size (see Appendix C for more details). The ks represents the dimension of the displacement matrix, for example, PM_x, PM_y. Analogously, Γ_2 can be represented in the following form:

$$\Gamma_2(P, S_2) = \frac{1}{Z_2 N_2} \left[\left(\boldsymbol{w}_1 \circledast \frac{U_{My} - \overline{U}_{py}}{\| \boldsymbol{U}_M - \overline{\boldsymbol{U}}_p \|} \right) \circledast \boldsymbol{w}_2 - \left(\boldsymbol{w}_2^\intercal \circledast \frac{U_{Mx} - \overline{U}_{px}}{\| \boldsymbol{U}_M - \overline{\boldsymbol{U}}_p \|} \right) \circledast \boldsymbol{w}_1^\intercal \right], \tag{3.16}$$

where

$$\overline{U}_P = \frac{1}{N_2} \sum_{S_2} U$$

is the local average velocity within region S_2 surrounding a point P, N_2 is the number of grid points in the region S_2, Z_2 is the average of $\| \boldsymbol{PM} \|$ inside the region S_P, and $\overline{U}_{px}$ and $\overline{U}_{py}$ are the x- and y-components of the $\overline{U}_p$ defined within region S_2, respectively. The Γ method and the CGM use the same region S for calculating Γ_1 and Γ_2. However, the velocity values at the grid points which are close to the corners of the region S do not significantly influence the true vortex boundary, and, therefore, use of those points in the average calculation of $\overline{U}_p$ may lead to an unwanted contribution to the Γ_2

calculation. As a result, we apply a different geometry to the subset of the region S—that is, S_2—when calculating Γ_2 in the AGM. For a point P in the domain, the region S_2 includes the vortex center, its boundary, and neighboring regions that truly contribute to the actual value of the local average velocity.

The size of region S_1 is defined in such a way that it encompasses only one vortex at each time. In other words, a set of vortices originally found in region S will be separated into individual vortices where each vortical structure is within a region S_1. To achieve this, S_1 is set to have the same size as its convolution kernel. Figure 3.2 presents the detection results obtained by the AGM, using two different single kernel sizes—that is, $ks = 3$ (blue contours) and $ks = 7$ (orange contours)—in a region where multivortices are close to each other. The AGM with $ks = 3$ detected five vortex structures, and the AGM with $ks = 7$ (orange) only detected two of them. It is also clear that the center's locations (indicated by the cross signs) change slightly depending on the kernel size. This is due to the fact that a larger S_1 area, such as the one for $ks = 7$, leads to the contribution of irrelevant flow areas when computing Γ functions, thereby introducing errors.

The boundaries shown in Fig.3.2 are obtained by applying the advanced Γ_2 method to a different region S_2. The first step in defining S_2 is to consider $S_2 = S_1$; that is, S_2 is rectangular and has the same size as the kernel used for the calculation of Γ_1. In this way, we obtain an initial vortex boundary, which is defined by Γ_2. Then, we analysed the velocity points surrounding point P just outside the initial vortex boundary. To check whether a point M outside the initial boundary meets the Γ_2 criteria, we recalculate Γ_2 with the changed S_2 region to encompass the M-point. In other words, we take into account the M point contribution to the local $\overline{U}_P$. This procedure is repeated until the points outside the new boundary do not obey the original Γ_2 criteria. The slightly different shapes for the two vortices identified by the use of $ks = 3$ and $ks = 7$ are related to their identified center's location and corresponding customized S_2 (see Section 3.3 for more details).

3.2.5 The Estimation of the Z Value

In order to compare the detection performance between the AGM and the CGM, both were applied to the same magnetoconvection simulation data set (Fig. 3.1). The corresponding values of the Γ_1 and Γ_2 fields calculated for the horizontal velocity field are shown in Figure 3.3. To analyse the influence of Z_1 and Z_2, first, we calculate the AGM using $Z_1 = Z_2 = 1$, and we call the map of those Γ values the 'unnormalised' AGM (see Figure 3.3 (a) and (c), respectively). For comparison, panels (b) and (d) of the same figure display the

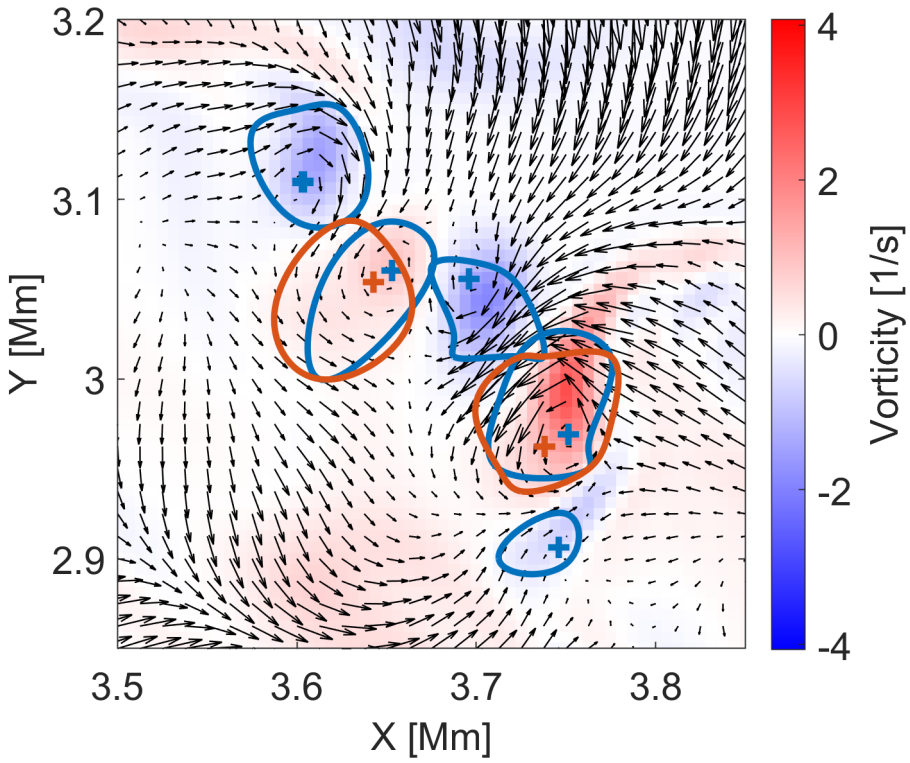


Figure 3.2: *Vortex detection by the AGM using different kernel sizes (ks). The blue contours (cross) indicate the vortex boundaries (centers) obtained by applying the AGM with* $ks = 3$*. The orange contours (cross) indicate the vortex boundaries (centers) obtained by applying the AGM with* $ks = 7$*.*

results obtained by the CGM. Both the AGM and CGM were applied using the same kernel size $ks = 11$; that is, each separable convolution kernel has the same size of 11×11. Both the AGM and CGM provide the same spatial distribution of Γ_1 and Γ_2 functions, differing only in the absolute values. The CGM provides the same range as the classical Γ method, which is expected since one can easily recover the original Γ method from CGM functions. The separation of kernels introduced by the AGM changes the maximum and minimum range of its functions if the correct values of Z_1 and Z_2 are not applied, but it does not change the Γ functions' spatial distribution.

Figure 3.4 presents distributions of ratios (i.e., C_1 and C_2) between the unnormalised AGM and CGM for Γ_1 and Γ_2. The mean values of the C_1 and C_2 distributions are 4.625 and 4.5, respectively. In other words, the Γ values detected by the unnormalised AGM are on average $\mu(C_i)$ times larger than the corresponding Γ values detected by the CGM. Therefore, in general, to obtain the AGM within the same range as the CGM for $ks = 11$, one should apply the values obtained for $\mu(C_1)$ and $\mu(C_2)$ to Z_1, Z_2, respectively. The variance of the $C_{1,2}$ distributions reflects how good the approximation provided by the $Z_{1,2}$ values to the $\| \boldsymbol{PM} \|$ values; that is, if $Var(C_{1,2}) = 0$, $Z_{1,2}$ is identical to $\| \boldsymbol{PM} \|$. In Appendix D, we present the same analysis for the case $ks = 33$.

Figure 3.5 presents a series of mean ratios of C_1 (blue squares) and C_2 (orange circles) versus the corresponding size of the convolution kernel (ks). Both dependencies show near-linear behaviour which indicates that the mean ratio increases with ks.

Based on the analysis performed, we can conclude that Z_1 and Z_2 can be chosen to be equal to the mean values of C_1 and C_2. In other words, such assumptions will provide the AGM functions within the expected range of -1, 1. Moreover, in cases where the difference between Z_1 and Z_2 is small enough, one can set $Z_1 = Z_2$, which can simplify the algorithm further. Our results also indicate that the selection of kernel size will influence the detection efficiency (see Equation (C.1)) and accuracy. For instance, by comparing Figs. 3.3 and Figs. D.1, we have found that the small-scale processes are better captured in Fig. 3.3.

The vortex detection by the AGM can be optimized by using an adaptive version of the AGM, which is based on a sequence of different kernel sizes, for example, 3, 5, 7, 9, and 11, and so on. Ideally, the kernel is located at the center of the potential vortical structure. For small odd-size kernels—that is, $ks = 3$—this may lead to nondetection (a failure to detect), and the use of a sequence of growing kernel sizes increases the chance of detection. This is illustrated in Figure 3.6(a), where the AGM with $ks = 5$ detected the vortex center (blue cross) outside the grid points. The same region would have a

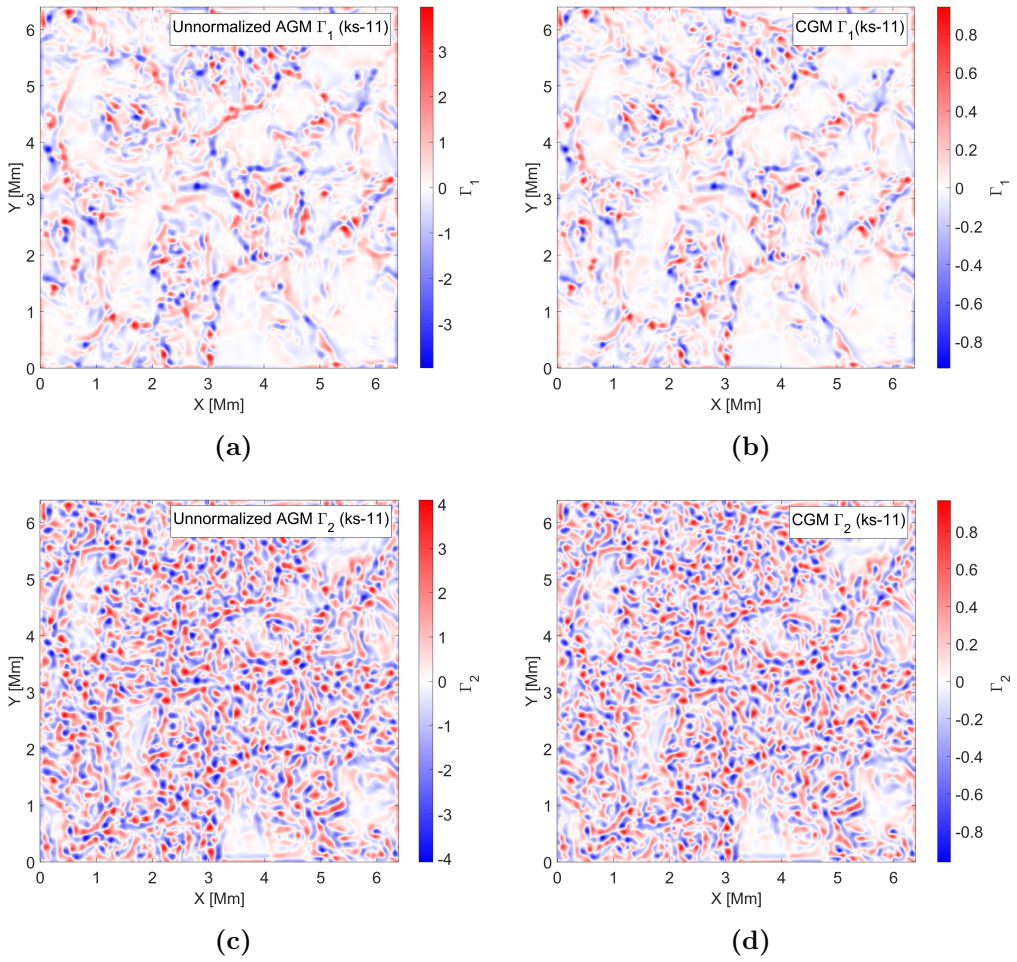


Figure 3.3: *The distribution of the* Γ *functions values obtained by the AGM ((a), (c)) and CGM ((b), (d)) which were applied to the photospheric horizontal velocity field, as shown in Figure 3.1. For both cases, a convolution kernel size* $ks = 11$ *was applied. As here the values of* Z_1 *and* Z_2 *for the AGM were equal to 1, the AGM was unnormalised.*

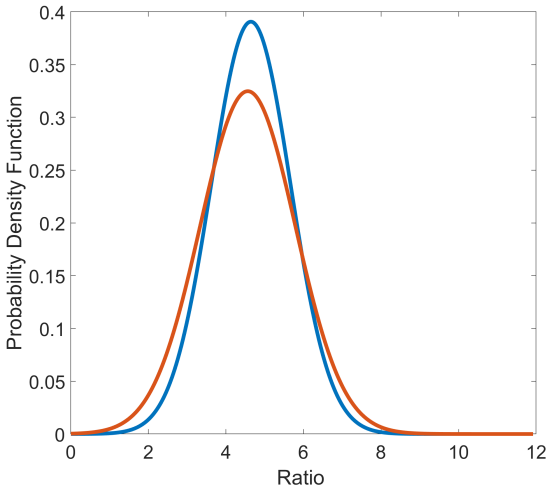


Figure 3.4: *The distribution of the ratios between the unnormalised AGM (Figure 3.3 panels(a) and (c)) and CGM (Figure 3.3 panels(b) and (d)) for* Γ_1 *(*C_1 *is shown in blue) and* Γ_2 *(*C_2 *is shown in red). For both cases, the convolution kernel is equal to 11.*

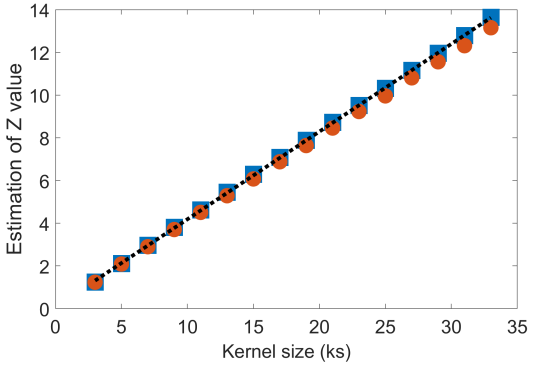


Figure 3.5: *The relationship between the estimated Z values and convolution kernel size (ks). Blue squares indicate the mean ratios of* C_1 *and orange circles correspond to the mean ratios of* C_2*. The black dashed line is the linear regression between them.*

Γ_1 value lower than the threshold when applying the AGM with $ks = 3$, as illustrated in Figure 3.6(b), where the corresponding Γ_1 values of the grid points surrounding the vortex center are indicated with black dash arrows. The blue and red dash squares show two kernel domains that can be applied when calculating the corresponding Γ_1 values. When the kernel size increases to $ks = 5$, which is depicted in Figure 3.6(c), the Γ_1 value is above the threshold and the vortex center is finally detected.

As discussed above, several kernel sizes may yield values of Γ_1 greater than the threshold; therefore, the result with the maximum value of Γ_1 will be selected to be the detection result. The maximum of the Γ_1 value indicates the most accurate location of the vortex center, as illustrated by Figure 3.2 and 3.6. As the vortex boundary is influenced by the kernel size (see Figure 3.2), the final area of the vortex is affected by the choice of the kernel. Therefore, some statistical features of the vortex, like size and center properties, are affected as one applies the most appropriate kernel size to describe a given vortex. For example, larger kernel sizes are better suited to describe vortices presenting greater areas and the opposite is found for small kernels. Therefore, varying the kernel size for each vortex, as done by the adaptive AGM, provides us with better identification and leads to the more accurate statistical results of the vortex parameters.

3.3 The Circular Sector Algorithm and the Whole-Circular Sector Algorithm

This Section introduces two boundary detection algorithms dedicated to calculating a more precise boundary for each detected vortex when applying with the AGM: the circular sector algorithm (CSA) and the whole-circular sector algorithm (WCSA).

The CSA algorithm is effectively illustrated in Figure 3.7. Initially, the AGM calculates the initial boundary of an example vortex (depicted in gray in panel (a)) by setting $S_1 = S_2$. The blue cross represents the detected vortex center, while the red cross marks a point on the initial boundary. The black dashed line denotes the current displacement vector for the red cross. Subsequently, the CSA systematically reevaluates all velocity points outside the initial boundary using the boundary criterion and a series of customized S_2. Panel (b) illustrates the outer neighboring points of the current red cross in yellow, with one highlighted in green to exemplify the corresponding customized S_2. In the CSA, the customized S_2 region comprises two subset parts within the entire velocity domain. The first part encompasses the velocity field

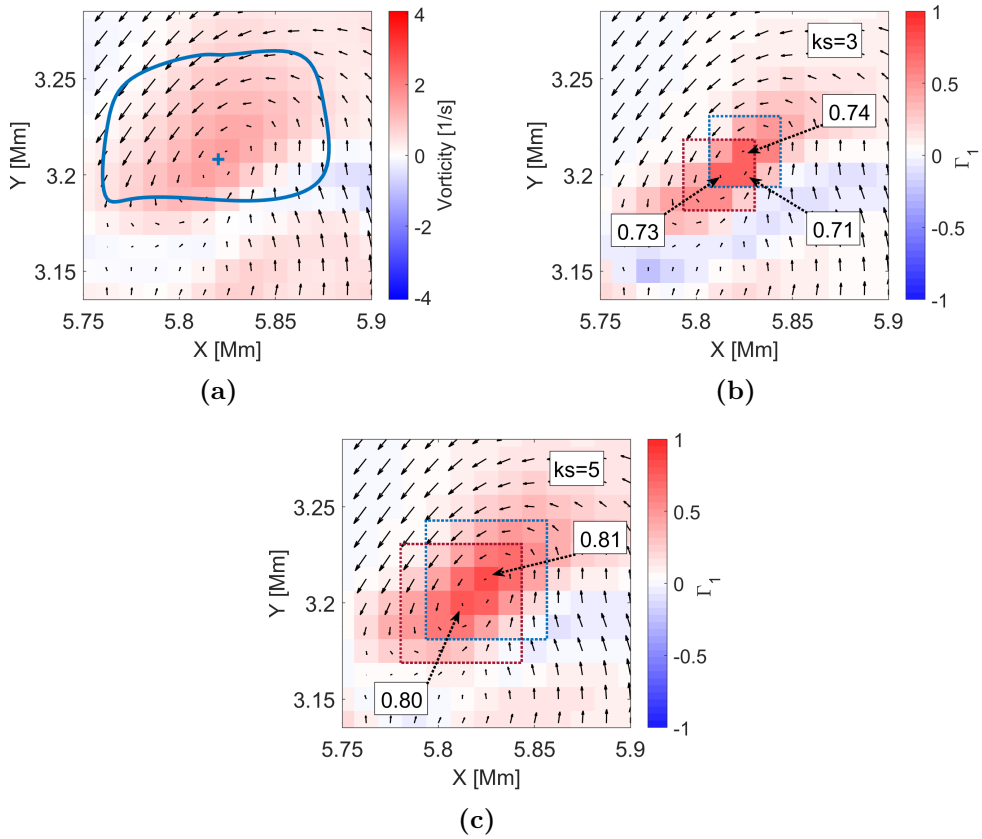


Figure 3.6: *(a) Vortex detected with the adaptive AGM. The vortex center and boundary are indicated with a blue cross and contour, respectively. (b) The* Γ_1 *map is detected with* $ks = 3$ *of panel (a). The blue and red squares indicate two* 3×3 *kernel domains that were used to calculate* Γ_1*. (c) The* Γ_1 *map is detected with* $ks = 5$ *of panel (a). The blue and red square indicate two* 5×5 *kernel domains that were used to calculate* Γ_1*.*

surrounding the vortex center (depicted by orange points in panel (c)). The second part relates to the velocity field near the current displacement vector (indicated by blue points in panel (d)). Notably, the first subset velocity field remains constant for all outer candidate velocity points, while the second part varies for different candidates. Panel (e) provides an example of the second part (blue) for a different candidate (green). After examining all nearest outer velocity points of the initial boundary, the CSA proposes a new vortex region whenever these points satisfy the boundary criteria. In the given example, the newly identified vortex region is enclosed in magenta in panel (f). Additionally, a green boundary is proposed based on the magenta one. The CSA iteratively explores and proposes new vortex regions until no outer velocity points meet the criterion.

The WCSA follows a procedure analogous to the CSA for calculating the boundary of a detected vortex. Unlike the CSA, which considers the velocity field surrounding the vortex center and a customized radial part of the velocity field (see Figure 3.7(c), (d), and (e)), the WCSA incorporates all velocity points enclosed by the current boundary. In Figure 3.8, the orange velocity points enclosed by the initial boundary are taken into account when calculating the Γ_2 value for the green point in the WCSA. This calculation involves considering the contribution of all velocity points to the entire vortex region.

In Figure 3.9, the AGM's final results using the CSA and the WCSA are represented by the green and dark red boundaries, respectively. An evident overestimation issue arises with the CSA, as indicated by the green boundary enclosing some weak vorticity regions. This issue stems from the CSA's assumption that the vortex is axisymmetric, where each radial sector velocity field of each S_2 is independently calculated. Consequently, this assumption neglects the constraints and contributions of the entire vortex region. The CSA performs well under axisymmetric conditions, where the vortex and its outer velocity field are axisymmetric. However, it may yield improper boundaries when these conditions are not met. The WCSA, which considers all velocity points within the vortex region, produces a more conservative boundary, preventing overestimation issues. Nevertheless, the WCSA tends to extract a circular or axisymmetric vortex boundary. To address this, we introduce the combination algorithms using both CSA and WCSA, resulting in the light blue boundary shown in Figure 3.9. This combination algorithm employs CSA to calculate the Γ_2 value, then computes the Γ_2 value using WCSA, comparing the results. If both Γ_2 values are close to each other and exceed the threshold, the algorithm continues exploring the outer region. However, if the CSA's Γ_2 value significantly exceeds that of the WCSA, the algorithm terminates the calculation. In Appendix E, we calibrate the algorithm using a vortex with a

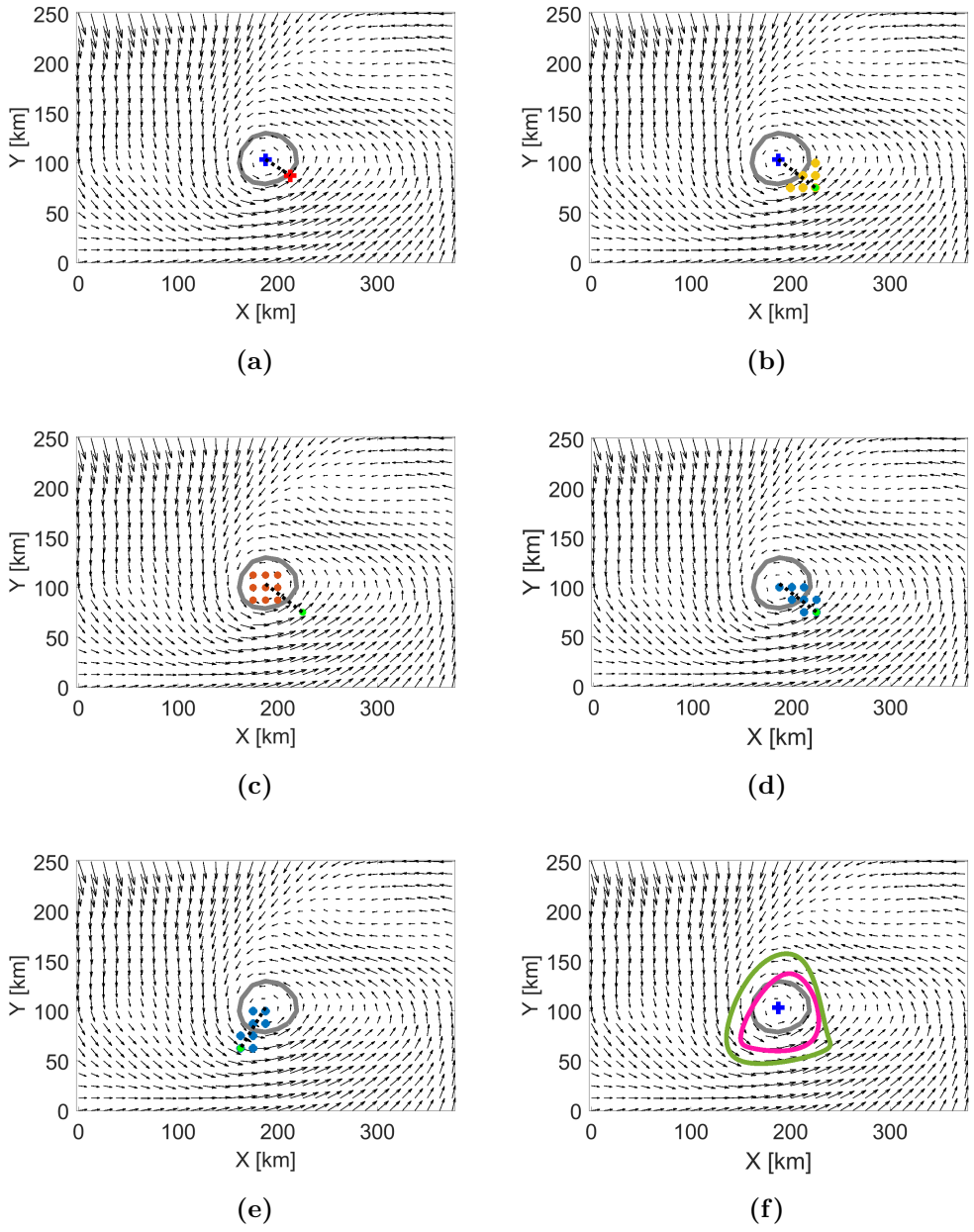


Figure 3.7: *Illustration of the CSA. (a) Initial boundary (gray) calculated by the AGM when setting $S_1 = S_2$. A blue cross indicates the vortex center, and a red cross denotes one point on the boundary. (b) Neighboring points (yellow and green) of the red cross. (c) Neighboring points of the vortex center (orange). (d) Velocity points between the vortex center and the green point. (e) Velocity points between the vortex center and another green point. (f) The magenta boundary is proposed based on the initial boundary, and the green boundary is proposed based on the magenta one.*

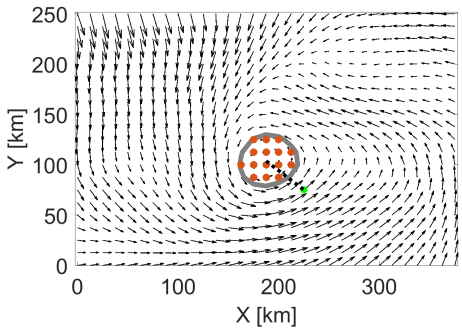


Figure 3.8: *Velocity points (orange) within the initial boundary (gray) of the same example shown in Figure 3.7.*

known boundary.

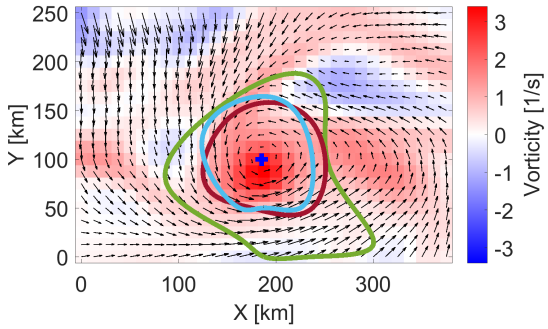


Figure 3.9: *The CSA extracts the green boundary, and the WCSA extracts the dark red boundary. The light blue boundary is extracted by the combination algorithm of the CSA and the WCSA.*

3.4 Vortex Detection Results and Statistical Analysis

In this Section, the vortex detection was performed with the use of the AGM method with a maximum of five different convolution kernel sizes. The candidate for a vortex center is located at the center of the kernel, and, therefore, it is surrounded by an even number of grid points in the x- and y-directions, which correspond to ks =3, 5, 7, 9, and 11. For convenience, we set $Z_1 = Z_2$ (see Section. 3.2.5). For different kernel sizes, the values of Z_1 and Z_2 were

set equal to the values of the linear regression function (see Figure 3.5). For example, for $ks = 11$, these values correspond to $Z_1 = Z_2 = 4.597$. The threshold of $|\Gamma_1|$ is set equal to 0.75. As in the classical Γ method, a candidate for a vortex center will be considered a true center if $|\Gamma_1| > 0.75$ and if $|\Gamma_1|$ value at that point is larger than in other surrounding grid points. Similarly, for Γ_2, we apply the threshold of $2/\pi$.

3.4.1 Detection Results

The AGM algorithm was applied to the whole simulated photospheric velocity field (see Figure 3.1), but here we will focus on the detection results of a single vortex. Figure 3.10 shows the identification obtained by applying the AGM with $ks = 7$. In all the panels, the vortex boundary is indicated by a red curve, and we see that the AGM provides a boundary that aligns well with the velocity streamlines (shown in blue). Based on the velocity field arrows (shown in black), it is clear that the vortex center, represented by a red cross, is also well detected by the AGM. The radial velocity component changes sign with the flow, as displayed in panel (a), and the tangential velocity component decreases from the boundary toward the center (see panel (b)). As expected, the vortex covers the area with the maximum vorticity, panel (c), and the high value of the angular velocity component, panel (d). In particular, this vortex is located in a downflow region, panel (e), and in the region of the low compressible plasma, panel (f).

The averaged (over angular directions) radial and tangential velocity profiles as a function of time and radius for the vortex shown in Figure 3.10 are shown in Figure 3.11(a) and (b), respectively. The vertical axis represents the distance from the vortex center to the mean maximum vortex radius (green dashed line), and the horizontal axis indicates the lifetime scale of the vortex. The radial and tangential velocity components show a maximum value just after the vortex appearance, and then they tend to decrease as the vortical structure disappears. The value of the tangential velocity component is approximately five times larger than the corresponding radial velocity component. Furthermore, the tangential component has larger values closer to the boundary; that is, plasma decelerates as it approaches the center. This result is in good agreement with previous results (Silva et al. (2020)) The change in sign of the radial velocity component close to the average boundary (see Figure 3.11(a)) indicates that the AGM detection might have slightly overestimated the vortex boundary. The AGM identifies vortices at each time frame, but it does not automatically provide the time evolution of a detected vortical structure. To analyse the vortex behaviour during its lifetime, it is necessary

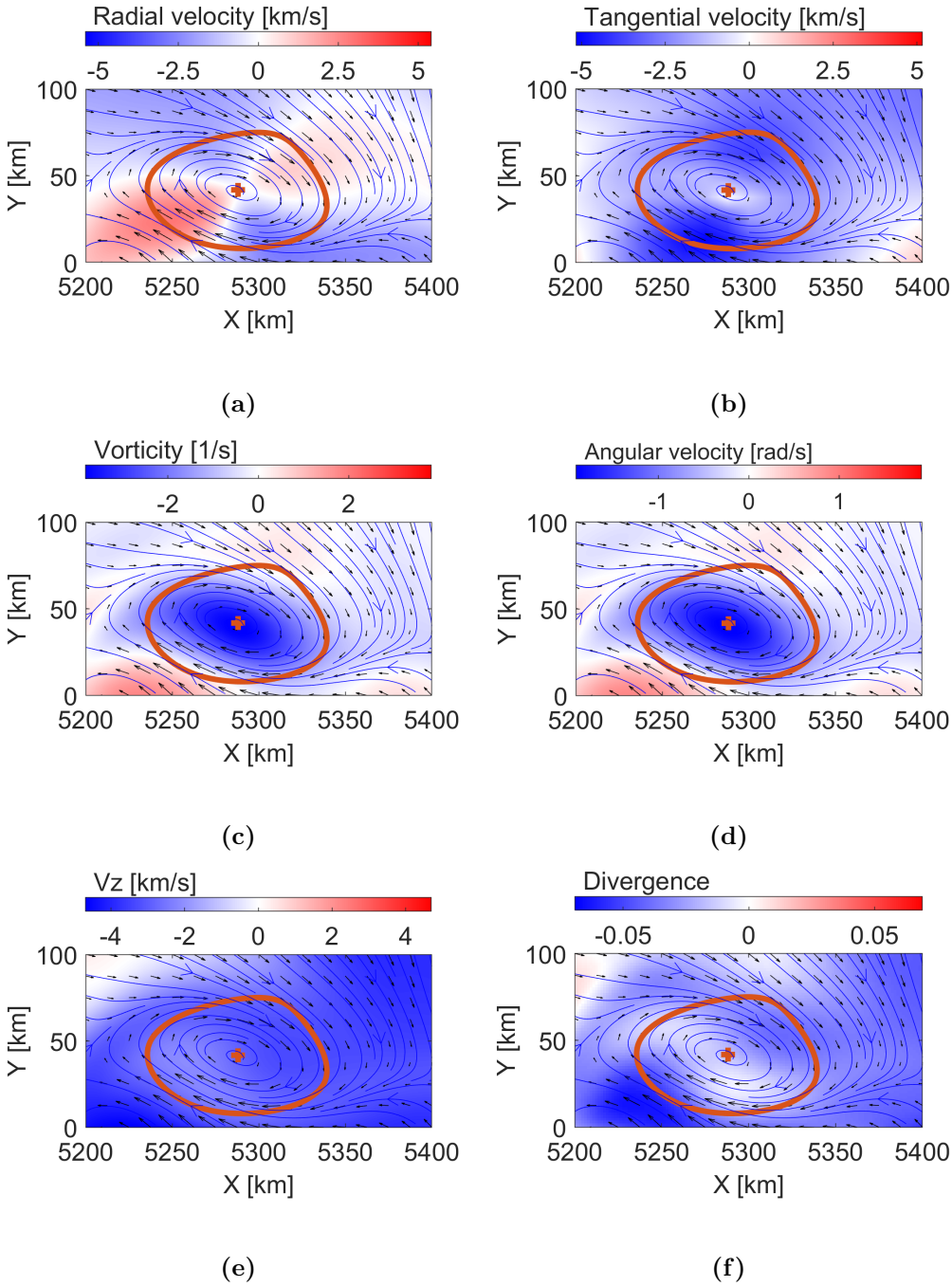
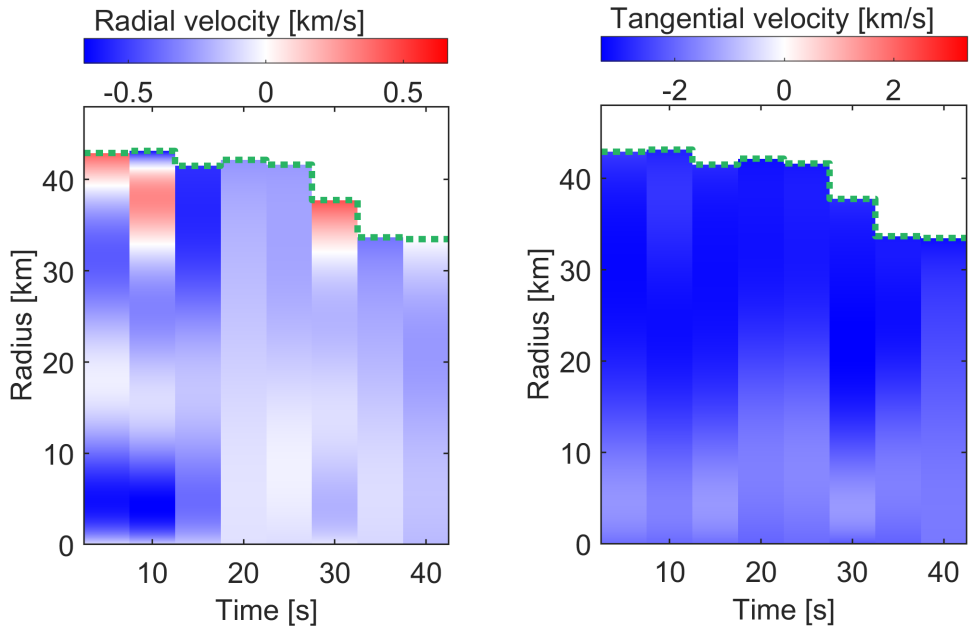


Figure 3.10: *The vortex number 1 as it was detected by the AGM using ks = 7 at the moment $t = t_0$. The orange plus sign indicates the vortex center, and the orange contour depicts the corresponding boundary. The direction of the horizontal velocity field is indicated by the streamlines and the arrows in blue. The background is coloured by (a) the radial velocity component, (b) the tangential velocity component, (c) the vorticity, (d) the angular velocity component, (e) the vertical component of the velocity field, (f) the divergence of the velocity.*



(a) **(b)**

Figure 3.11: *The radial and tangential velocity components (averaged over angular directions) as a function of time and radius are shown in panels (a) and (b), correspondingly. The vertical axis indicates the distance from the vortex center to the mean vortex radius calculated at the corresponding time (green dashed line). The horizontal axis indicates the lifetime scale of the vortex shown in Figure 3.12.*

to make sure that tracking of the vortex center and the boundary is applied to the same vortex over a period of time. To satisfy this, it was assumed that the maximum displacement distance of the vortex center between two successive frames of this numerical simulation would be 50 km (equivalent to four pixels of the simulation grid). This strategy is based on Giagkiozis et al. (2018), where the speed of the vortex center was assumed to be near the speed of sound in the photosphere (i.e., 10 km/s, given by Nordlund et al., 2009). Therefore, if in two consecutive frames, a vortex centre is detected within 50 km from another vortex centre in the previous frame, they are the same vortical structure.

The temporal evolution of the boundary and center of vortex number 1 is presented in Figure 3.12. The position of the vortex center is mostly oscillating during its lifetime (i.e. between t_0 and t_7) around the initial position. This means that the vortex does not show significant displacement in space. The boundary tends to shrink close to the end of the vortex lifetime (t_7). The decrease in the vortex area is shown in Figure 3.12(b). The blue and orange curves depict the area and average radius of the vortex as functions of time, respectively. An identified average radius around 35-45 km is in accordance with previous findings for lower photospheric vortices in MURaM magneto-convection simulations (Silva et al., 2020; Aljohani et al., 2022). The vorticity and Γ_1 as functions of time are shown in Figure 3.12(c). This result can be explained in the following way. The vorticity does not start at zero, due to the fact that the velocity field lines start to curl even before the detection by Γ_1. Due to the threshold, the vortex is only considered to start to exist when its Γ_1 is greater than the threshold. Close to the end of its lifetime, there is a steep decrease in the vorticity magnitude, which starts at $t_5 = 30$ s close to the vortex decay at $t_7 = 40$ s. As expected, the Γ_1 value (orange curve) shows a similar behavior.

3.4.2 Statistical Results

First, to evaluate how different kernel sizes affect the statistics of the identified vortices, the AGM with $ks = 3, 7$ and 11 and the Γ method were applied to the same simulation data. Figure 3.13 summarizes the obtained results. The green, yellow, and purple curves correspond to the results of AGM $ks = 3, 7, 11$, respectively. The blue curve indicates the instantaneous detection result by the adaptive AGM. The Γ method identification results are shown in red. All results show a similar curve trend as a function of time, but they provide a different number of detections. Among the different ks applied for the AGM, the lowest average number of vortex detections—that is, on average 51.1 at

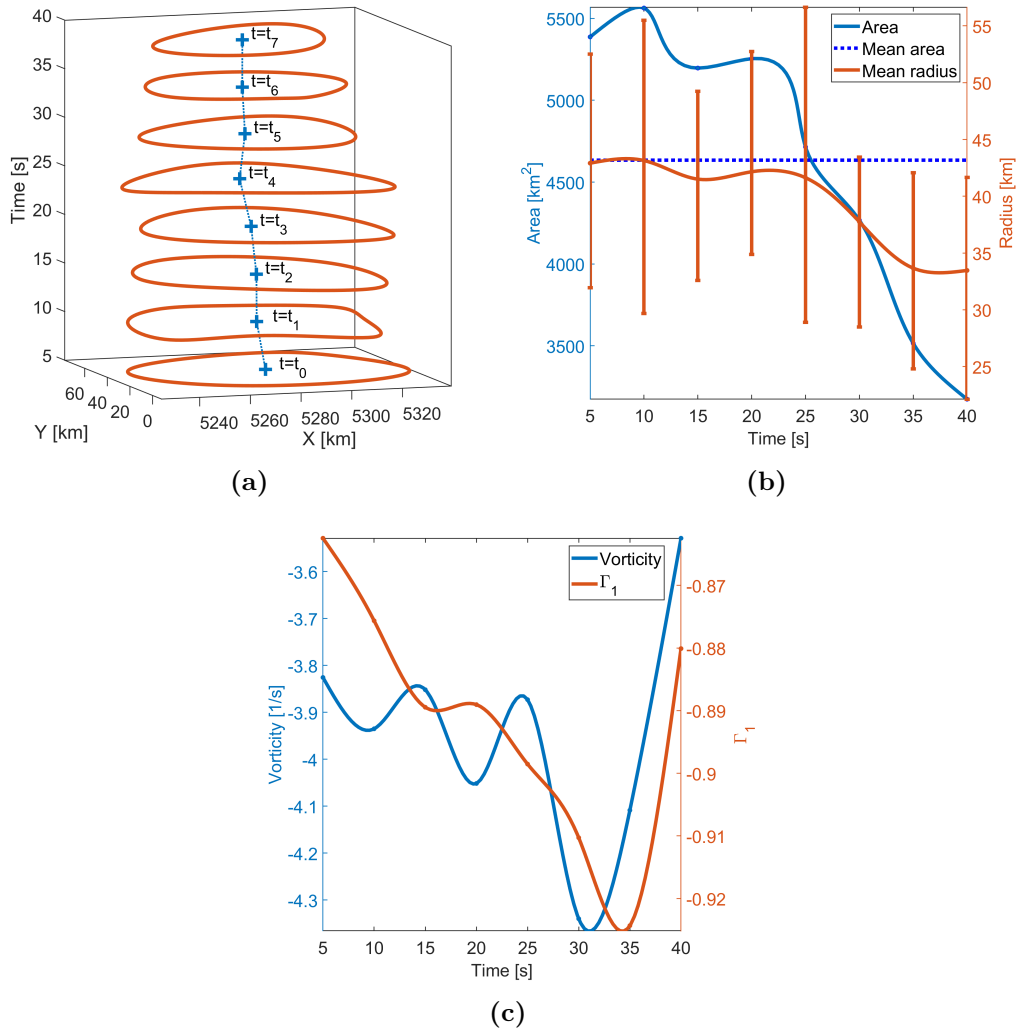


Figure 3.12: *The time-dependent behaviour of the vortex number 1, which is shown in Figure 3.10. (a) A series of detected centers of the vortex (blue plus sign) and its boundaries (orange contour) are shown for times $t_0 - t_7$ from the bottom to the top. (b) The evolution of the vortex area (blue curve) and mean radius (orange curve). The error bar shows the maximum and minimum of the radius at the corresponding time. The dark blue dashed line indicates the mean area of the vortex. (c) The evolution of the vorticity at the center of the vortex (blue curve) and the corresponding evolution of Γ_1 value (orange curve).*

each time frame—is obtained when using $ks = 11$, followed by $ks = 7$, with 95.1 detections and $ks = 3$, with 124.6. The Γ method presents results similar to the AGM with $ks = 3$, with an average detection number of 112.4. The higher number of detected vortices by the adaptive AGM appears to be due to the use of different kernel sizes in the same region of interest (ROI). This makes it possible to detect vortices even if they are very close to each other. By comparing the total number of detected vortices, it was found that the Γ method detects only 69.7% of the actual vortical structures in the ROI compared with the AGM. The AGM $ks = 3$ identifies 77.3%, while AGM $ks = 7$ and AGM $ks = 11$ identifies 59% and 31.6%, respectively. These percentage values can be used to provide an estimate of the actual number of vortical structures present in the photospheric flow and also the number of nondetections of each method applied.

In total 3390 vortices were identified for a time interval of 625 s by the adaptive AGM. Most of the detected vortices tend to anchor around a local region with slight spatial displacement during their lifetime. Figure 3.14(a) shows the spatial and temporal behaviour of a number of identified vortices in the ROI. The evolutions of their boundaries and centers are plotted as a function of time (vertical axis). A zoom-in of the selected area (3×3 Mm) is shown in Figure 3.14(b). Identified vortical motions have similar behaviours to vortex number 1 shown in Figsure 3.12; that is, little spatial displacement in time and a tendency toward boundary shrinking. The majority of the detected vortices have a lifetime below 50 s, which justifies why it is hard to see 150 vortices in Figure 3.14. We have only two vortices presenting a lifetime larger than 200 s, which is in accordance with previous studies, for example, Giagkiozis et al. (2018) and Silva et al. (2021). From Figure 3.13, it is clear that the number of instantaneous detections of vortices by the AGM is larger than the detections by the Γ method, but, at the same time, the lifetimes of the vortices detected by the AGM are also larger. Therefore, the total number of vortices detected by the AGM (3390) is not considerably larger than the number detected by the Γ method (3211).

3.4.3 Comparison of Identification Results Obtained by the AGM and the Γ Method

In this Section, we compare the vortex detection performance between the adapted AGM and the Γ method. Figure 3.16 shows the distributions of various vortex parameters—that is, lifetime, mean area, and diameter—detected by the adaptive AGM (shown in blue) and the Γ method (shown in red).

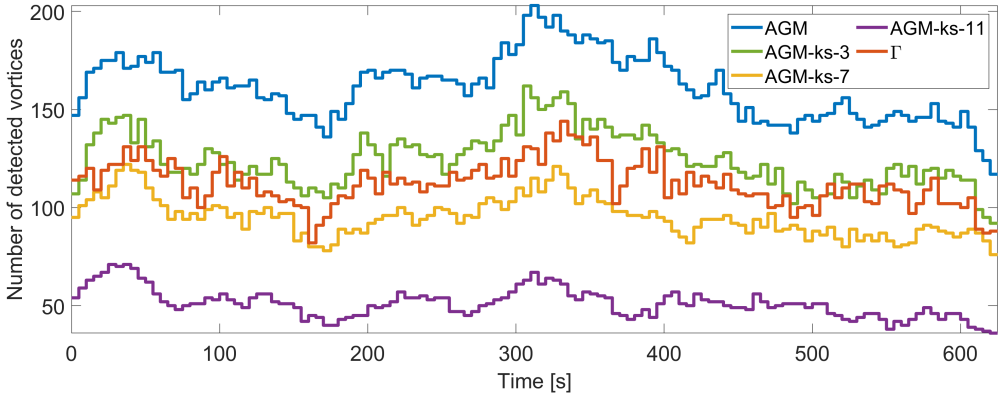


Figure 3.13: *The instantaneous number of vortices detected by the AGM with the use of different kernel sizes: ks = 3 (green), ks = 7 (yellow), and ks = 11 (purple). The results obtained by the adaptive AGM (ks = 3, 5, 7, 9, and 11) are shown in blue, and those by the Γ method are shown in red.*

These distributions were obtained by analysing 625 s of numerical data. The left column panels, that is, Figure 3.16(a), (c), and (e), show the distribution of counterclockwise vortices, and the right panels (b), (d), and (f) depict the clockwise direction vortices. The adaptive AGM detected slightly more vortices than the Γ method; that is, 3390 versus 3211, respectively. The ratio between the counterclockwise and clockwise vortices is nearly 1 : 1: the AGM identified 1697 counterclockwise and 1693 clockwise; the Γ method detected 1608 counterclockwise and 1603 clockwise vortices. The distributions for the detected number, lifetime, mean area, and diameter are independent of the vortex's direction. There are small differences at the tail of the distributions; however, these may appear due to the limitations of the size of our time sequences and do not represent the real physical properties of the vortices. The average lifetimes of detected vortices are 29.72 s for the adaptive AGM and 21.82 s for the vortices detected by the Γ method. This difference relies on the fact that the AGM performs better on the identification of both small- and large-scale vortices when varying kernel size. As it has been shown, close to its lifetime, the vortex tends to shrink and a small convolution kernel enables the AGM to capture this behaviour better than the Γ method. Also, the vortex may start at a smaller size that is not easy to capture by the Γ method.

The mean area distributions are shown in Figure 3.16(c) and (d), and they display nearly Gaussian distribution behaviour. The vortex's mean area detected by the Γ method is 7.28×10^3 km^2 and the corresponding value given

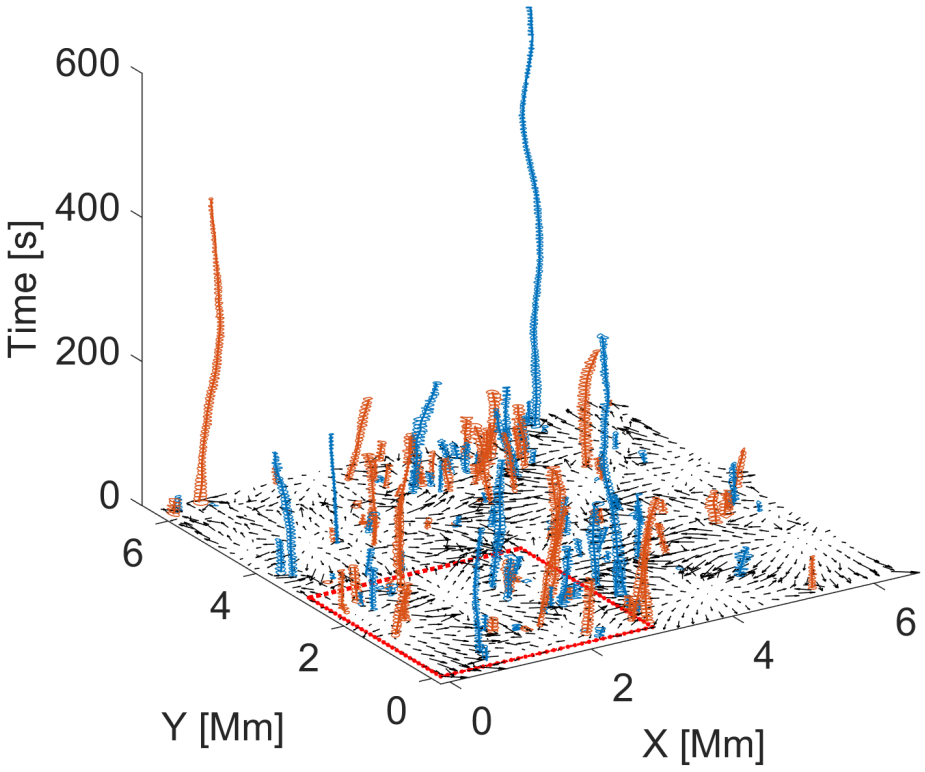


Figure 3.14: *Time evolution of 150 detected vortices (some of them are less visible, due to their short lifetimes; see Figure 3.16) in simulation data. The size of the analysed region is 6.4 Mm × 6.4 Mm. The vertical axis indicates the time, and the black arrow indicates the horizontal velocity field at the initial moment of time. The vortices were detected by the AGM using ks=3,5,7,9 and 11. The orange and the blue contours indicate clockwise and counterclockwise rotation vortices, respectively. The close view of the region enclosed by a red rectangle is presented in Figure 3.15.*

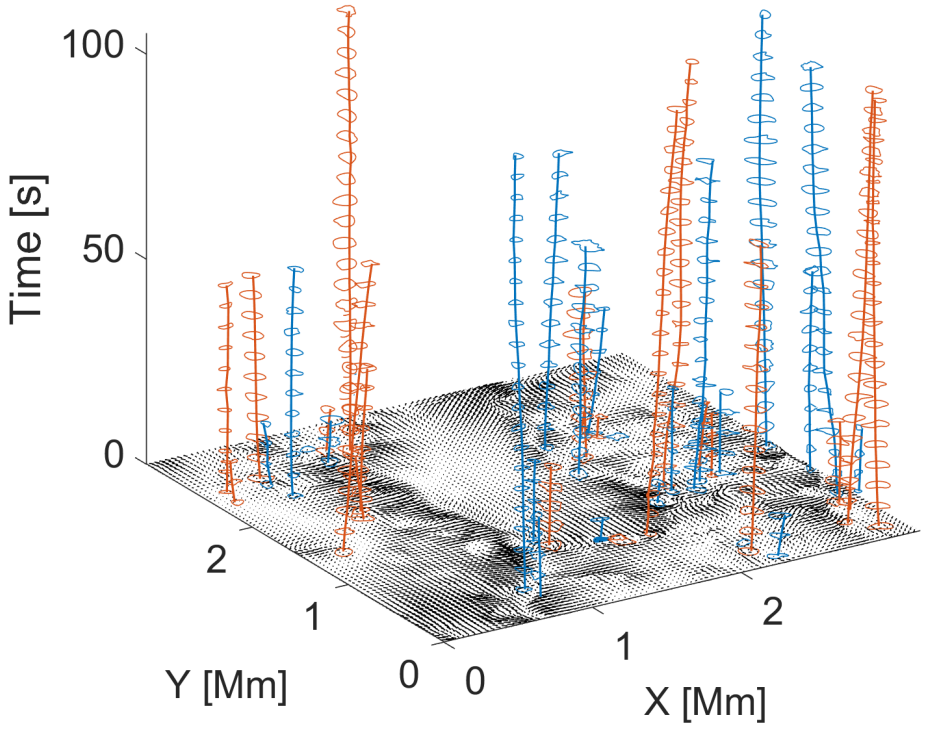


Figure 3.15: *The close view of the analysed region enclosed by the red rectangle in Figure 3.14.*

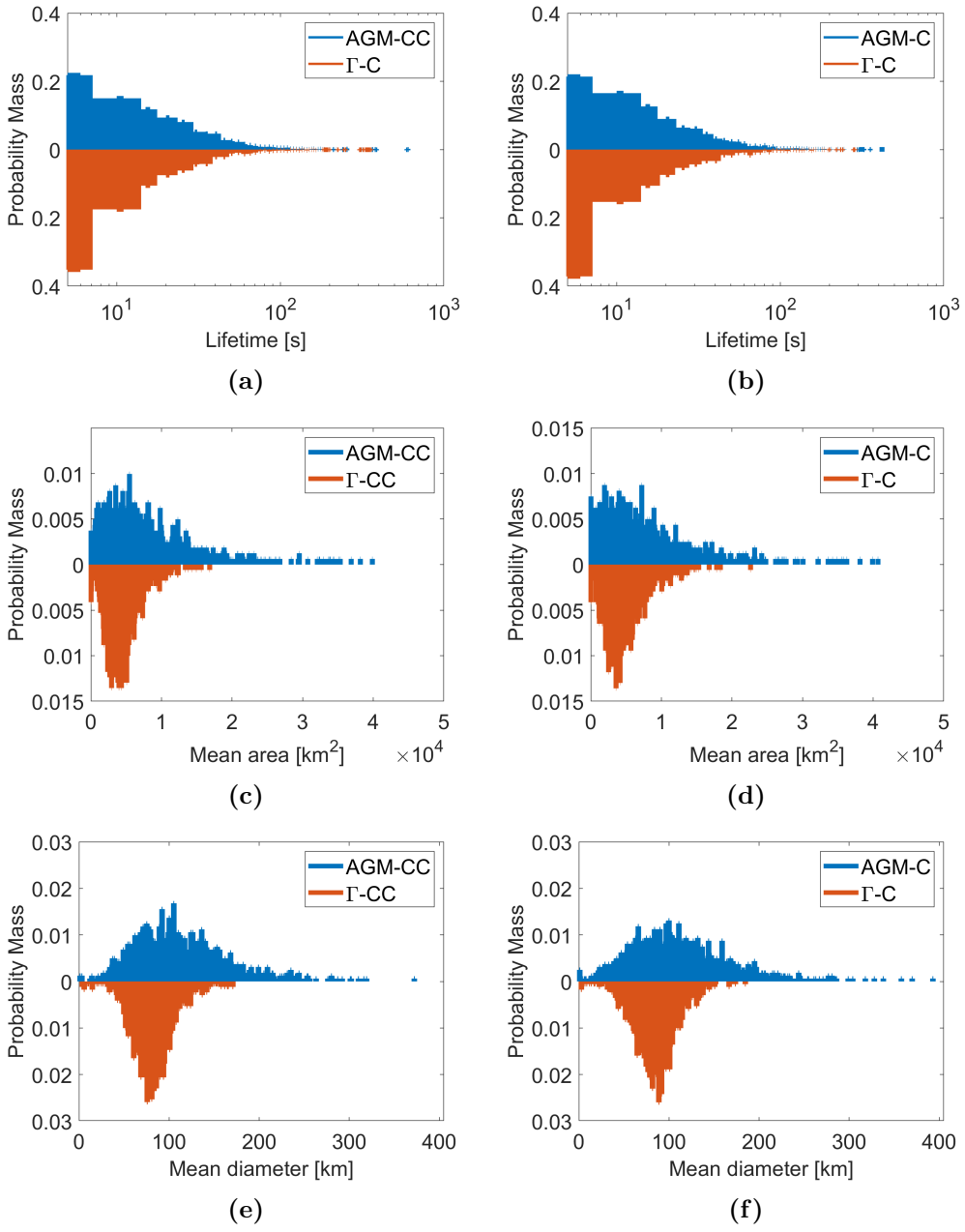


Figure 3.16: *Statistical comparison between vortices detected by the AGM (blue) and the Γ method (red) in numerical data. (a) Probability Mass function of the lifetimes of detected counterclockwise ('CC') vortices. (b) Probability Mass function of the lifetimes of detected clockwise ('C') vortices. (c) Probability Mass function of the mean area of detected counterclockwise ('CC') vortices. (d) Probability Mass function of the mean area of detected clockwise ('C') vortices. (e) Probability Mass function of the mean diameter of detected counterclockwise ('CC') vortices. (f) Probability Mass function of the mean diameter of detected clockwise ('C') vortices.*

by the AGM is 4.67×10^3 km^2. On the other hand, the tail distribution indicates that the AGM has a higher number of defections presenting larger areas. Another pair of Gaussian-like distributions is obtained for the mean diameter of the detected vortex, as depicted in Figure 3.16(e) and (f). The AGM gives a smaller value of the detected vortex's mean diameter than the Γ method; that is, AGM, 83.51 km versus Γ, 110.31 km.

The difference in sizes and shapes for vortices detected by the AGM and the Γ method is illustrated in Figure 3.17, which presents the detections obtained for a selected time frame, $t = 35$ s. The color bar indicates the vorticity scale of the region, and the arrows indicate the direction of the horizontal velocity. The AGM detection is depicted in blue, and the Γ method vortex identification is shown in red. We selected three ROIs, that is, R1, R2, and R3, for comparison. The AGM-detected vortex boundaries were always smoother than those provided by the Γ method, and most of the AGM detections do not overlap, even if vortices are near each other. The Γ method exhibits a sparse detection result within the same region; that is, it has several nondetection issues when the vortices are too close to each other. Also, the Γ method tends to overestimate the vortex region compared to the AGM, but it also depends on the chosen threshold value.

3.4.4 Comparison of Observational Data Analysis with Numerical Data Results

In order to test the performance of the adaptive AGM method on the high-resolution observational data, we used the data set from the CRISP instrument mounted on the Swedish 1-m Solar Telescope (SST, Scharmer et al., 2003b, 2008b). The identification results were compared with the results of the Γ method applied to the same observational time sequence. The analysed data were collected between 08:07:24 and 09:05:46 UT on 2012 June 21, with a resolution 0."059 per pixel with a mean cadence of 8.25 s. Only part (8.5 $\times$ 8.5 Mm) of the whole domain (55" $\times$ 55"), with a of duration 2425.5 s, was selected for analysis. The adaptive AGM detected 1449 vortices within the ROI, with 709 clockwise and 746 counterclockwise vortices. In contrast, the Γ method detected 982 vortices (464 clockwise and 518 counterclockwise). The ratio between the number of vortices with different directions of rotation (clockwise/counterclockwise) is similar for both methods; that is, 0.95 for the adaptive AGM and 0.9 for the Γ method. Therefore, the flexibility of utilizing different kernel sizes in the adaptive AGM when analysing observational data shows good improvement in the identification of small-scale vortices, especially

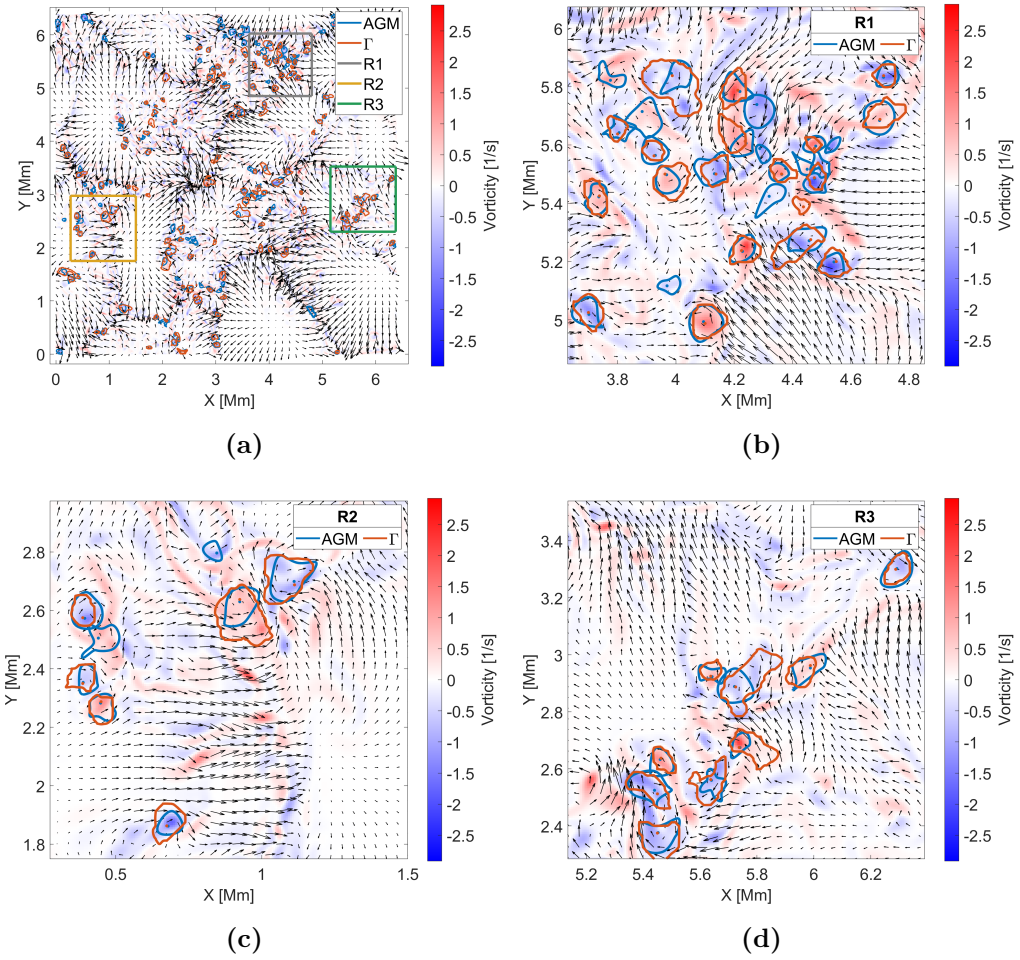


Figure 3.17: *Comparison of detected results between the adaptive AGM method and the* Γ *method. The blue and orange contours indicate vortices detected by using the adaptive AGM and* Γ *method, respectively. Blue dots indicate vortices center identified by the AGM, while orange dots indicate those identified by the* Γ *method. The black arrow indicates the horizontal velocity field. Three ROIs (R1, R2, and R3) were selected to zoom in on and are presented with the details in (b), (c), and (d). (a) Spatial domain when time* $= 35$ *s. (b) Zoom-in view of region R1. (c) Zoom-in view of region R2. (d) Zoom-in view of region R3.*

those with weak vorticity, which may be missed by the Γ method. This is similar to the results obtained from the analysis of numerical magnetoconvection simulations (see Section 3.4.3).

The mean lifetime of the vortices detected by both methods is nearly equal, with 11.79 s for the adaptive AGM and 11.68 s for the Γ method. We also found that the mean area of the vortices detected by the adaptive AGM (0.164 Mm^2) is comparable to the mean area of the vortices detected by the Γ method (0.2 Mm^2). Additionally, the adaptive AGM provided a smaller mean vortex diameter than the Γ method; that is, 504 km and 540.5 km, respectively (see Figure 3.18).

Comparison of Figure 3.16 and 3.18 shows that the ability of the adaptive AGM to detect smaller- and larger-scale vortices is preserved even in the case of the more irregularly shaped vortices. Although the observation data have a lower resolution (42.75 × 42.75 km^2) compared with the numerical data (12.5 × 12.5 km^2), the mean lifetime of the vortices found by both methods is rather similar. The observed vortex lifetimes are most likely longer since crucial parts of the vortex lifetimes (the beginning and end) are missed by the lower resolution. A key difference found between the results of the numerical and observational analysis is that the number of detected vortices in CRISP data is around 12 times smaller than that obtained from the 'StellarBox' simulations (see Figure 3.13 and 3.19). This is in accordance with the difference in the spatial resolution, as the resolution of the simulation data is approximately 12 times higher than that of the observational data. Figure 3.20 shows the identified vortices in the observational data at $t = 231$ s. The Zoom-in views of the three selected regions, namely, R1, R2, and R3, are shown in panels (b), (c), and (d). The adaptive AGM provided more precise vortex detection and showed better performance in comparison with the Γ method in regions with weak vorticity (see Figure 3.20 (d)).

3.5 Conclusions

In this work, we have presented a new and improved method for vortex identification, based on the previously proposed Γ method (Graftieaux et al. (2001)). The CGM (Zigunov et al., 2020) was extended to provide Γ_2 functions, and the CGM was used to develop the AGM. We have proved that the CGM is mathematically the same as the Γ method and that the AGM provides similar maps to the CGM. The advantage of using the AGM relies on choices of kernel sizes, which allows the detection of vortices at both small and large scales. Moreover, the adaptive AGM, which uses different kernel sizes for identification, provides

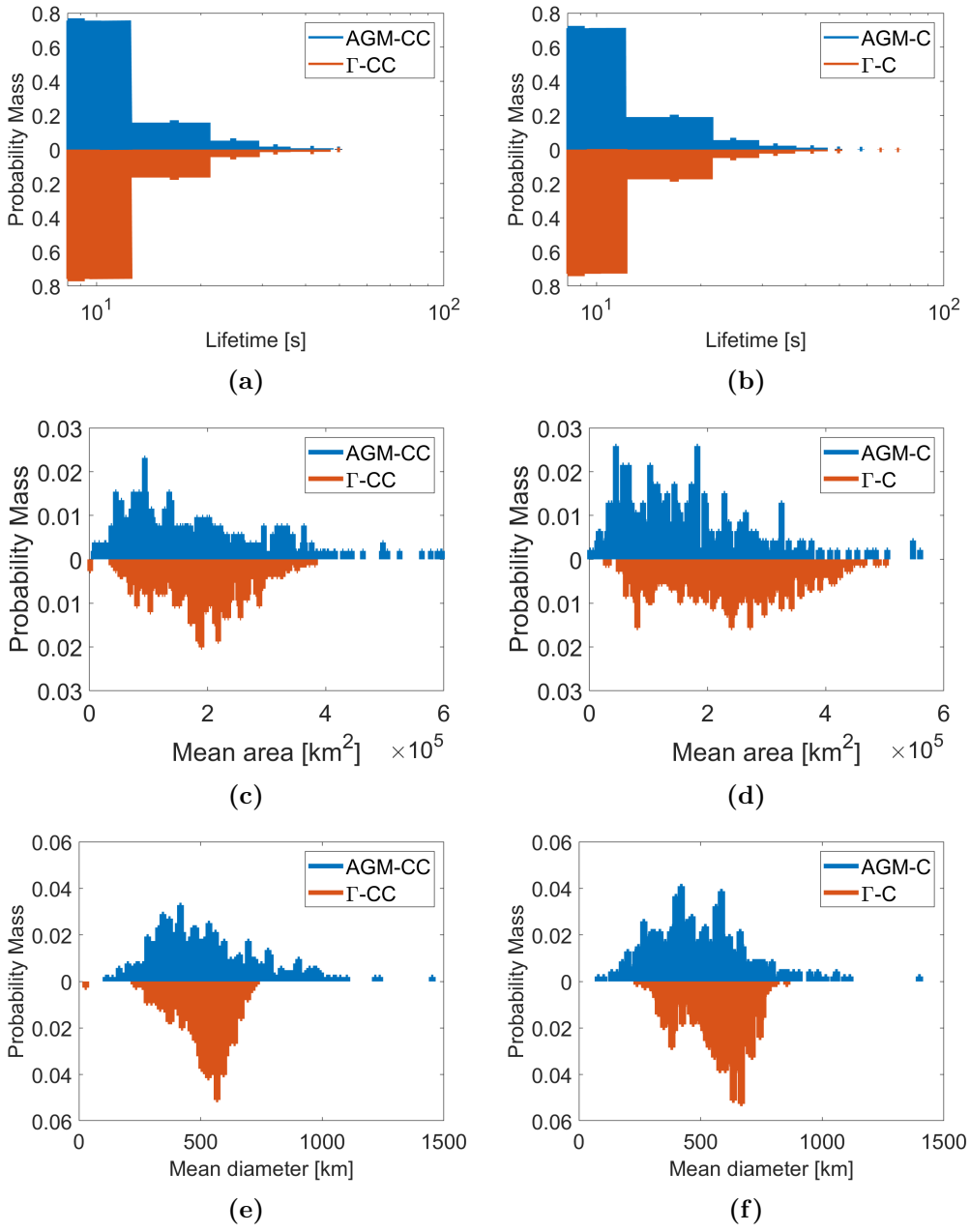


Figure 3.18: *Statistical comparison between vortices detected by the adaptive AGM (blue) and the* Γ *method (red) in observational data. Three results are shown: lifetime probability mass function of the detected counterclockwise ('CC') (a) and clockwise ('C') (b) vortices; Mean area probability mass function of the detected counterclockwise ('CC') (c) and clockwise ('C') (d) vortices; and mean diameter probability mass function of the detected counterclockwise ('CC') (e) and clockwise ('C') (f) vortices.*

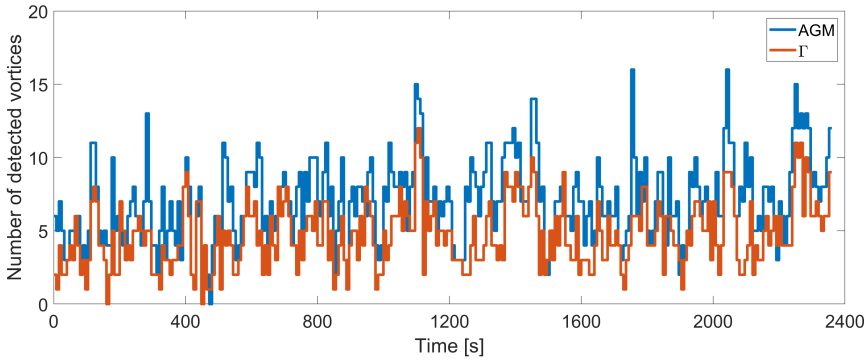


Figure 3.19: *The instantaneous number of vortices detected in the observational data (CRISP) by the adaptive AGM and the* Γ *method. The results obtained by the adaptive AGM with kernel sizes* $ks = 3$*, 5, 7, 9, and 11 are shown in blue, and those obtained by the* Γ *method are shown in red.*

a smoother and more accurate vortex center and a smoother boundary and is more successful at preventing nondetections. For better identification results, the Γ method requires the customized tuning of parameters separately; that is, for small- and large-scale vortices. Hence, this method is onerous and may lead to overfitting for a specific vortex. Since the AGM is better suited to detecting both small-scale and large-scale vortex behavior, the values of the mean area and diameter given by the AGM are more reliable. Note that the adaptive AGM algorithm selects the optimal vortex profiles and produces the most accurate results but is more computationally expensive. Although AGM limits the nondetections caused by threshold choices, there are still limitations on the detections caused by the application of the threshold. More precisely, to be considered as a vortex center, the grid point should present a minimum curl of velocity field lines, and this may lead to the nondetection of weak vortices. Another issue is that it is based on the Γ method, which was developed for divergence-free flows, and therefore the AGM may not work as well in compressible plasma. However, both Giagkiozis et al. (2018) and our results indicate that the Γ method and the AGM perform well in the flow conditions found in the solar atmosphere.

Using the adaptive AGM, we detected and analysed vortices from a numerically simulated photosphere obtained using the 3D radiative MHD code 'StellarBox' (see e.g., Kitiashvili et al., 2012a; Wray et al., 2015). We found that the vortical structures had a size, a mean radius of about 41.75 km, which was similar to previous vortex studies based on MURaM simulation (Silva

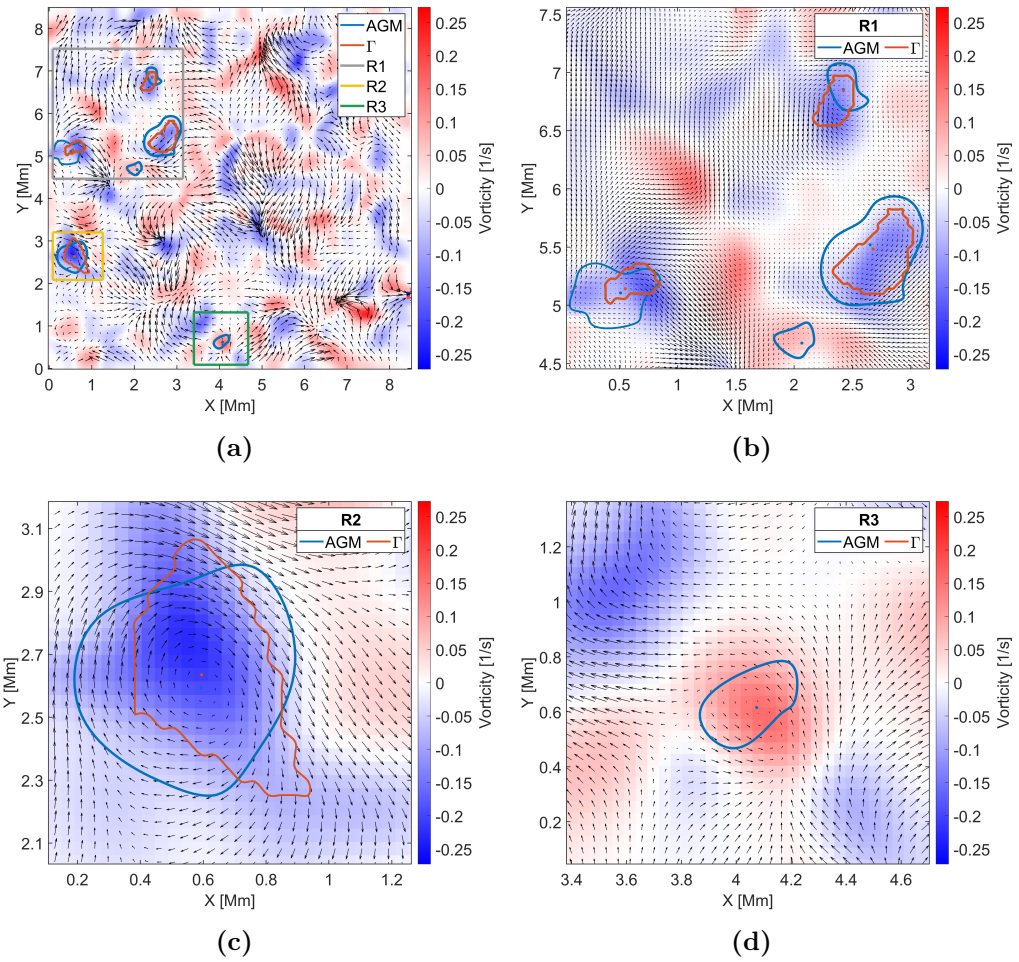


Figure 3.20: *Results of vortex detection by the adaptive AGM and the* Γ *methods. Both methods were applied to observational (CRISP) data at* $t = 231$ *s (a). The blue and red contours indicate vortices detected by the adaptive AGM and the* Γ *methods, respectively. The detected vortex centers from both methods are shown in blue and red colors as well. The black arrows indicate the horizontal velocity field. Three ROIs (R1, R2, and R3) are selected to zoom in on and are presented in panels (b), (c), and (d).*

et al., 2020; Aljohani et al., 2022). Therefore, our results suggest that the different vortex detection methods were comparable in calculating the average vortex size. However, the tail distribution of the mean radius indicates that the Γ method and the AGM lead to considerable differences in the maximum size of vortices detected in the photosphere. Another discrepancy was found for the average lifetimes, as the adaptive AGM indicated an average lifetime that was much shorter than the one found using IVD in MURaM data (Silva et al., 2021). Those differences may be due to distinct definitions proposed for vortices by different methods, as well as the particularities of the simulation configurations. For instance, Giagkiozis et al. (2018) applied the Γ method to an observational data set and found the average lifetime of detected vortices to be around 17 s, whereas applying the same methodology to 'StellarBox' data, we found 21 s. However, in all the studies there is a general tendency toward most of the vortices being short-lived, regardless of the type of data set or identification methodology.

To further assess the adaptive AGM's detection ability, we applied the method to selected observational data obtained from CRISP instrument, found on the Swedish 1-m Solar Telescope (Scharmer et al., 2003b, 2008b). The AGM and Γ methods were applied to the same observational data set, and statistical results were compared to the previously obtained numerical results. It was found that the adaptive AGM extracted more vortices than the Γ method and was more capable of capturing both smaller- and larger-scale vortices. Both methods obtained comparable vortex sizes: the adaptive AGM, 0.164 Mm^2 versus the Γ method, 0.2 Mm^2. The adaptive AGM found a smaller mean diameter than that calculated by the Γ method, in accordance with the results from numerical data. Also, the adaptive AGM was able to detect both smaller- and larger-scale vortices, as indicated by the tails of the distributions in Figure 3.18, reproducing similar results to the numerical data. Therefore, the adaptive AGM works for the identification of realistic irregular vortices. The ratio of average lifetimes calculated by both methods was nearly equal to 1; however, compared with the corresponding statistical analysis with the numerical data, we found that the mean lifetime of detected vortices in the observational data is smaller than that in numerical data. Moreover, the instantaneous vortex detection rate in the observational data is smaller than that in the numerical data. The reason behind this is that the resolution of the observational data used for comparison is 12 times smaller than the numerical simulation data resolution and does not involve any limitations of the adaptive AGM.

The vortices extracted by the AGM tend to have an oval boundary with a vortex core located near the center of the vortical structure. By analysing

vortex properties, the mean tangential velocity component is larger than the corresponding radial velocity component, and the magnitude of the tangential velocity component increases when the component is further from the vortex center. This result is consistent with the previous finding by Silva et al. (2020). By tracking vortices detected by the AGM over a period of time, we found that the size of the vortical structure ceases to exist. The vorticity at the center of the vortex is not constant, oscillating as a function of time and steeply decaying close to the end of the vortex lifetime. Therefore, vortices exhibiting a fast vorticity decay are likely to be at the end of their lives. The velocity components, radial and tangential, also display a distinct radial profile at the beginning and the end of the life of a vortex. In particular, both components tend to decrease around the vortex center as the vortical structure is decaying. In summary, the developed adaptive AGM is able to precisely capture a number of essential aspects of the vortex geometry and dynamics. Therefore, the proposed automated algorithm can be used to recover crucial information on photospheric vortices and statistical results. In the era of high-resolution observational data—for example, DKIST (Rast et al., 2021), the European Solar Telescope (EST, Quintero Noda et al., 2022), and SULIS (Scullion et al., 2022)—we expect that the proposed algorithm and methodology will provide the necessary accuracy for the photospheric vortices detection.

Chapter 4

Vortex categorising for small-scale vortex in the solar atmosphere

4.1 Introduction

Solar vortices are dynamic and swirling structures that play a crucial role in shaping the complex dynamics of the solar atmosphere. They contribute to the magnetohydrodynamic (MHD) waves excitation (Schüessler, 1984; Attie et al., 2009; Fedun et al., 2011; Kitiashvili et al., 2011; Yadav et al., 2022) and plasma jets excitation (Kitiashvili et al., 2013; Iijima and Yokoyama, 2017; Snow et al., 2018; Skirvin et al., 2022) . Solar vortices also act as dynamic channels for energy transfer (Wedemeyer-Böhm et al., 2012; Shelyag et al., 2012; Yadav et al., 2021). These vortical motions extend across various scales and are found in different regions of the solar atmosphere. At smaller scales, short-lived vortices are believed to be abundant, although their direct observation remains challenging due to limitations in spatial resolution (Kato and Wedemeyer, 2017; Giagkiozis et al., 2018; Liu et al., 2019; Yadav et al., 2020). G-band observations from the Swedish Solar Telescope (SST; Scharmer et al., 2003a) have provided valuable insights, revealing the existence of swirling motions around the intergranular lane, characterized by sizes of approximately 500 km and lifetimes with an average lifetime of about 5 minutes (Bonet et al., 2008). On a larger scale (15 - 20 Mm), long-lasting (1-2 h) vortex flows at supergranular junctions have been identified through observations obtained from the Solar Optical Telescope/Filtergraph (SOT/FG) equipped on the Hinode satellite (Attie et al., 2009). Chromospheric observations using instruments

like the CRisp Imaging Spectro-Polarimeter (CRISP) at the SST have further unveiled the presence of chromospheric swirls, presenting a diverse array of vortices with scales ranging from 1.5 - 4 Mm and lasting for 5 - 10 minutes (Wedemeyer-Böhm et al., 2012).

With advances in the observation of solar vortices, the volume of observed vortice structures continues to increase. With such a large scale of data, there is an urgent need to develop an automatic, efficient, and practical analysis strategy for the observed solar vortices. In Chapter 3, we developed an effective vortex identification algorithm, the advanced Γ method (AGM). The adaptive version of the AGM is applied to identify and analyse the photospheric vortices in 3D realistic magnetoconvection numerical and observational data. We compare the detection result with the Γ Method. As a result, the AGM performs better on the vortex detection rate and extracts more accurate boundary profiles for each detected vortex. In other words, the AGM better captures the properties and time-dependent dynamic behavior of vortices.

Clustering analysis is fundamental in data analysis, particularly when data has no label. This analytical technique involves clustering data points with similar characteristics into clusters, revealing underlying patterns and structures within complex datasets. The importance of clustering analysis lies in its ability to discover hidden relationships and categorise intricate phenomena systematically. In the context of solar vortex studies, where vast and dynamic datasets are standard, clustering offers a method to categorise vortices based on shared features. This categorisation facilitates a clearer understanding of the diverse behaviors of solar vortices and enables researchers to identify commonalities and variations among different vortex groups. By unveiling inherent structures and patterns, clustering analysis enhances the efficiency of data interpretation, allowing researchers to go deeper into the investigation of solar vortices. Ultimately, the application of clustering in vortex studies contributes to a more informed and targeted exploration of these complex solar phenomena, advancing our comprehension of their roles and behaviors within the solar atmosphere.

Fundamental clustering techniques, such as the K-means algorithm (Lloyd, 1982), the Gaussian mixture models (GMM), and the density-based spatial clustering of applications with noise (DBSCAN, Ester et al., 1996), are widely applied in clustering analysis. The K-means algorithm is a centroid-based method that partitions data into K clusters by minimizing the total squared distances between data points and their assigned cluster centroids. While computationally efficient, the algorithm may be sensitive to the initialisation of cluster centroid and is limited in handling non-convex clusters. The GMM assumes that data points are generated from a mixture of Gaussian distribu-

tions, allowing for more flexible cluster shapes. However, the GMM requires estimating parameters, making it sensitive to initialization and susceptible to convergence to local optima. The DBSCAN, a density-based method, identifies clusters based on data point density, making it robust against noise and capable of discovering clusters of arbitrary shapes. Nevertheless, the DBSCAN may struggle with varying densities and require careful parameter tuning. In summary, these clustering techniques have different advantages and limitations, and their suitability depends on the data's characteristics and the analysis's goals. In Section 2.5.6, we have applied these clustering techniques to different datasets and evaluated their performance in various scenarios.

Inspired by the clustering analysis, we present the discrete Fréchet distance vortex visualisation method (DFDVVM) as an innovative approach dedicated to quantifying the specific time-dependent behaviour of solar vortices, mainly focusing on their vorticity evolution. Additionally, we introduce a vortex categorisation algorithm based on the K-means method to analyse solar vortices systematically. The solar vortices under examination are simulated by the Stellarbox (a 3D radiative MHD code), and the adaptive AGM identified them. This Chapter is organized as follows: Section 4.2 provides a detailed description of the discrete Fréchet distance vortex visualization method (DFDVVM). Subsequently, Section 4.3 presents the vortex categorisation method based on the K-means algorithm and applies it to solar vortices. The clustering results are thoroughly discussed in this Section. Finally, Section 4.4 concludes this Chapter, summarizing the essential findings and implications.

4.2 Methodology

4.2.1 Description of Simulation data

In Chapter 3, we applied the adaptive advanced Γ method (AGM) to a horizontal velocity field of the solar surface simulated by the StellarBox, a 3D radiative magnetohydrodynamic numerical modeling code, (see, e.g., Kitiashvili et al., 2012a; Wray et al., 2015). As a result, 3390 vortical structures were identified. The size of the simulated velocity field is 6.4×6.4 Mm, and the resolution in the horizontal directions is 12.5 km. The cadence of the simulation data is 5 s, and the duration of the simulation data is 625 s. The AGM has recorded the properties, such as the evolution of vorticity, diameter, and area of each detected vortex during its lifetime. An example of the time-dependent behavior of a detected vortex is shown in Figure 3.12. By investigating the evolution of vorticity profiles of each detected vortex, we find out that vortex exhibit

different vorticity patterns. In this Chapter, we perform cluster analysis on the vorticity behavior of those simulated solar vortices identified by the AGM in Chapter 3.

4.2.2 The K-means Algorithm

In this Chapter, we applied the K-means clustering technique to classify specific time-dependent behavior, that is, the evolution of vorticity of the detected vortex. Lloyd (1982) introduced the K-means algorithm dedicated to partitioning a data group while assigning each observation into K clusters.

In the K-means algorithm, let X represent an n-dimensional data set, and the size of the data set is N, that is, $X = \{x_i | i = 1, ..., N\}$; let $C = \{C_j | j = 1, ..., K\}$ stands for the K clusters in the algorithm and the j^{th} cluster is denoted by C_j. The objective of the K-means algorithm is to optimize the within-cluster sums of squares (WCSS) while allocating data into K clusters. The objective function of the K-means algorithm is defined below

$$J(C) = min \sum_{j=1}^{K} \sum_{x_i \in C_j} \parallel x_i - \mu_j \parallel^2, \tag{4.1}$$

where the centroid of C_j is denoted by μ_j. The K-means algorithm optimizes the corresponding clustering results through an iterative procedure. The iterative process is below:

1. Initialize K cluster centroids.
2. Compute the pairwise distance between each observation and each centroid.
3. Allocate each observation to the cluster with the nearest centroid.
4. Compute the means of the observations within each cluster and update all cluster centroids.
5. Repeat steps 2 to 4 until the clustering membership stays mostly the same or exceeds the maximum number of iterations.

The K-means algorithm requires determining the objective cluster number before the calculation. The common ways to determine the optimal number of clusters are the elbow point method, the Silhouette value criterion, and the Calinski-Harabasz criterion.

The Elbow Point Method

The K-means algorithm optimises the clustering result while minimizing the overall WCSS value. Therefore, the value of the WCSS will continue to decrease during clustering. The elbow point method refers to tracking the value of the WCSS as a function of the cluster numbers. The plot of the function would exhibit an elbow shape. The location of the elbow point relates to the optimal cluster number.

Silhouette Value Criterion

The Silhouette value criterion, introduced by Rousseeuw (1987), is used to evaluate the clustering result when computing the silhouette value of each observation. The definition of the Silhouette value is below

$$S_i = \frac{b_i - a_i}{max(a_i, b_i)}, \tag{4.2}$$

where a_i is the average distance between data i and the other data within the same cluster; b_i is the minimum value of the average distance between data i and other data from different clusters. The range of the Silhouette Value is between -1 and 1. A high positive value implies that the current data are well allocated. In contrast, a low negative value implies improper data clustering, which might belong to another cluster. Therefore, the Silhouette value can be used to determine the optimal cluster number of the given data set. The optimal cluster number corresponds to most data having a positive Silhouette value.

Calinski-Harabasz Criterion

The Calinski-Harabasz criterion, or the variance ratio criterion (VRC), is developed by Caliński and Harabasz (1974) and is utilized to assess the optimal number of clusters for a given dataset. The Calinski-Harabasz criterion is defined as follows:

$$VRC_k = \frac{SS_b}{SS_w} \times \frac{(N-k)}{(k-1)}, \tag{4.3}$$

where SS_b refers to the between-cluster sum of squares and SS_w refers to the within-cluster sum of squares; N is the total number of the observations; k is the number of clusters.

The definition of between-cluster sum of squares in VRC is

$$SS_b = \sum_{i=1}^{k} n_i \parallel m_i - m \parallel^2, \tag{4.4}$$

where n_i is the number of data within i^{th} cluster; m_i denote the centroid of the i^{th} cluster; m denote the total mean of the data. The SS_b is used to evaluate whether cluster and cluster are well split. A larger value of it indicates clusters are well separated.

The definition of within-cluster sum of squares in VRC is

$$SS_w = \sum_{i=1}^{k} \sum_{x \in C_i} n_i \parallel x - m_i \parallel^2, \tag{4.5}$$

where x refers to a observation; C_i represent the i^{th} cluster. The SS_w is used to quantify the data variation within their clusters. A low SS_w implies that clusters are compact. In contrast, a high SS_w implies that the data set is diverse.

From Equation (4.3), we know that both the SS_w and SS_b have been taken into account for evaluating the clustering result. A result with a large SS_b and a small SS_w indicates well-clustering. Therefore, the optimal cluster number of given data is related to the highest Calinski-Harabasz value.

4.2.3 The Fréchet Distance

The Fréchet distance is designed by Fréchet (1906) and applied to calculate the similarity between two curves. The following intuitive example is most often used to explain the concept: A man is walking with his dog while they walk along two different paths; their walking speed may vary, and moving back is not allowed. The minimum length of any dog leash during their walk is defined as the Fréchet distance between them.

Let P and Q denote two polygon curves; the formal definition of Fréchet distance between these curves is expressed as follows:

$$\delta_F(P, Q) = \inf_{\alpha,\beta} \max_{t \in [0,1]} \parallel P(\alpha(t)) - Q(\beta(t)) \parallel, \tag{4.6}$$

where $\alpha(t)$ and $\beta(t)$ represent non-decreasing functions with respect to t. The initial point of P is denoted as $\alpha(0)$, and $\alpha(1)$ signifies its corresponding endpoint. Similarly, $\beta(0)$ and $\beta(1)$ denote Q's starting point and endpoint, respectively. In the definition of the Fréchet distance, the concept of similarity between two curves incorporates the evolution of their respective points as functions of t.

Continuing with P and Q, two sequences of points are used to represent the endpoints on the two curves. Sequence $\sigma(P) = (u_1, u_2, \cdots, u_p)$ denotes each endpoint of P, where the start point of P is u_1 and the endpoint is

u_p. Similarly, sequence $\sigma(Q) = (v_1, v_2, \cdots, v_q)$ represents each endpoint of Q, where the start point of Q is v_1 and the endpoint is v_q. In addition to these two sequences, a new sequence, coupling L, is defined to represent specific distinct pairs between $\sigma(P)$ and $\sigma(Q)$.

$$(u_{a_1}, v_{b_1}), (u_{a_2}, v_{b_2}), \cdots, (u_{a_m}, v_{b_m}).$$

where u_{a_1} and v_{b_1} corresponds to u_1 and v_1, respectively; u_{a_m} and v_{b_m} corresponds to u_p and v_q, respectively. Moreover, for all $i = 1, \cdots, p$, $a_{i+1} = a_i$ or $a_{i+1} = a_i + 1$; for all $i = 1, \cdots, q$, $b_{i+1} = b_i$ or $b_{i+1} = b_i + 1$. Therefore, the order of the points in $\sigma(P)$ and $\sigma(Q)$ has been considered. The length of the coupling L is defined as the length of the longest link among all distinct pairs shown in L:

$$\| L \| = \underset{i=1,\cdots,m}{max} d(u_{a_i}, v_{b_i}), \tag{4.7}$$

where $d(u_{a_i}, v_{b_i})$ represent the Euclidean distance between point u_{a_i} and v_{b_i}. Therefore, the discrete Fréchet distance between P and Q is:

$$\delta_{dF}(P, Q) = min \{ \| L \|, \text{L is a coupling between P and Q} \}. \tag{4.8}$$

To compute the discrete Fréchet distance, Eiter and Mannila (1994) develop a simple and highly efficient algorithm, that is, the algorithm **dF**, to do the calculation. The pseudocode of the algorithm **dF** is provided in Appendix F. Furthermore, Appendix G presents a detailed example to illustrate the calculation procedure of the discrete Fréchet distance when applying the algorithm **dF**.

In summary, the Fréchet distance between P and Q, the corresponding discrete Fréchet distance, and the function **dF** are interconnected in the following manner:

$$\delta_F(P, Q) \text{ approximated by } \delta_{dF}(P, Q) = \mathbf{dF}(P, Q). \tag{4.9}$$

The discrete Fréchet distance exhibits the following properties: suppose P and Q represent two discrete polygonal sequences. If $\delta_{dF}(P, Q) = 0$, it implies that P is identical to Q. In this context, we employ the discrete Fréchet distance to measure the difference in the vorticity evolution of each identified vortex.

4.2.4 The Discrete Fréchet Distance Vortex Visualization Method

In this Section, we present a novel method for vortex analysis known as the discrete Fréchet distance vortex visualization method (DFDVVM). The

DFDVVM utilizes the discrete Fréchet distance as a metric to quantify distinctions in the time-dependent behavior, specifically focusing on the evolution of vorticity of detected vortices. This method facilitates the visualization of differences between solar vortices, contributing to a more comprehensive exploration of their characteristics.

A series of essential preprocessing steps are undertaken to ensure robust analysis before applying the discrete Fréchet distance to measure the difference in the vorticity pattern between vortices. Firstly, an approximation step involves fitting a polynomial curve to approximate the vorticity evolution, preserving the vortex's underlying trend. The vorticity as a function of time of an example clockwise vortex is presented in Figure 4.1(a). The lifetime of this vortex is 65 s. The blue data indicate the vorticity evolution of the example vortex, and the yellow data is the corresponding polynomial curve fitting. Subsequently, a normalization step accounts for the diverse temporal scales observed in solar vortices, ensuring consistency across different scales and directions by rescaling lifetime and vorticity to the range $[0, 1]$. Lastly, a sampling step finalizes the preparation, where discrete points are uniformly sampled from the normalized data, concisely representing the vorticity pattern. Each set of normalized data is sampled with 21 discrete points, as exemplified in Figure 4.1(b). This systematic approach ensures a robust and uniform basis for measuring and comparing vorticity patterns across varied temporal scales and directions.

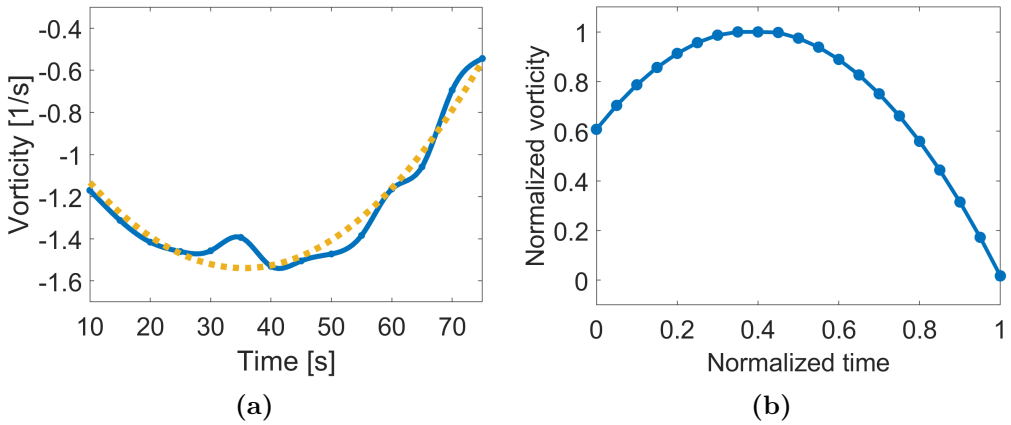


Figure 4.1: *(a) The vorticity evolution at the center of an example clockwise vortex (blue curve) and the corresponding polynomial curve-fitting (yellow dashed curve). (b) The normalized vorticity evolution corresponds to panel (a).*

The choice of the discrete Fréchet distance as the metric for comparing two vorticity data series is based on the consideration that the order of points in the discrete data series is introduced into the calculation. This unique property of the discrete Fréchet distance holds particular significance in capturing the intricate temporal evolution process of solar vortices. In the DFDVVM, three artificial vorticity evolution data series are introduced for comparison, as depicted in Figure 4.2. The purple data represent a vortex with a stable vorticity evolution, the red data correspond to a vortex with an increasing vorticity evolution, and the yellow data illustrate a vortex with a decreasing vorticity evolution. Each artificial data series is treated as a distinct dimension in the DFDVVM analysis. These three dimensions measure the discrete Fréchet distances between processed vorticity data and these three artificial data series, respectively. For instance, the discrete Fréchet distance between processed vorticity data and the purple data is treated as one of the three dimensions. Therefore, these three dimensions focus on measuring the overall vorticity evolution of the vorticity data.

Recognizing that vorticity behavior may vary across different time intervals, we further divide each discrete sampling series into seven groups, each containing three consecutive discrete sampling points. For clarity, in Figure 4.2, black dashed lines indicate the division of discrete sampling series. Discrete Fréchet distances between processed vorticity data and those three data series are then calculated in these corresponding small time intervals, respectively. For example, one of those calculations is to compute the discrete Fréchet distance between the first three consecutive discrete sampling points of a vorticity data series and the purple data's first three consecutive discrete sampling points. These segment calculations result in additional measurements, and the DFDVVM treats these measurements as additional dimensions. Notably, different from the first three dimensions, these additional dimensions emphasize measuring different local vorticity evolution of the vorticity data. Therefore, there have 21 additional dimensions.

Combining the 21 additional dimensions described above with the initial three dimensions, the DFDVVM operates in a 24-dimensional space, termed the DFD space, to comprehensively describe the differences in each vorticity evolution data series. This sophisticated strategy facilitates an in-depth investigation into vorticity variations across temporal scales, enhancing the method's capability to capture intricate patterns in solar vortices.

The adaptive AGM has identified 3390 vortices in the simulation data; 1365 of those vortices have a lifetime below 10 s, while the cadence of the simulation data is 5 s. The DFDVVM requires the analysed vortices to have a lifetime at least longer than twice the cadence of the simulation data. There-

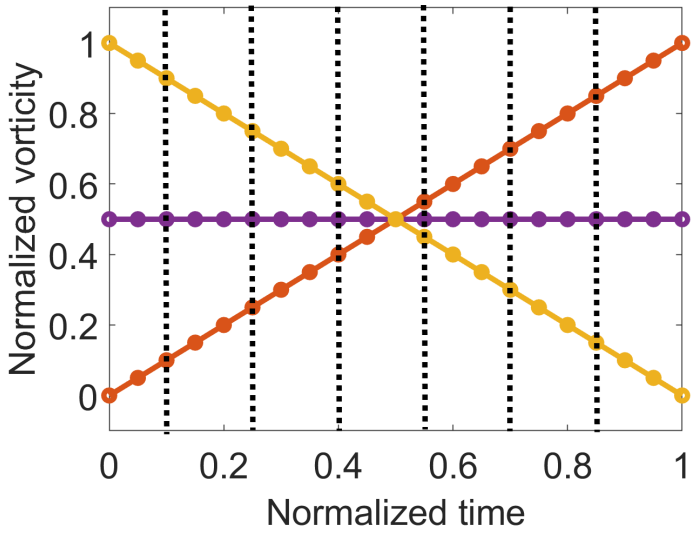


Figure 4.2: *Three artificial normalized discrete vorticity evolution data series. The purple data represent a vortex exhibiting a stable vorticity evolution in its lifetime. The red data represent a vortex with an increasing vorticity evolution in its lifetime. The yellow data represents a vortex with a decreasing vorticity evolution in its lifetime. The black dashed lines show that each vorticity data is separated into seven groups with three sampling points.*

fore, only the remaining 2025 vortices fulfill this requirement. Utilizing the 24 dimensions described previously, the DFDVVM calculates the position of each vortex in the DFD space. Figure 4.3 illustrates a 3D subspace of the DFD space, with Axis 1, Axis 2, and Axis 3 representing the discrete Fréchet distances between the vorticity evolution of the detected vortex and the purple, yellow, and red data from Figure 4.2, respectively. In this subspace, each green point signifies the unique vorticity evolution of an individual vortex, while the purple, yellow, and red points correspond to the artificial vorticity evolution patterns. The blue point corresponds to the clockwise vortex shown in Figure 4.1. Vortices with similar vorticity patterns are positioned closer in the DFD space. By selecting specific dimensions, the DFDVVM provides an intuitive and compact visualization of the differences in vorticity behavior across various temporal and spatial scales. Thus, the DFDVVM offers an effective means of visually exploring the distinctions in vorticity behavior among vortices of different temporal and spatial scales.

4.3 The Vortex Clustering Results and Statistical Analysis

To analyse the vorticity pattern of detected vortices and categorise them, we employ the K-means algorithm to cluster each vorticity pattern measured in the DFD space by the DFDVVM. The reason for clustering with the K-means algorithm is that the given data are relatively uniform distributed and positioned close to each other in the DFD space (see Figure 4.3); it is not intuitive to identify the difference in density of the distributed data or assume whether mixtures Gaussian models govern the distribution of clusters within the data set; as a result, clustering methods, such as the density base method (e.g., DBSCAN) and mixture models method (e.g., GMM) fail to distinguish different pattern or identify similar pattern into a same cluster; moreover, density based method and mixture models method might yield different clustering result for the given data in each independent test; therefore, the K-means algorithm is applied as the location of each vortex in the DFD space corresponds to their difference between each other.

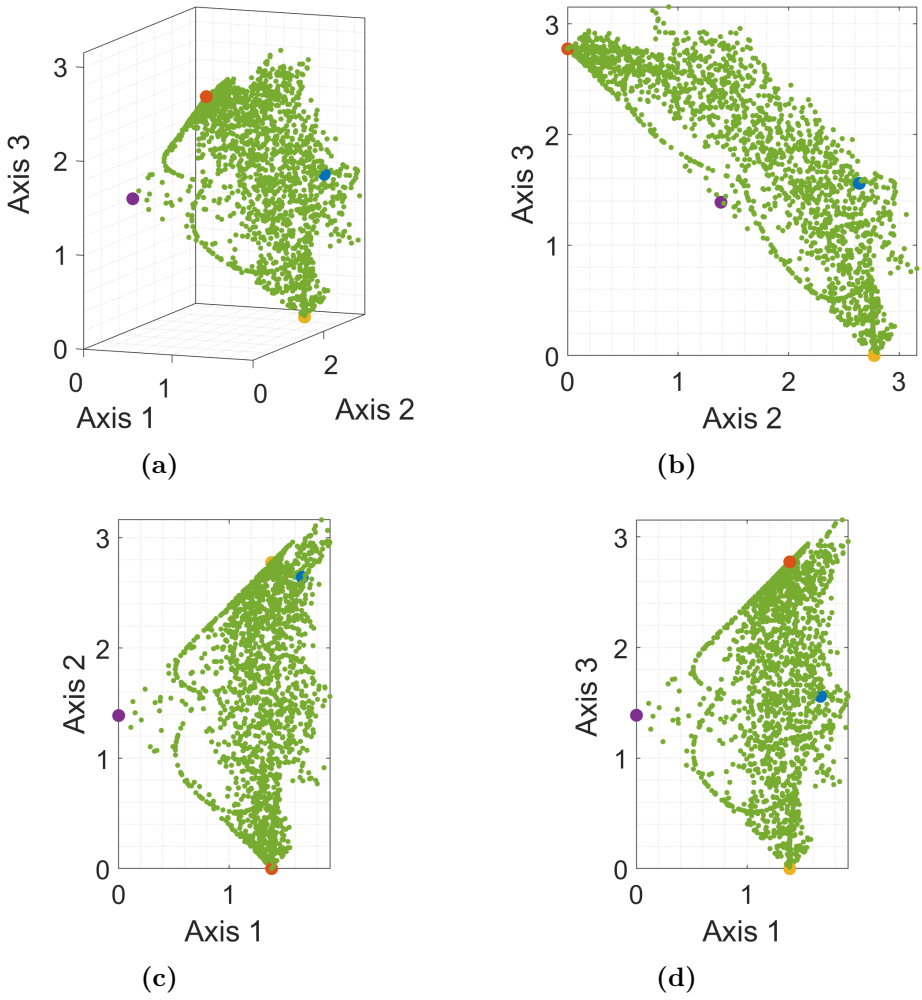


Figure 4.3: *The spatial distribution of vorticity evolution (green points) of detected vortices in the first three dimensions of the DFD space. Axis 1, Axis 2, and Axis 3 are the three dimensions that calculate the discrete Fréchet distance between the vorticity evolution of the detected vortex and the purple, yellow, and red data shown in Figure 4.2, correspondingly. The purple, yellow, and red points correspond to the purple, yellow, and red data in Figure 4.2, respectively. The blue point corresponds to the example clockwise vortex shown in Figure 4.1.*

4.3.1 The Determination of the Optimal Cluster Number

The first step in applying the K-means algorithm is determining the optimal cluster number (k) for the given data. To achieve this, we employ the elbow point method, the Silhouette value criterion, and the Calinski-Harabasz criterion to evaluate the clustering result of different cluster numbers and determine the optimal one with the best result. Figure 4.4 shows the within-cluster sum of square (WCSS) as a function of cluster numbers (k). The elbow plot shows that the elbow point is located around $k = 5$, which implies that the optimal choice of the cluster number of the given data is around $k = 5$.

Figure 4.5 and Figure 4.6 display the mean Silhouette value and the Calinski-Harabasz value of 100 independent tests of the K-means algorithm with different cluster numbers (k). The blue, red, green, and purple data correspond to $k = 4$, 5, 6, and 7, respectively. Figure 4.5 shows the red data ($k = 5$) consistently demonstrated the highest and most stable mean Silhouette value (around 0.46) over 100 independent tests, indicating robust cluster assignments and stable memberships. In contrast, the mean Silhouette values for $k = 4$, 6, and 7 were smaller, exhibiting more significant fluctuations (see blue, green, and purple data in Figure 4.5). Therefore, the Silhouette value criterion suggests that the optimal cluster number for the given data is $k = 5$.

In Figure 4.6, all data shows a stable Calinski-Harabasz value in this 100 test. The red data ($k = 5$) have a smaller Calinski-Harabasz value than the blue data ($k = 4$). Both have a more considerable Calinski-Harabasz value than green ($k = 6$) and purple ($k = 7$) data. As noted before, the Calinski-Harabasz criterion suggests that the optimal cluster number of the given data corresponds to the highest Calinski-Harabasz value. Therefore, the Calinski-Harabasz criterion suggested that $k = 4$ or 5 is a suitable choice for the cluster number, with a slight preference for $k = 4$. Although the Calinski-Harabasz criterion suggested $k = 4$, the overall stability of cluster membership, lower WCSS value, and elbow point at $k = 5$ led us to select $k = 5$ as the optimal cluster number for obtaining stable and meaningful clustering results.

4.3.2 The Vortex Clustering Results

In this Section, we employ the K-means algorithm with a cluster number (k) of 5 to analyse the vorticity evolution data of each detected vortex measured by the DFDVVM. As previously explained, the DFDVVM quantifies the vorticity evolution of each vortex in a 24-dimensional space (DFD space). Similar to Figure 4.3, Figure 4.7 displays a three-dimensional subspace of the DFD

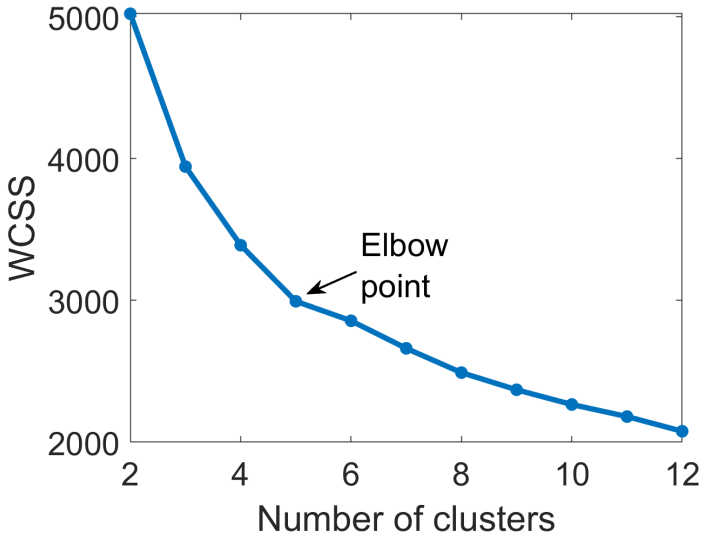


Figure 4.4: *The within-cluster sum of square (WCSS) as a function of cluster numbers of the K-means algorithm.*

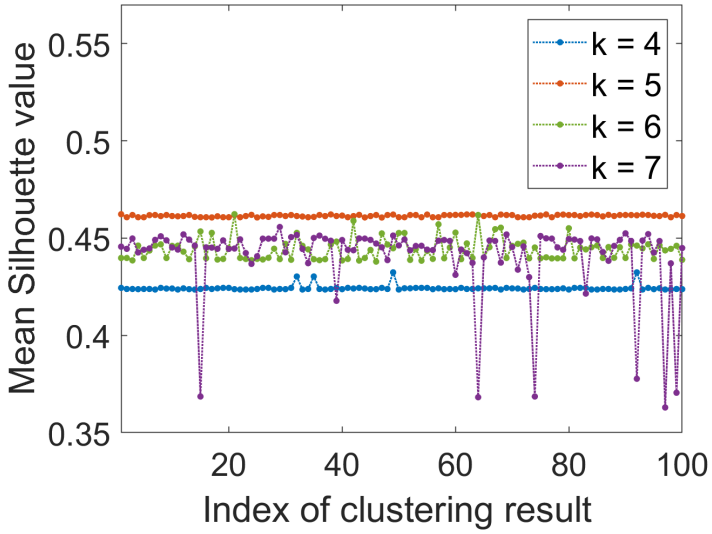


Figure 4.5: *The mean Silhouette value of the K-means algorithm with different number of clusters (k) in 100 independent tests.*

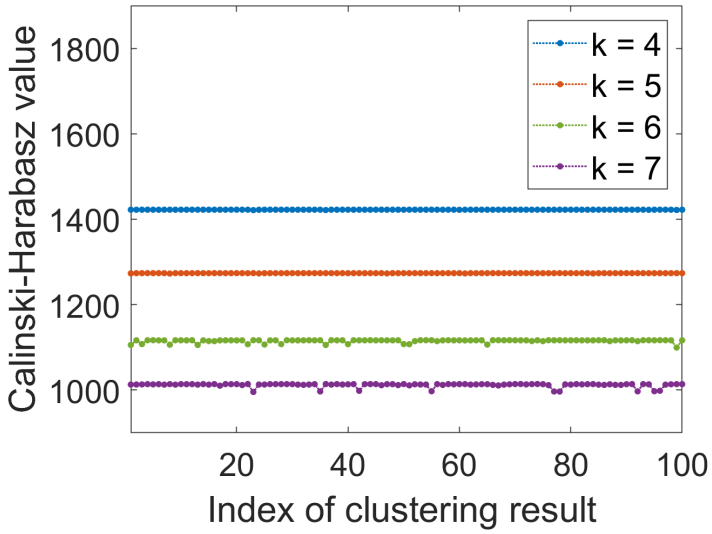


Figure 4.6: *The Calinski-Harabasz value of the K-means algorithm with different numbers of clusters (k) in 100 independent tests.*

space, where these three dimensions correspond to the first three dimensions of the DFD space. Each point in this space represents an individual vortex's vorticity evolution, and five distinct colors are used to indicate the clusters suggested by the K-means algorithm. The black crosses within each cluster signify the centroids calculated by the algorithm, representing the average vorticity pattern of those clusters. Figure 4.8 presents these distinctive average vorticity patterns in associated color. Vortices in cluster 1 (blue) demonstrate a declining vorticity evolution throughout their lifetime. In contrast, cluster 2 (yellow) vortices exhibit an increasing vorticity behavior. The green cluster (cluster 3) displays a vorticity pattern with two phases: an initial increase peaking around 70% of their lifetime, followed by a decrease in the remaining 30%. Similarly, cluster 4 (red) vortices exhibit increasing vorticity in the first 35% of their lifetime, with a subsequent decrease in the remaining 65%. Finally, vortices in cluster 5 (purple) show a relatively stable vorticity behavior characterized by slight fluctuations throughout their lifetime.

In Figure 4.9, we plot the corresponding Silhouette value of the vorticity evolution of each detected vortex. The black dashed line indicates the mean Silhouette value (0.4618) of the overall data. According to the Silhouette value criterion, data with high positive values implies that they are near the current cluster centroid, and this value corresponds to the data allocation quality.

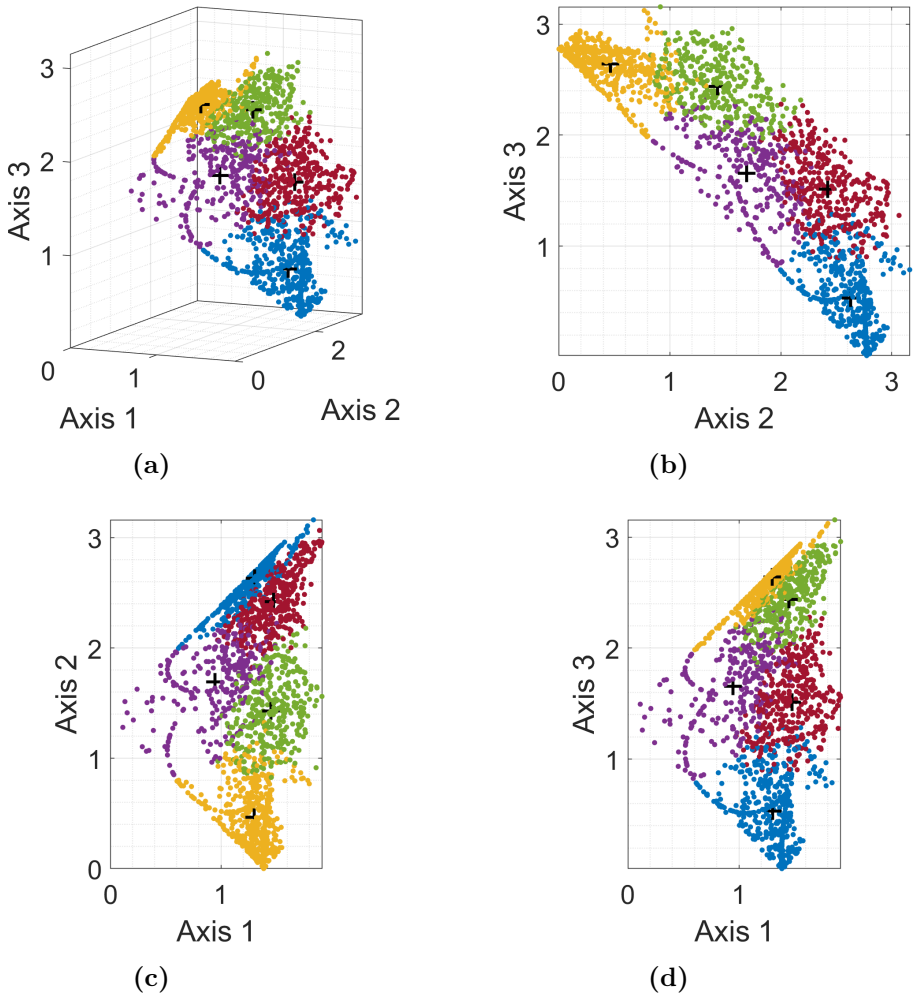


Figure 4.7: *The clustering result of detected vortices' vorticity evolution (points) in the first three dimensions of the DFD space. Axis 1, Axis 2, and Axis 3 are the three dimensions that calculate the discrete Fréchet distance between the vorticity evolution of the detected vortex and the purple, yellow, and red data shown in Figure 4.2, correspondingly. Five colors are used to denote the five clusters: cluster 1 (blue), cluster 2 (yellow), cluster 3 (green), cluster 4 (red), and cluster 5 (purple). The black crosses within each cluster denote the corresponding cluster centroid.*

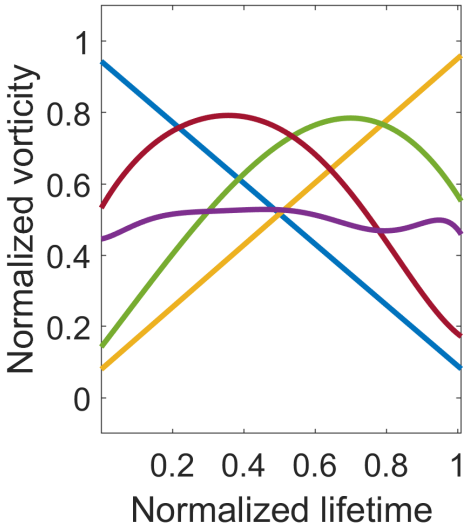


Figure 4.8: *The five average vorticity patterns that are suggested by the K-means algorithm. The blue, yellow, green, red, and purple data corresponds to clusters 1, 2, 3, 4, and 5 shown in Figure 4.7, respectively.*

Most data have a positive Silhouette value higher than 0.45, implying the K-means algorithm with k = 5 allocates them well. Vortices allocated in clusters 1 and 2 have no negative Silhouette value, implying their allocation is not ambiguous. In contrast, some vortices in clusters 3, 4, and 5 exhibit slight negative Silhouette values, suggesting that these specific vorticity patterns are positioned near the border regions of their respective clusters. This proximity results in a subtle overlap with neighboring clusters. In other words, there is some degree of proximity between those vorticity patterns, and the magnitude of the Silhouette value quantifies this proximity. In Figure 4.7, clusters 3, 4, and 5 are situated next to each other, with a minor overlap region. The width of each cluster in the Silhouette plot reflects the quantity of data assigned to those clusters. Interestingly, clusters 1 and 2, while having a similar quantity of data, exhibit larger widths compared to clusters 3, 4, and 5.

Examining the widths of the clusters in the Silhouette plot provides additional insights into the data distribution. Notably, the varying widths of the clusters in the Silhouette plot (Figure 4.9) provide insights into differences in data distribution within each cluster. Despite having a similar quantity of data and nearly no negative Silhouette value, clusters 1 and 2 exhibit larger widths than clusters 3, 4, and 5. This difference in the widths of each cluster and the

Silhouette value may indicate variations in the internal relation of vorticity patterns within these clusters. In other words, vortices within clusters 1 and 2 have less variation in the vorticity pattern than clusters 3, 4, and 5.

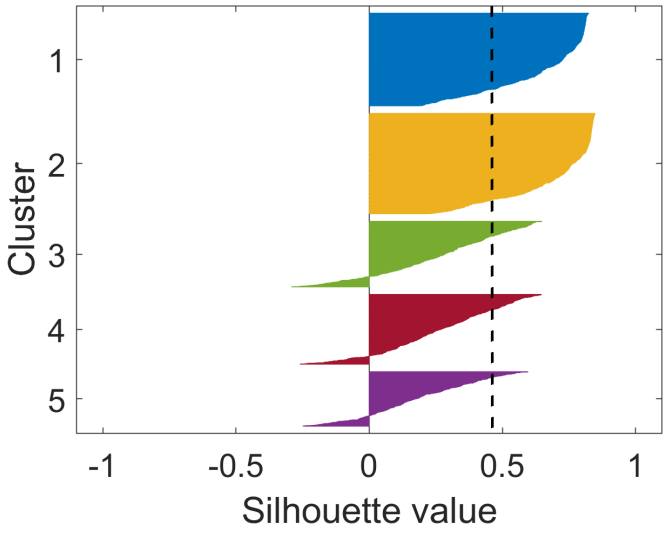


Figure 4.9: *The Silhouette plot of the five clusters. Cluster 1, 2, 3, 4, and 5 corresponds to the blue, yellow, green, red, and purple clusters shown in Figure 4.7, respectively. The black dashed line indicates the overall data's average Silhouette value (0.4618).*

4.3.3 Statistical Results

This Section presents the statistical results of five clusters suggested by the clustering algorithm. Figure 4.10, 4.11, 4.12 shows the distributions of various vortices parameters of the five clusters, including lifetime, mean area, and mean diameter. Here, two colors were used to indicate the direction of the vortex: blue for counterclockwise (CC) and orange for clockwise(C).

Table 4.1 summarizes the parameters of each cluster, revealing a nearly 1:1 ratio between counterclockwise and clockwise vortices in every cluster. Notably, Clusters 1 (blue) and 2 (yellow) exhibit the highest quantity of vortices, each comprising around 500 vortices. Similarly, Clusters 3 (green) and 4 (red) contain approximately 360 vortices each, while Cluster 5 (purple) has the fewest vortices at 288. These quantities align with the widths of the corresponding clusters observed in the Silhouette plot (see Figure 4.9).

The distributions of lifetime, mean area, and mean diameter are independent of the vortex's direction. An interesting observation is that the average vorticity evolution of Clusters 1 and 2 exhibits mirror symmetry about x = 0.5, as does the evolution of Clusters 3 and 4 (see Figure 4.8). These pairwise clusters also have similar numbers of vortices. Specifically, the mean lifetimes of Clusters 1 and 2 are 34.1 s and 39.1 s, respectively, while Clusters 3 and 4 have longer mean lifetimes at 48.2 s and 45.1 s, respectively. Cluster 5 has the most considerable mean lifetime, reaching 65.2 s.

Cluster	1	2	3	4	5
Number of counterclockwise	249	259	163	185	151
Number of clockwise	241	272	184	184	137
Mean lifetime [s]	34.1	39.1	48.2	45.4	65.2
Mean area $[km^2]$	4733	4989	5164	4932	5376
Mean diameter $[km]$	83.74	86.11	86.78	85.47	87.55

Table 4.1: Physicial properties of five different clusters

The mean area distributions of the five clusters exhibit a near Gaussian distribution (see Figure 4.11). Cluster 1 has the smallest mean area among the five clusters, measuring 4.73×10^3 km^2. Clusters 2, 3, and 4 have similar mean areas, and they are larger than those of cluster 1, with values of 4.89×10^3 km^2, 5.64×10^3 km^2, and 4.93×10^3 km^2, respectively. Cluster 5 has the largest mean area, 5.37×10^3 km^2. The mean diameter distributions of the five clusters also exhibit a Gaussian-like pattern (see Figure 4.12). All clusters present mean diameters in close proximity: 83.74 km (Cluster 1), 86.11 km (Cluster 2), 86.78 km (Cluster 3), 85.47 km (Cluster 4), and 87.55 km (Cluster 5).

The instantaneous number of vortices for each cluster is depicted in Figure 4.13. The blue, yellow, green, red, and purple data correspond to Clusters 1, 2, 3, 4, and 5, respectively. The purple data indicates that Cluster 5 has the highest average number of vortex detections, averaging 34.5 at each time frame. Cluster 2 closely follows with 32.1 detections. Clusters 1 and 4 show similar detection rates, with 27.64 versus 26.1 detections, respectively. Cluster 3 exhibits the lowest average number of vortex detections, with 24.7 detections.

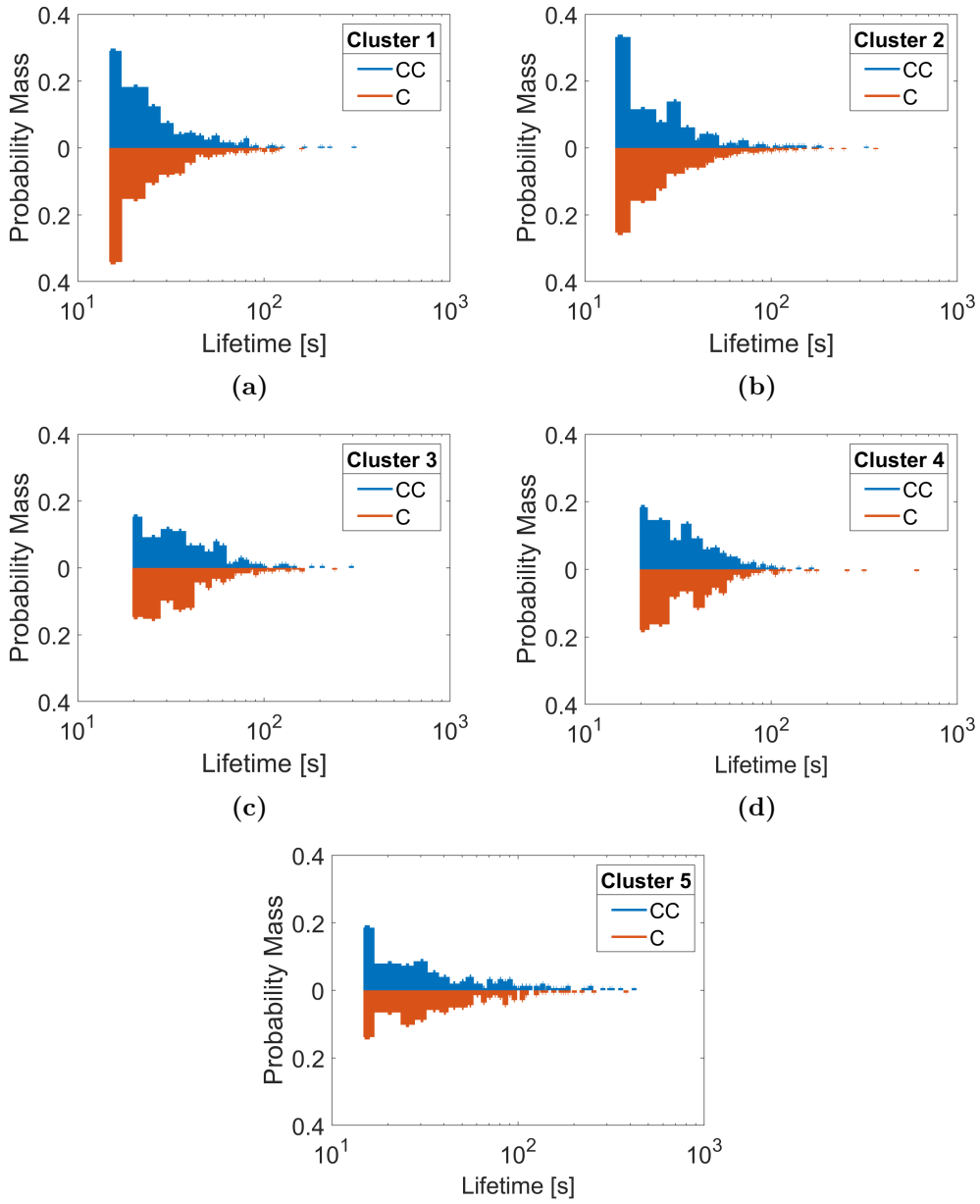


Figure 4.10: *Probability Mass function of lifetime of detected vortices of different clusters. The blue indicates a counterclockwise direction (CC), and the orange indicates a clockwise direction (C).*

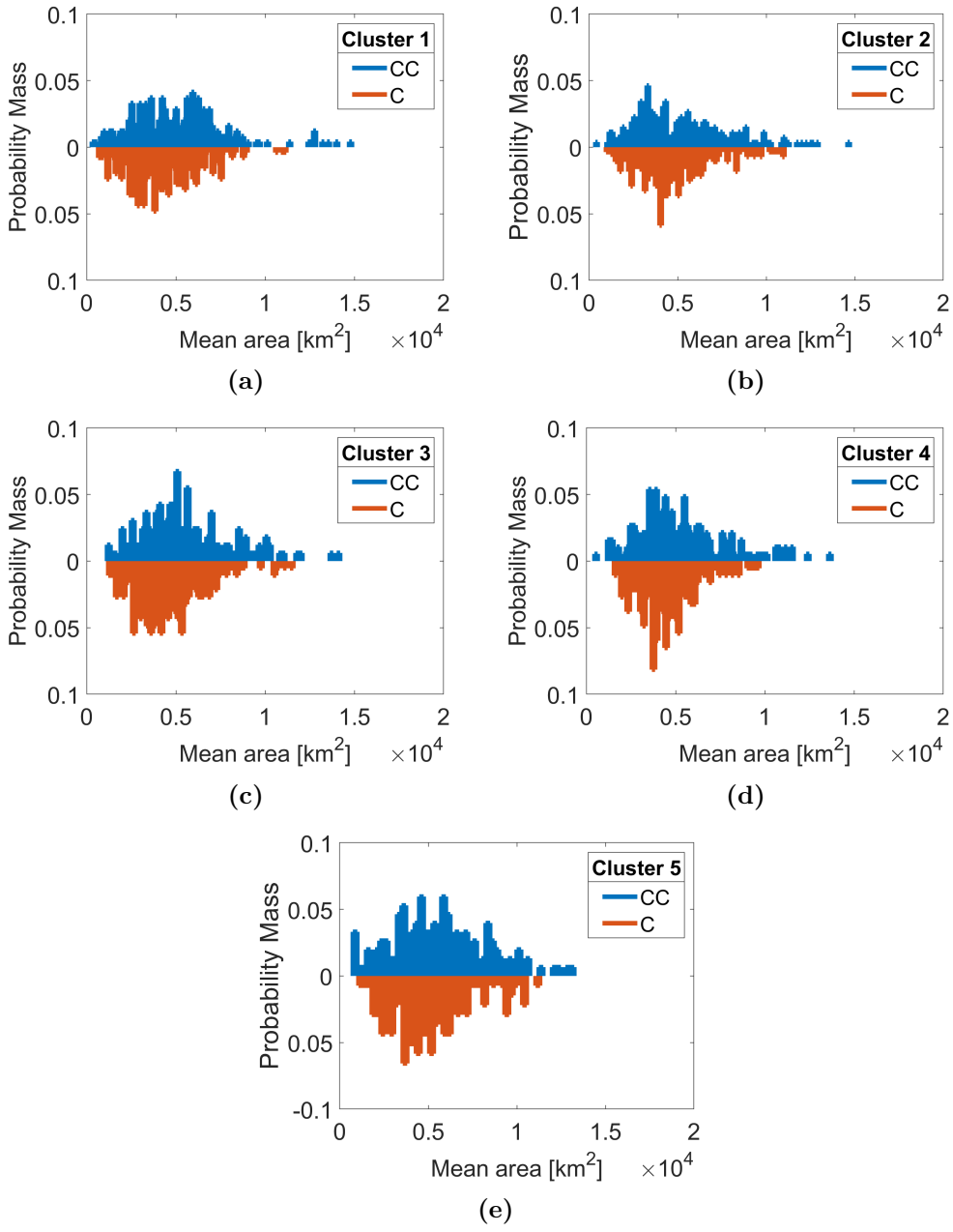


Figure 4.11: *Probability Mass function of mean area of detected vortices of different clusters. The blue indicates a counterclockwise direction (CC), and the orange indicates a clockwise direction (C).*

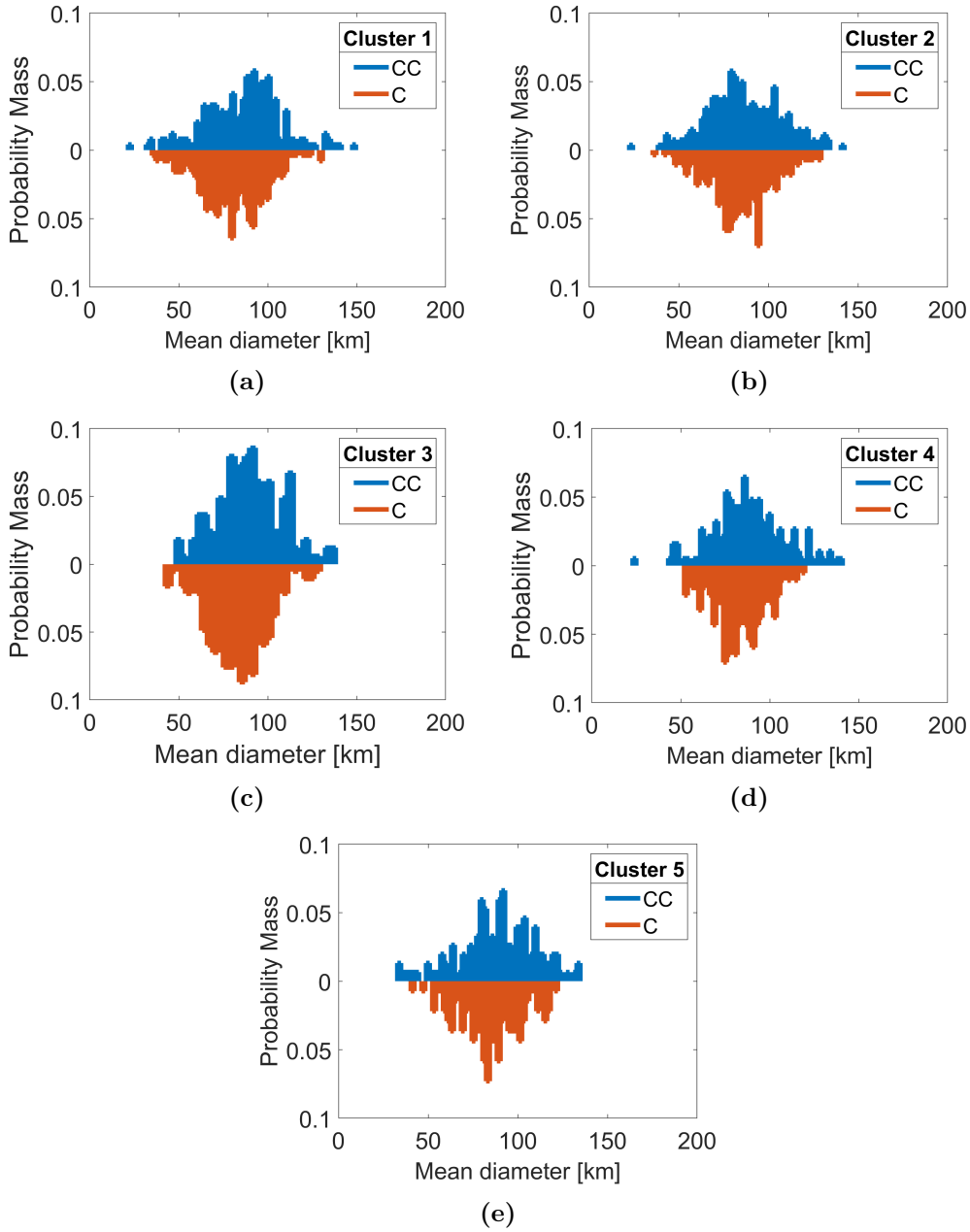


Figure 4.12: *Probability Mass function of mean diameter of detected vortices of different clusters. The blue indicates a counterclockwise direction (CC), and the orange indicates a clockwise direction (C).*

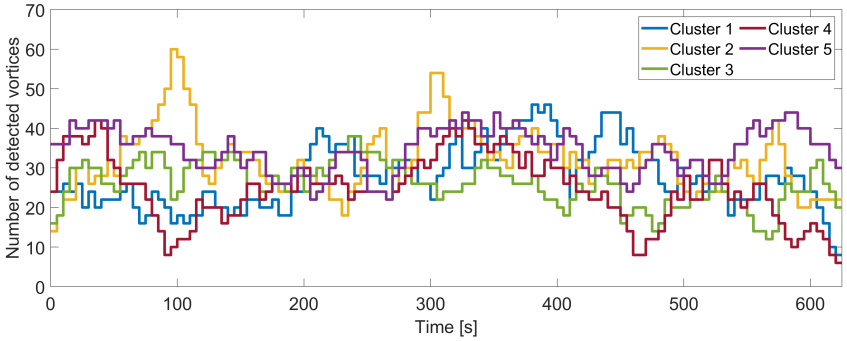


Figure 4.13: *The instantaneous number of vortices of the five clusters.*

4.4 Conclusion

In this study, we explored the complex domain of solar vortex dynamics, employing innovative methodologies to investigate the time-dependent behaviors of vortices simulated by 'StellarBox.' The discrete Fréchet distance vortex visualization method (DFDVVM) emerged as a powerful tool, providing a detailed comprehension of vorticity evolution in a 24-dimensional space (the DFD space). The DFDVVM not only measures variations in the vorticity patterns of solar vortices but also offers an intuitive visualization of these differences in selected dimensions.

The vortex categorising strategy, featuring the application of the K-means algorithm, has demonstrated its capability to organize vortices into meaningful clusters. In this Chapter, the vortex categorising algorithm is applied to cluster solar vortices processed by the DFDVVM. The careful evaluation of clustering results, considering the elbow point method, Silhouette value criterion, and Calinski-Harabasz criterion, led us to the optimal cluster number of $k = 5$. This categorisation revealed distinct clusters with varying quantities of vortices, summarizing the five most representative vorticity patterns.

The statistical analyses of these clusters provided insights into key parameters, including lifetime, mean area, and mean diameter. The average vorticity evolution of clusters that exhibit mirror symmetry to each other added a unique layer to our findings.

The DFDVVM, combined with the K-means-based vortex categorising strategy, opens new avenues for comprehending solar vortices observed in various temporal and spatial scales. Our approaches provide a practical framework for researchers in the field, offering unique insights into the relations and variations between and within vortex clusters.

In conclusion, our study contributes innovative methods and meaningful perspective to analysing solar vortex dynamics. The richness of information extracted by the DFDVVM and the categorising achieved through vortex clustering pave the way for a deeper understanding of these fascinating solar phenomena, advancing the frontier of solar vortice research.

Chapter 5

Conclusions

5.1 Overview of thesis

This thesis is devoted to developing practical and effective analysis methods that assist in investigating solar vortices. Chapter 1 introduced the background of solar vortices, including the observation of solar vortices, reviews of possible formation mechanisms, solar vortices' role in the solar atmosphere, and related analysis.

In Chapter 2, we first reviewed the fundamental properties and principles of differential on which most vortex identification techniques depend. Then, we reviewed several famous vortex identification techniques. We also discussed essential clustering techniques and evaluated their performance in different scenarios.

Chapter 3 introduced a highly efficient and practical vortex identification method, the advanced Γ method based on the Γ method. Additionally, two novel vortex boundary extraction techniques were proposed: the circular sector algorithm and the whole-circular sector algorithm.

In Chapter 4, we extend our analysis methods further by proposing the discrete Fréchet distance vortex visualisation method (DFDVVM) and applying clustering techniques to explore the vorticity evolution of detected vortices in the DFD space.

5.2 Summary of results

5.2.1 Summary of Chapter 3

In Section 3.2, the advanced Γ method (AGM) is developed and applied to identify solar vortices in both simulation and observational data. Using a series of different kernel sizes, the adaptive AGM provides better identification of the vortices structures than the Γ method and prevents nondetection of the vortex as much as possible. Section 3.3 introduced the circular sector algorithm (CSA) and the whole-circular sector algorithm (WCSA). These two algorithms are devoted to extracting a smooth and accurate boundary of vortice structures. The vortex boundary proposed by the CSA and the WCSA is much more accurate since they only consider the regions responsible for a vortex boundary. Therefore, vortex behavior detected by the adaptive AGM is more reliable than those detected by the Γ method. The AGM is applied to detect and analyse the dynamics behavior of vortices simulated by the StellarBox, a 3D radiative MHD code, and the result is discussed in Section 3.4. In particular, an individual vortex's detection is discussed, and the overall statistical result obtained by the two methods, i.e., the adaptive AGM and the Γ method, is compared. It shows that the adaptive AGM has detected more vortices than the Γ method, and the lifetime recorded by the adaptive AGM is slightly longer than the Γ method. Both identification methods show vortices tend to be short-lifetime, which is consistent with other previous simulation research(such as vortex analysis based on MURaM simulation(e.g., Silva et al. (2020); Aljohani et al. (2022))). Moreover, we also analysed observational data obtained from CRISP, and the result of the analysis is compared with the simulation data.

5.2.2 Summary of Chapter 4

In Section 4.2, we introduced the discrete Fréchet distance vortex visualisation method (DFDVVM). It is a novel approach dedicated to quantifying the vorticity patterns of detected solar vortices in the DFD space (a designed 24-dimensional space). In this Chapter, we analyse solar vortices simulated by Stellabox, previously identified by the adaptive AGM in Chapter 3. The DFDVVM is applied to these detected vortices, revealing that the method systematically captures the vorticity characteristics of each vortex in the DFD space. Furthermore, the spatial distribution of vortices in the DFD space correlates with the differences in their vorticity patterns. In Section 4.3, a vortex clustering algorithm inspired by K-means clustering was introduced to group

solar vortices based on their vorticity patterns. This algorithm was subsequently applied to the solar vortices processed by the DFDVVM. At first, we systematically evaluated the clustering results with different numbers to determine the optimal clustering number for the given data, ensuring robust and meaningful categorization. Then, we applied the algorithm with the optimal cluster number to cluster detected vortices. Our result shows that the vortex clustering algorithm has effectively categorized vortices with similar vorticity characteristics into the same clusters. The outcomes of this categorization are summarized, including the average vorticity patterns for each cluster and corresponding statistical results. These analysis methods enhance our comprehension of the time-dependent behaviors of solar vortices, and they are expected to contribute valuable insights into the complex dynamics of these phenomena.

5.3 Future work

The research of this thesis is dedicated to developing practical and efficient analysis techniques that assist in the identification of solar vortices and the investigation of their properties. The investigation presented in the thesis is the first step toward the goal and far away from the objective. Therefore, this research can be extended in the following directions in the future:

In Chapter 3, the advanced Γ method (AGM) is presented and applied to extract the contour of the solar vortex in simulation and observational data. As a result, the method has detected most of the vortex compared to the original Γ method. Each vortex boundary extracted by the method aligned well with the streamline, i.e., the AGM method can accurately capture the vortex geometry and dynamics. The next step is to apply the AGM to high-resolution observational data obtained from DKIST, EST, and other instruments. As noted, the nondetection issue is prevented when utilizing the AGM, and the vortex boundary is much more accurate; therefore, future work is also related to utilizing the mighty computing power of deep learning and neural networks with label data provided by the AGM.

In Chapter 4, the discrete Fréchet distance vortex visualisation method (DFDVVM) and the vortex clustering algorithm were presented and applied to those simulated solar vortices identified by the AGM in chapter 3. In particular, these two algorithms have shown their powerful capability to measure solar vortices, categorize them into different groups, and summarize the corresponding dynamic patterns of each cluster. In this thesis, the DFDVVM quantifies the vorticity pattern of detected vortices in the DFD space. The

future work is related to further developing the DFDVVM, which considers other essential properties of solar vortices and finds a new way to represent the solar vortices.

Appendices

Appendix A

The General Format of the Convolution Kernels

$$PM_x = \begin{bmatrix} -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \\ \vdots & & & & \vdots & & & & \vdots \\ -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \\ -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \\ -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \\ -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \\ -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \\ \vdots & & & & \vdots & & & & \vdots \\ -n & \cdots & -2 & -1 & 0 & 1 & 2 & \cdots & n \end{bmatrix}, \tag{A.1}$$

$$PM_y = \begin{bmatrix} -n & \cdots & -n & -n & -n & -n & -n & \cdots & -n \\ \vdots & & & & \vdots & & & & \vdots \\ -2 & \cdots & -2 & -2 & -2 & 1 & 2 & \cdots & -2 \\ -1 & \cdots & -1 & -1 & -1 & 1 & -1 & \cdots & -1 \\ 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 1 & 1 & 1 & \cdots & 1 \\ 2 & \cdots & 2 & 2 & 2 & 2 & 2 & \cdots & 2 \\ \vdots & & & & \vdots & & & & \vdots \\ n & \cdots & n & n & n & n & n & \cdots & n \end{bmatrix}, \tag{A.2}$$

$$\| \boldsymbol{PM} \| = \sqrt{(PM_x).^2 + (PM_y).^2}$$

$$= \begin{bmatrix} \sqrt{2n^2} & \cdots & \sqrt{n^2+1^2} & n & \sqrt{n^2+1^2} & \cdots & \sqrt{2n^2} \\ \vdots & & & \vdots & & & \vdots \\ \sqrt{n^2+2^2} & \cdots & \sqrt{5} & 2 & \sqrt{5} & \cdots & \sqrt{n^2+2^2} \\ \sqrt{n^2+1^2} & \cdots & \sqrt{2} & 1 & \sqrt{2} & \cdots & \sqrt{n^2+1^2} \\ n & \cdots & 1 & 0 & 1 & \cdots & n \\ \sqrt{n^2+1^2} & \cdots & \sqrt{2} & 1 & \sqrt{2} & \cdots & \sqrt{n^2+1^2} \\ \sqrt{n^2+2^2} & \cdots & \sqrt{5} & 2 & \sqrt{5} & \cdots & \sqrt{n^2+2^2} \\ \vdots & & & \vdots & & & \vdots \\ \sqrt{2n^2} & \cdots & \sqrt{n^2+1^2} & n & \sqrt{n^2+1^2} & \cdots & \sqrt{2n^2} \end{bmatrix}, \tag{A.3}$$

where $n \in \mathbb{Z}^+$.

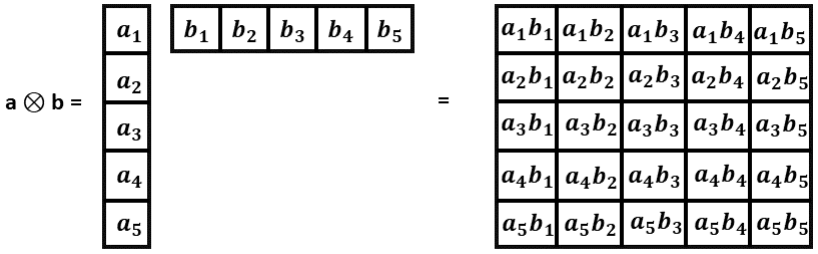
Appendix B

Proof of Separable Convolution Kernel

If $h(m, n)$ is a matrix of size $m \times n$ and its rank is equal to 1, then matrix h can be represented as the outer product of two vectors, $\boldsymbol{w}_1(m, 1)$ and $\boldsymbol{w}_2(1, n)$ as

$$h(m, n) = \boldsymbol{w}_1(m, 1) \otimes \boldsymbol{w}_2(1, n), \tag{B.1}$$

where $\boldsymbol{w}_1(m, 1)$ is a column vector of size $m \times 1$ and $\boldsymbol{w}_2(1, n)$ is a row vector of size $1 \times n$.



(a)

Figure B.1: *The outer product between column vector* $\boldsymbol{a}$ *and row vector* $\boldsymbol{b}$*.*

Gonzalez et al. (2004) have shown that the outer product $\boldsymbol{a} \otimes \boldsymbol{b}$ can be represented as convolution $\boldsymbol{a} \circledast \boldsymbol{b}$ ('padded' convolution operation). The graphical explanation of this is presented in Figures B.1 and B.2. Here, $\boldsymbol{a}$ and $\boldsymbol{b}$ are column and row vectors of size 5, respectively. Figure B.1 shows the result of

the outer product between $\boldsymbol{a}$ and $\boldsymbol{b}$; that is, a matrix of 5×5 elements. The convolution procedure is detailed in Figure B.2(a)–(d), where the operation performed in this example is referred to as the 'padded convolution'. The first step of this operation is to rotate the row vector $\boldsymbol{b}$ by 180° and perform multiplication sequentially between the rotated row vector and the first element of the column vector to compute each element in the first row of the output matrix; see Figure B.2(a). Then, the same procedure is applied between the rotated row vector $\boldsymbol{b}$ and the column vector $\boldsymbol{a}$'s second element (Figure B.2(b)) to compute the second row of the output matrix. This procedure continues until the last element of the column vector (Figure B.2(c)). Figure B.2(d) shows the result of the convolution.

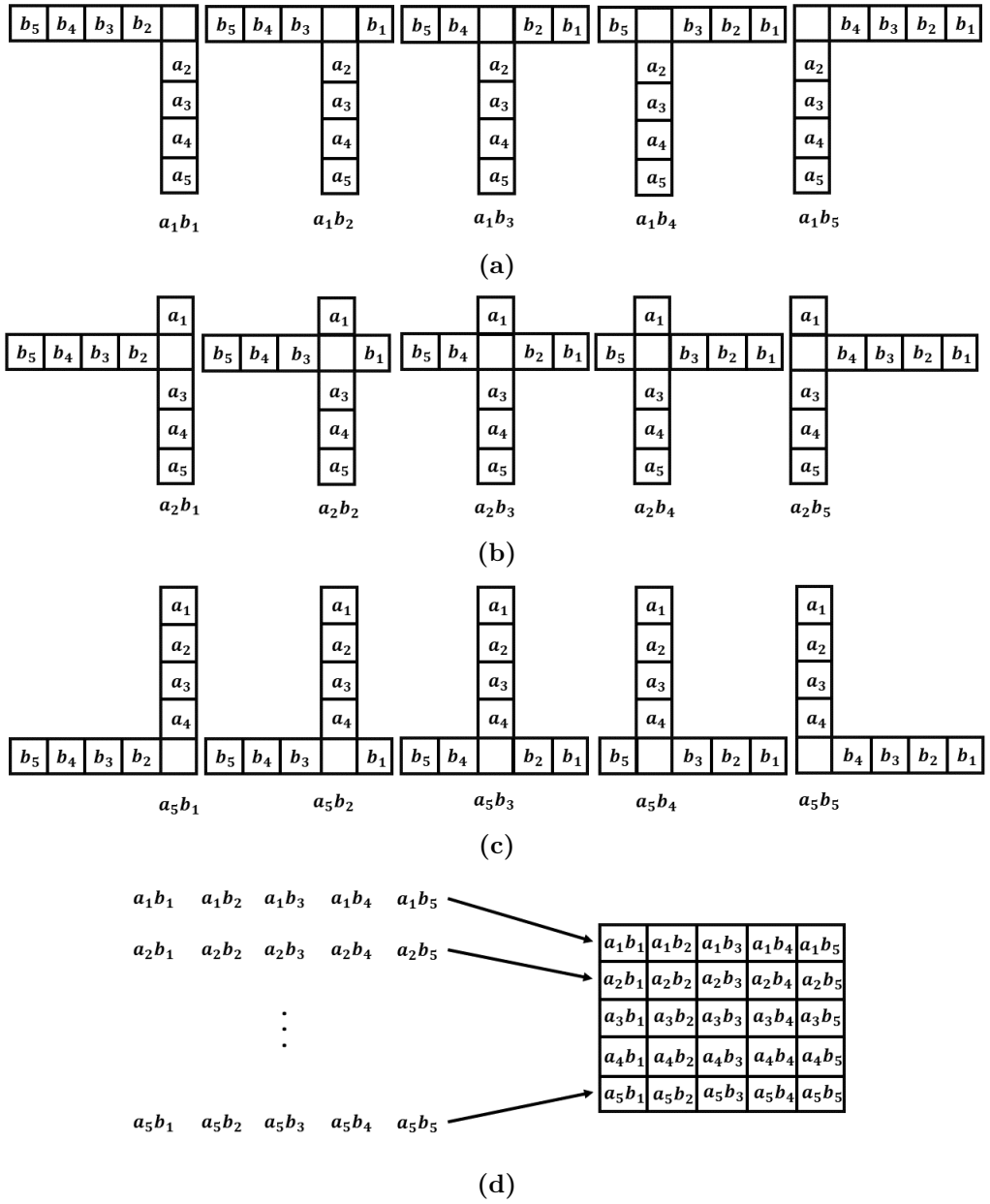


Figure B.2: *Convolution between column vector* $\boldsymbol{a}$ *and row vector* $\boldsymbol{b}$*. (a) The output of the first row of the convolution between column vector* $\boldsymbol{a}$ *and row vector* $\boldsymbol{b}$*. (b) The output of the second row of the convolution between column vector* $\boldsymbol{a}$ *and row vector* $\boldsymbol{b}$*. (c) The output of the fifth row of the convolution between column vector* $\boldsymbol{a}$ *and row vector* $\boldsymbol{b}$*. (d) The output of the convolution between* $\boldsymbol{a}$ *and* $\boldsymbol{b}$*.*

Appendix C

Computational Cost

When applying the convolution operation between a matrix of size $M \times N$ and a separable convolution kernel of size $m \times n$, the result will require approximately $M \times N \times m \times n$ multiplications and additions without using the kernel's separability. If the separability of the kernel is being applied—that is, the above kernel is being decomposed into a column vector of size $m \times 1$ and a row vector of size $1 \times n$ - then the calculation will have two steps. The first step is applying the convolution between the matrix and the column vector, and the corresponding multiplications and additions are about $M \times N \times m$. The second step will perform the convolution between the output result of the first step (the output size of the first step remains $M \times N$) and the row vector. The second step will have $M \times N \times n$ multiplications and additions. Thus, the total multiplications and additions of the calculation when applying the kernel's separability will be $MN(m+n)$. When $m = n$, the convolution kernel is square. By comparing the computational complexity between the direct and the separable convolution, the estimated computational advantage is

$$\frac{MNmn}{MN(m+n)} = \frac{mn}{(m+n)} = \frac{m}{2}. \tag{C.1}$$

This ratio means that the AGM Γ_1 is expected to be $\frac{m}{2}$ times faster than the CGM Γ_1 in each convolution operation between the velocity domain and each separable kernel.

Appendix D

Unnomalized AGM versus CGM ($ks = 33$)

Figure D.1 displays a set of Γ value intensity maps when applying the AGM on the simulation data shown in Figure 3.1, and the corresponding kernel size is set equal to 33. In Figure 3.3, panels in the left column correspond to the unnormalized AGM; panels in the right column correspond to the CGM. Figure D.2 displays the corresponding distributions of ratios between the unnormalized AGM and CGM when using $ks = 33$. The mean values of C_1 (shown in blue) and C_2 (shown in red) are 13.63 and 13.15, respectively. The variance of C_1 (shown in blue) and C_2 (shown in red) is 8.97 and 9.13, respectively.

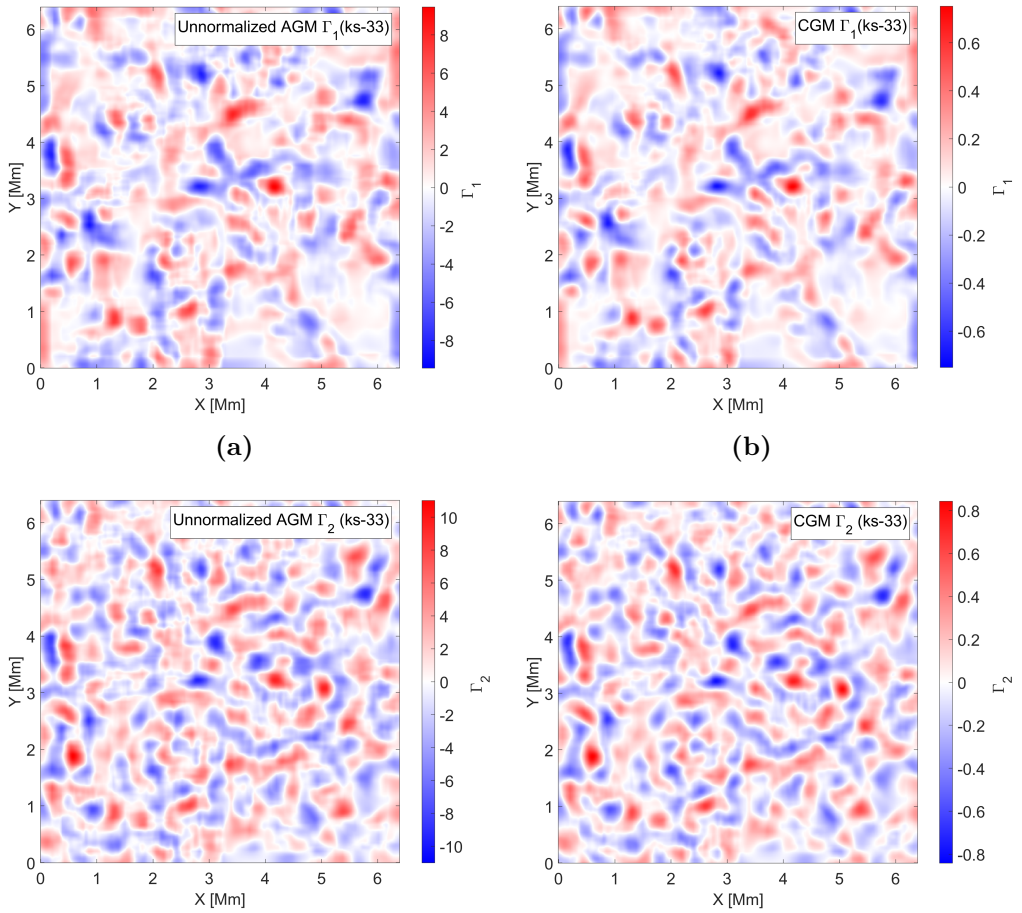


Figure D.1: *The distributions of the* Γ *functions values obtained by the AGM ((a) and (c)) and CGM ((b) and (d)) that were applied to the photospheric horizontal velocity field are shown in Figure 3.1. For both cases, a convolution kernel size* $ks = 33$ *was applied. As here, the values of* Z_1 *and* Z_2 *for the AGM were equal to 1, and the AGM was unnormalised.*

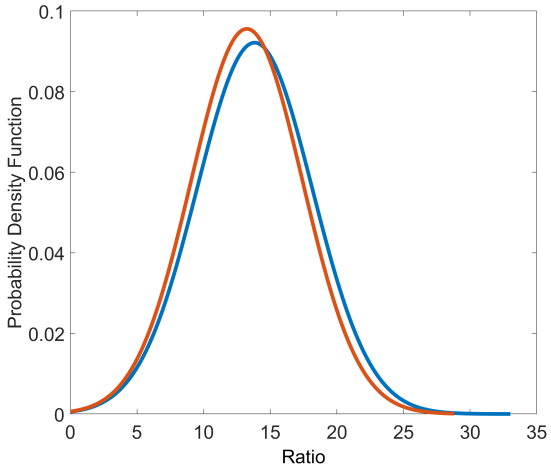


Figure D.2: *The distribution of the ratios between the unnormalized AGM (Figure D.1(a) and (c)) and CGM (Figure D.1 (b) and (d)) for* Γ_1 *(*C_1 *is shown in blue) and* Γ_2 *(*C_2 *is shown in red). For both cases the convolution kernel is equal to 33.*

Appendix E

The Calibration of the CSA and the WCSA

In this Appendix, we calibrate the CSA and WCSA using a vortex with a known boundary. Figure E.1 displays a horizontal velocity field with five axisymmetric vortices, each having a diameter of 13 pixels. The colormap represents vorticity. The blue contour and the blue cross denote the central vortex's boundary and center, respectively. The only difference among these vortices lies in their rotation—the central vortex is counterclockwise, and the four peripheral vortices are clockwise. The AGM employs various odd kernel sizes (ranging from $ks = 3$ to 31 pixels) to calculate the center and boundary of the central vortex.

Figure E.2 illustrates the corresponding Γ_1 values of the vortex center of the central vortex based on different kernel sizes. For an axisymmetric vortex, the Γ_1 value of the center is 1 (indicated by the blue line), and the black dashed line indicates the Γ_1 threshold (0.75). The red curve depicts the Γ_1 values calculated by the AGM for different kernel sizes. Initially, the Γ_1 value increases with the kernel size, peaking at 0.9906 with $ks = 13$ pixels. Subsequently, the Γ_1 value decreases with further increases in kernel size. Notably, all calculated Γ_1 values are below 1 because the square kernel domain is not axisymmetry in all directions.

Figure E.3 displays the Γ_1 value map for four different kernel sizes (5, 13, 21, 27) of the velocity field. In each panel, the black square indicates the kernel domain used to calculate the Γ_1 value at the vortex center. Panel (a) illustrates an example with the entire kernel domain within the vortex region. Panel (b) depicts a kernel domain (13 pixels $\times$ 13 pixels) comparable to the actual size of the vortex region. In this case, the calculated Γ_1 value is closest to 1. However, it remains below 1 because the square kernel domain introduces

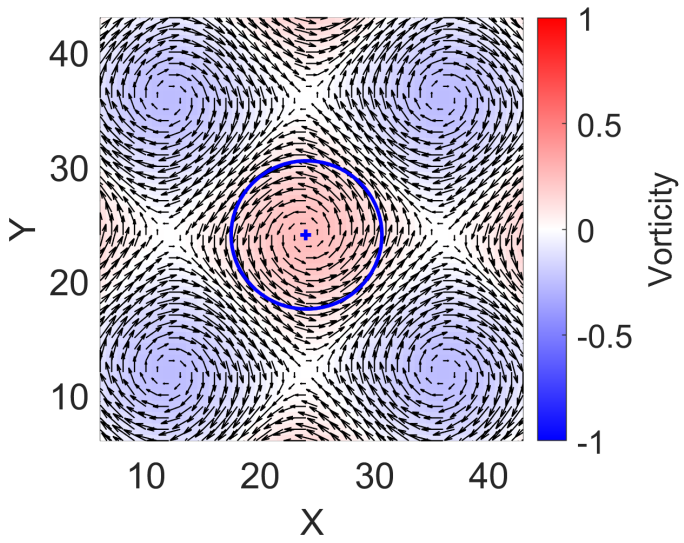


Figure E.1: *Horizontal velocity field (black arrows) of five axisymmetric vortices. The central vortex's center and boundary are indicated by the blue cross and blue contour, respectively.*

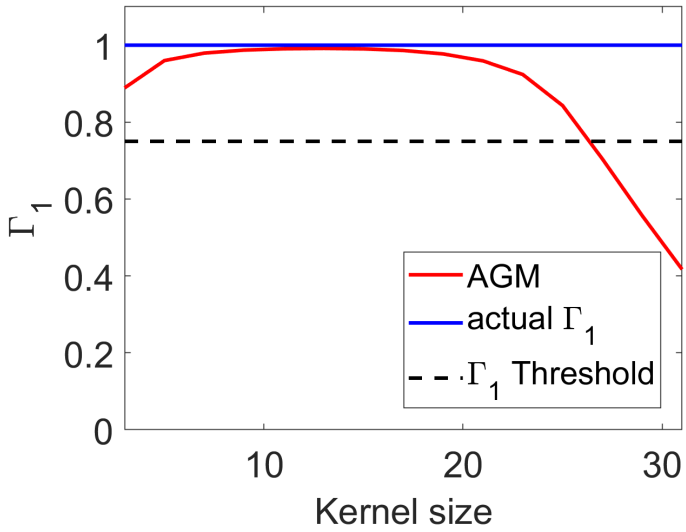


Figure E.2: Γ_1 *value at the center of the central vortex as a function of kernel sizes. The blue line indicates the axisymmetric vortex's actual* Γ_1 *value at the center. The red curve indicates the* Γ_1 *value calculated by the AGM with different kernel sizes. The black dashed line indicates the threshold of the* Γ_1 *function.*

velocity field points at the four corners, which are outside the vortex region. Panels (c) and (d) show results when the corresponding ks is larger than the vortex region, implying that the kernel domain includes peripheral velocity points (bias) outside the actual boundary during calculation. Consequently, these panels' Γ_1 value is less than 1. When the kernel domain is significantly larger than the vortex region (see panel (d)), the adaptive AGM may fail in detection ($ks = 27$ pixels and $\Gamma_1 = 0.704$). Therefore, the adaptive AGM's initial step for vortex detection involves determining a suitable kernel size for different vortices. This kernel size corresponds to the maximum Γ_1 value among different kernel sizes.

Figure E.4 presents the corresponding Γ_2 value map when setting $S_1 = S_2$ in the calculation. The white boundaries in each panel indicate the initial boundary calculated by the AGM with the corresponding kernel size, and the blue contour represents the actual boundary of the central vortex. Different kernel sizes yield varying initial boundaries. In panel (a), the initial boundary is identical for the central vortex and the four spherical vortices. The reason is that all kernel domains are within the vortex region; therefore, no bias is introduced in calculating the initial boundary. However, there is a substantial difference between the initial and actual boundaries. In comparison, the initial boundaries in panel (b) are larger than those in panel (a). The difference between the initial shown in this panel and the actual boundaries is reduced. Initial boundaries shown in panels (c) and (d) have a larger difference from the actual boundary compared to the initial boundaries shown in panels (b). Additionally, the initial boundaries of the four peripheral and central vortices are not identical in these two panels. This discrepancy arises from the influence of the velocity field outside the vortex region on calculating the initial boundary. In other words, errors are introduced in the calculation. The original Γ method does not consider this circumstance, introducing bias and resulting in an inaccurate vortex region calculation. The CSA and WCSA address these issues using the customized S_2 region to calculate Γ_2.

It is necessary to calibrate the CSA and WCSA before applying them to vortex detection. The calibration involves applying the AGM with a proper kernel size (e.g., $ks = 13$ pixels) to detect the axisymmetric vortex (e.g., the diameter is 13 pixels) and adjusting CSA and WCSA parameters (such as the customized S_2 region) until the AGM produces a boundary close to the actual boundary. In Figure E.5, the green contour indicates the boundary detected by the AGM with the assistance of CSA and WCSA after calibration. Compared to the initial boundary (white), there is less difference between the green and blue contours.

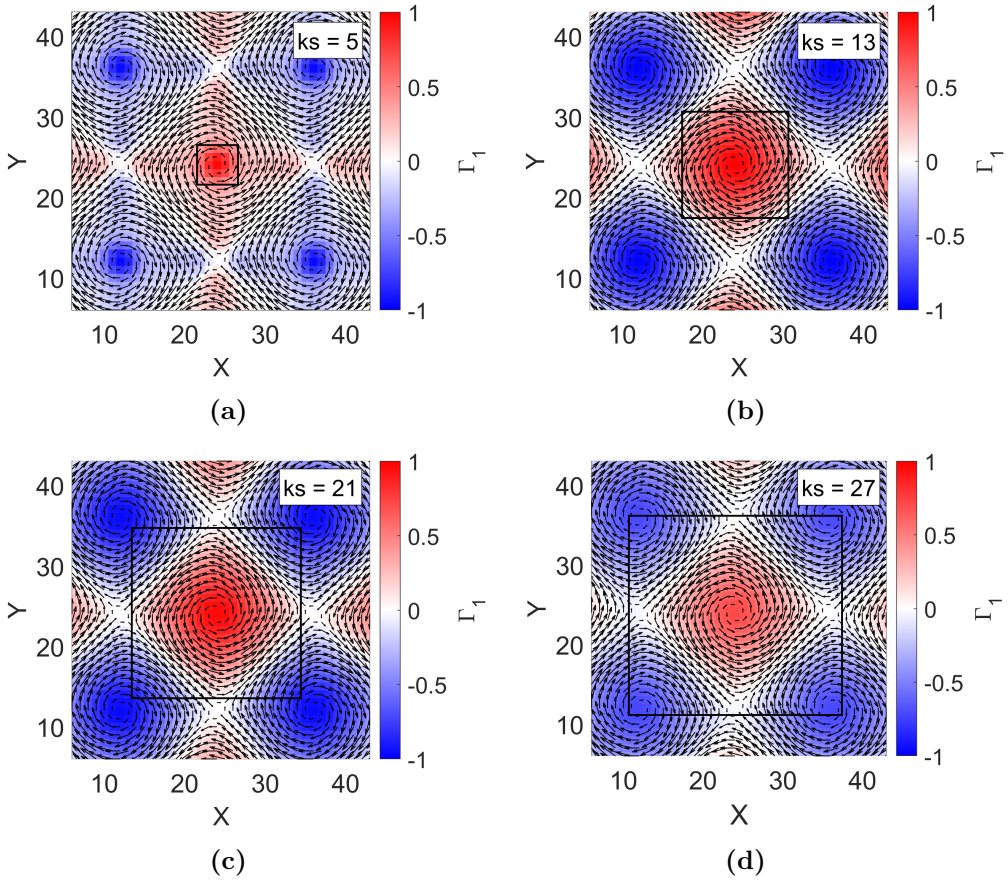


Figure E.3: *The distributions of the* Γ_1 *functions values obtained by the AGM (with* $ks = 5$*, 13, 21, and 27) that were applied to the horizontal velocity field shown in Figure E.1. The black square in each panel indicates the kernel domain that is used to calculate the* Γ_1 *value at the center of the central vortex. (a)* $ks = 5$*; (b)* $ks = 13$*; (c)* $ks = 21$*; (d)* $ks = 27$*.*

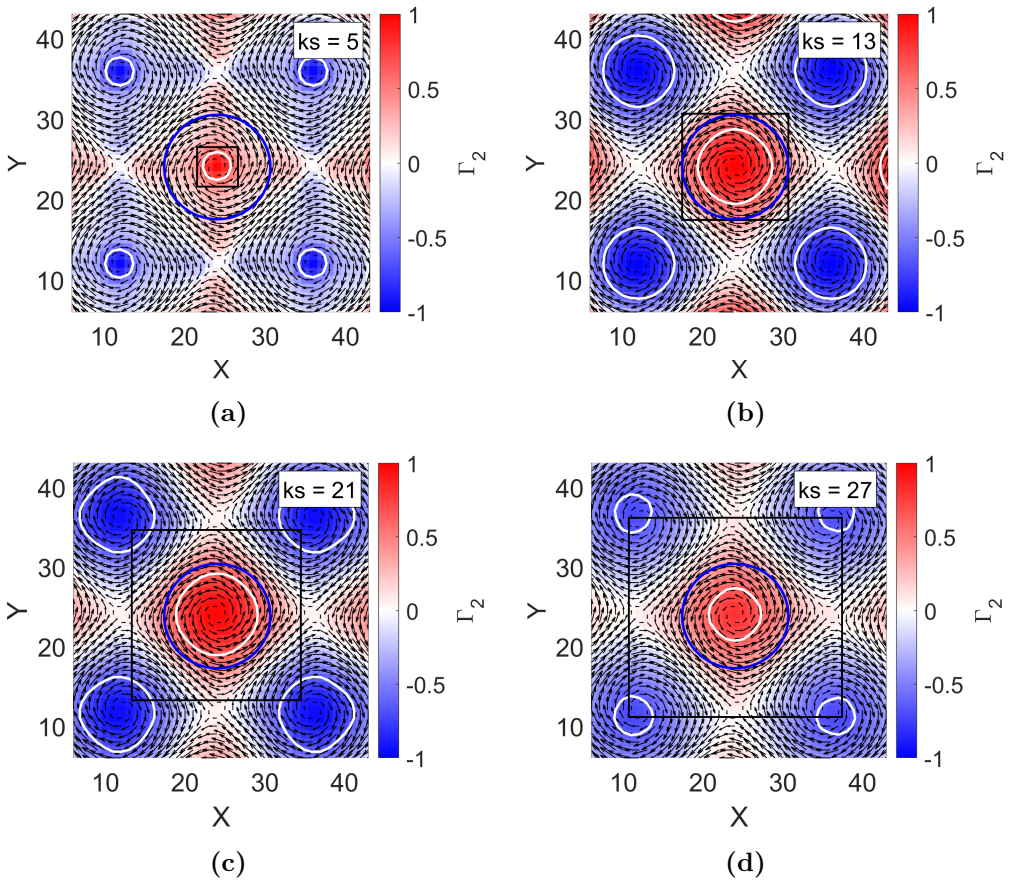


Figure E.4: *The distributions of the Γ_2 functions values obtained by the AGM (with $ks = 5$, 13, 21, and 27) that were applied to the horizontal velocity field shown in Figure E.1. White contours indicate the initial boundaries of each vortex. The blue contour in each panel indicates the actual boundary of the central vortex. The black square in each panel indicates the kernel domain that is used to calculate the initial boundary of the central vortex. (a) $ks = 5$; (b) $ks = 13$; (c) $ks = 21$; (d) $ks = 27$.*

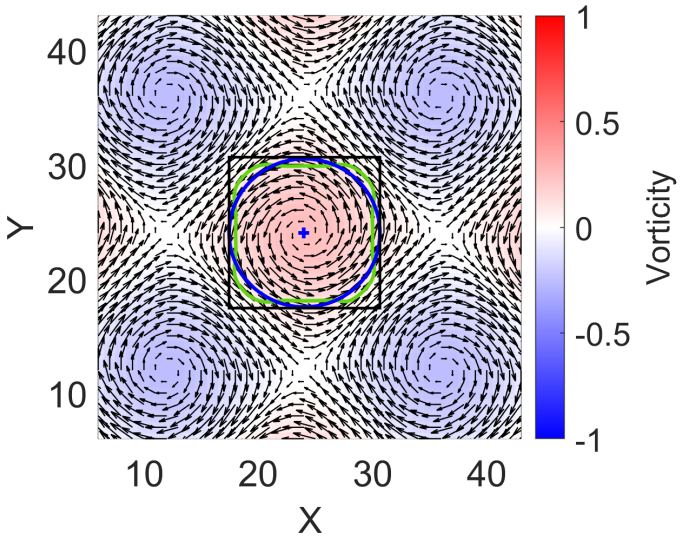


Figure E.5: *The black arrows indicate the horizontal velocity field of five axisymmetric vortices. The center and boundary of the central vortex are indicated by the blue cross and blue contour, respectively. The AGM calculates the green contour using the CSA and the WCSA. The black square indicates the kernel domains that are used to calculate the* Γ_1 *value at the vortex center.*

Appendix F

The Algorithm dF

Figure F.1 is the pseudocode of the algorithm **dF**. Here, $c(p, q)$ refers to calculating the Euclidean distance between point p and q; $ca(p, q)$ is a $p \times q$ matrix, denoting the calculation of the coupling length $\| L \|$ for each distinct pair in P and Q. Each element of it is initialized to -1 before the calculation. The discrete Fréchet distance is equal to $ca(p, q)$ through the recursive operation.

```
Function dF(P, Q): real;
  input: polygonal curves P = (u_1, ..., u_p) and Q = (v_1, ..., v_q).
  return: δ_dF(P, Q)
  ca: array [1..p, 1..q] of real;
  function c(i, j): real;
    begin
      if c(i, j) ≻ −1 then return ca(i, j)
      elseif i = 1 and j = 1 then ca(i, j) := d(u_1, v_1)
      elseif i ≻ 1 and j = 1 then ca(i, j) := max{c(i − 1, j), d(u_i, v_1)}
      elseif i = 1 and j ≻ 1 then ca(i, j) := max{c(1, j − 1), d(u_1, v_j)}
      elseif i ≻ 1 and j ≻ 1 then ca(i, j) := max{min(c(i − 1, j), c(i − 1, j − 1), c(i, j − 1)), d(u_i, v_j)}
      else ca(i, j) = ∞
    return ca(i, j);
    end; /* function c*/
```

Figure F.1: *The algorithm* ***dF*** *(Eiter and Mannila, 1994).*

Appendix G

Computing the Discrete Fréchet Distance with the use of the Algorithm dF

This Section provides an example of calculating the discrete Fréchet distance using the algorithm **dF**. For illustrative purposes, two 2D curves are considered. In Figure G.1, the polygonal curve P is shown in red and has four endpoints, while the polygonal curve Q is in blue and has three endpoints. The coordinates of each point for the two curves are provided below:

$$P : [p_1(1,1), p_2(2,1), p_3(3,2), p_4(4,1)],$$
$$Q : [q_1(1,0), q_2(2,0), q_3(3,0)].$$

The following steps outline how to calculate the discrete Fr'echet distance between curves P and Q:

1. Calculate the Euclidean length of all distinct coupling pairs between P and Q. In Figure G.2, $d_(p_i, q_j)$ denotes the length between point p_i and q_j;

2. Calculate the coupling length $\parallel L \parallel$ by applying the algorithm **dF**. In Figure G.3, $D(p_i, q_j)$ denotes the coupling length between point p_i and q_j;

3. Calculate the discrete Fréchet distance after considering all endpoints in P and Q. Thus,

$$\delta_{dF}(P, Q) = \mathbf{dF}(P, Q) = D(p_4, q_3) = 2.$$

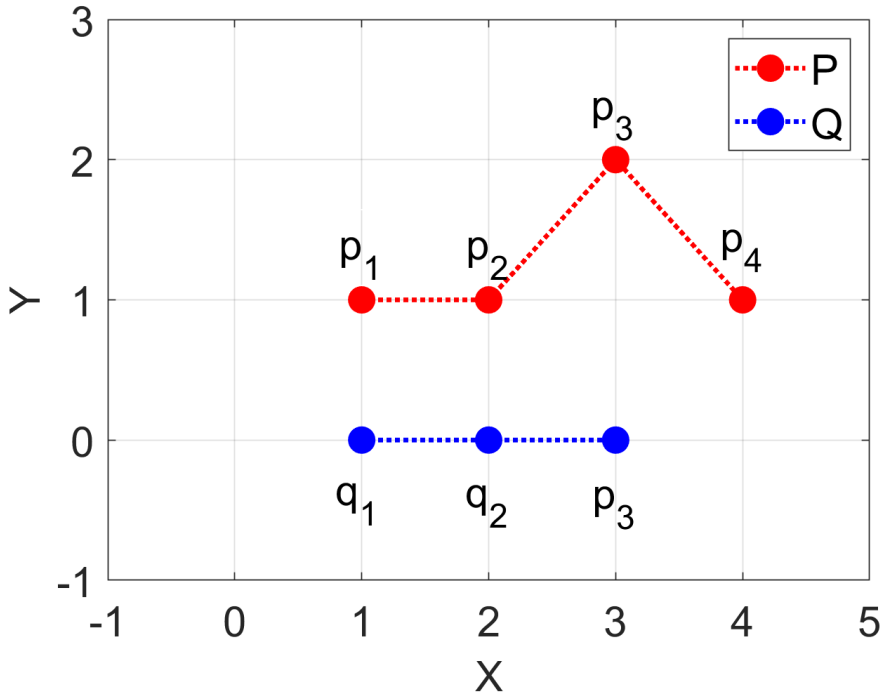


Figure G.1: *Two discrete polygonal curves: P (shown in red) and Q (shown in blue).*

$d_{p,q} = \|p_i - q_j\|$	q_1	q_2	q_3
p_1	$d_{p_1,q_1} = 1$	$d_{p_1,q_2} = 1.41$	$d_{p_1,q_3} = 2.23$
p_2	$d_{p_2,q_1} = 1.41$	$d_{p_2,q_2} = 1$	$d_{p_2,q_3} = 1.41$
p_3	$d_{p_3,q_1} = 2.82$	$d_{p_3,q_2} = 2.23$	$d_{p_3,q_3} = 2$
p_4	$d_{p_4,q_1} = 3.16$	$d_{p_4,q_2} = 2.23$	$d_{p_4,q_3} = 1.41$

Figure G.2: *Euclidean length of all distinct coupling pairs between P and Q.*

$$D_{p_1,q_1} = max\{d_{p_1,q_1}\} = d_{p_1,q_1}$$
$$D_{p_1,q_2} = max\left\{d_{p_1,q_2}, min\{D_{p_1,q_1}\}\right\} = d_{p_1,q_2}$$
$$D_{p_1,q_3} = max\left\{d_{p_1,q_3}, min\{D_{p_1,q_2}\}\right\} = d_{p_1,q_3}$$
$$D_{p_2,q_1} = max\left\{d_{p_2,q_1}, min\{D_{p_1,q_1}\}\right\} = d_{p_2,q_1}$$
$$D_{p_2,q_2} = max\left\{d_{p_2,q_2}, min\{D_{p_1,q_2}, D_{p_2,q_1}, D_{p_1,q_1}\}\right\} = d_{p_2,q_2} = d_{p_1,q_1}$$
$$D_{p_2,q_3} = max\left\{d_{p_2,q_3}, min\{D_{p_1,q_3}, D_{p_2,q_2}, D_{p_1,q_2}\}\right\} = d_{p_2,q_3}$$
$$D_{p_3,q_1} = max\left\{d_{p_3,q_1}, min\{D_{p_2,q_1}\}\right\} = d_{p_3,q_1}$$
$$D_{p_3,q_2} = max\left\{d_{p_3,q_2}, min\{D_{p_2,q_2}, D_{p_3,q_1}, D_{p_2,q_1}\}\right\} = d_{p_3,q_2}$$
$$D_{p_3,q_3} = max\left\{d_{p_3,q_3}, min\{D_{p_2,q_3}, D_{p_3,q_2}, D_{p_2,q_2}\}\right\} = d_{p_3,q_3}$$
$$D_{p_4,q_1} = max\left\{d_{p_4,q_1}, min\{D_{p_3,q_1}\}\right\} = d_{p_4,q_1} = D_{p_3,q_1} = d_{p_3,q_1}$$
$$D_{p_4,q_2} = max\left\{d_{p_4,q_2}, min\{D_{p_3,q_2}, D_{p_4,q_1}, D_{p_3,q_1}\}\right\} = d_{p_4,q_2} = D_{p_3,q_2} = d_{p_3,q_2}$$
$$D_{p_4,q_3} = max\left\{d_{p_4,q_3}, min\{D_{p_3,q_3}, D_{p_4,q_2}, D_{p_3,q_2}\}\right\} = D_{p_3,q_3} = d_{p_3,q_3}$$

(a)

$D_{p,q} = (p_i, q_j)$	q_1	q_2	q_3
p_1	$\boldsymbol{D}_{p_1,q_1} = \boldsymbol{d}_{p_1,q_1} = \mathbf{1}$	$D_{p_1,q_2} = d_{p_1,q_2} = 1.41$	$D_{p_1,q_3} = d_{p_1,3q_3} = 2.23$
p_2	$D_{p_2,q_1} = d_{p_2,q_1} = 1.41$	$\boldsymbol{D}_{p_2,q_2} = \boldsymbol{d}_{p_2,q_2} = \boldsymbol{d}_{p_1,q_1} = \mathbf{1}$	$D_{p_2,q_3} = d_{p_2,q_3} = 1.41$
p_3	$D_{p_3,q_1} = d_{3,q_1} = 2.82$	$D_{p_3,q_2} = d_{p_3,q_2} = 2.23$	$\boldsymbol{D}_{p_3,q_3} = \boldsymbol{d}_{p_3,q_3} = \mathbf{2}$
p_4	$D_{p_4,q_1} = d_{p_4,q_1} = d_{p_3,q_1} = 3.16$	$D_{p_4,q_2} = d_{p_4,q_2} = d_{p_3,q_2} = 2.23$	$\boldsymbol{D}_{p_4,q_3} = \boldsymbol{D}_{p_3,q_3} = \boldsymbol{d}_{p_3,q_3} = \mathbf{2}$

(b)

Figure G.3: *Coupling length $\| L \|$ of each distinct pairs in P and Q.*

In algorithm **dF**, Equation (G.1) shown below is used to calculate the coupling length $\| L \|$ of the current location.

$$D_{p,q} = max\left\{d_{p,q}, min\left\{D_{p-1,q}, D_{p,q-1}, D_{p-1,q-1}\right\}\right\}. \tag{G.1}$$

This operation described suggests that the coupling length of the current endpoint in P and Q is determined by taking the maximum value between the length for the current location and the minimum value of all neighboring coupling lengths prior to the current location. Here, $D_{p-1,q}$, $D_{p,q-1}$, and $D_{p-1,q-1}$ represent the coupling lengths of the current location's neighbors, specifically coming from the three neighbor points: $(p-1,q), (p,q-1), (p-1,q-1)$, respectively.

Bibliography

Aljohani, Y., Fedun, V., Ballai, I., Silva, S. S. A., Shelyag, S., and Verth, G. (2022). New Approach for Analyzing Dynamical Processes on the Surface of Photospheric Vortex Tubes. , 928(1):3.

Attie, R., Innes, D. E., and Potts, H. E. (2009). Evidence of photospheric vortex flows at supergranular junctions observed by FG/SOT (Hinode). , 493(2):L13–L16.

Avrett, E. H. and Loeser, R. (2008). Models of the Solar Chromosphere and Transition Region from SUMER and HRTS Observations: Formation of the Extreme-Ultraviolet Spectrum of Hydrogen, Carbon, and Oxygen. , 175(1):229–276.

Balmaceda, L., Vargas Domínguez, S., Palacios, J., Cabello, I., and Domingo, V. (2010). Evidence of small-scale magnetic concentrations dragged by vortex motion of solar photospheric plasma. , 513:L6.

Barthol, P., Gandorfer, A., Solanki, S. K., and Schüssler, M., e. a. (2011). The Sunrise Mission. , 268(1):1–34.

Beckers, J. M. (1972). Solar Spicules. , 10:73.

Berkefeld, T., Schmidt, W., Soltau, D., Bell, A., and Doerr, H. P., e. a. (2011). The Wave-Front Correction System for the Sunrise Balloon-Borne Solar Observatory. , 268(1):103–123.

Bingert, S., Zacharias, P., Peter, H., and Gudiksen, B. (2008). On the Nature of Coronal Loops. In *European Solar Physics Meeting*, volume 12 of *European Solar Physics Meeting*, page 3.29.

Boehm-Vitense, E. (1984). Chromospheres, Transition Regions, and Coronas. *Science*, 223(4638):777–784.

Bonet, J. A., Márquez, I., and Sánchez Almeida, J., e. a. (2010). SUNRISE/IMaX Observations of Convectively Driven Vortex Flows in the Sun. , 723(2):L139–L143.

Bonet, J. A., Márquez, I., Sánchez Almeida, J., Cabello, I., and Domingo, V. (2008). Convectively Driven Vortex Flows in the Sun. , 687(2):L131.

Bourdin, P.-A., Bingert, S., and Peter, H. (2014). Coronal loops above an active region: Observation versus model. , 66:S7.

Brandt, P. N., Scharmer, G. B., Ferguson, S., Shine, R. A., Tarbell, T. D., and Title, A. M. (1988). Vortex flow in the solar photosphere. , 335(6187):238–240.

Bray, R. J., Loughhead, R. E., and Tappere, E. J. (1974). High-Resolution Photography of the Solar Chromosphere. XV: Preliminary Observations in Fe I λ6569.2. , 39(2):323–326.

Burton, W. M., Jordan, C., Ridgeley, A., and Wilson, R. (1971). The Structure of the Chromosphere-Corona Transition Region from Limb and Disk Intensities. *Philosophical Transactions of the Royal Society of London Series A*, 270(1202):81–98.

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3:1–27.

Canivete Cuissa, J. R. and Steiner, O. (2020). Vortices evolution in the solar atmosphere. A dynamical equation for the swirling strength. , 639:A118.

Chian, A. C. L., Silva, S. S. A., Rempel, E. L., Bellot Rubio, L. R., Gošić, M., Kusano, K., and Park, S.-H. (2020). Lagrangian chaotic saddles and objective vortices in solar plasmas. , 102(6):060201.

Chian, A. C. L., Silva, S. S. A., Rempel, E. L., Gošić, , M., Bellot Rubio, L. R., Kusano, K., Miranda, R. A., and Requerey, I. S. (2019). Supergranular turbulence in the quiet Sun: Lagrangian coherent structures. , 488(3):3076–3088.

Chong, M. S., Perry, A. E., and Cantwell, B. J. (1990). A general classification of three-dimensional flow fields. *Physics of Fluids A: Fluid Dynamics*, 2(5):765–777.

Cucitore, R., Quadrio, M., and Baron, A. (1999). On the effectiveness and limitations of local criteria for the identification of a vortex. *European Journal of Mechanics - B/Fluids*, 18(2):261–282.

Dakanalis, I., Tsiropoula, G., Tziotziou, K., and Kontogiannis, I. (2022). Chromospheric swirls. I. Automated detection in Hα observations and their statistical properties. , 663:A94.

Dakanalis, I., Tsiropoula, G., Tziotziou, K., and Koutroumbas, K. (2021). Automated Detection of Chromospheric Swirls Based on Their Morphological Characteristics. , 296(1):17.

De Pontieu, B., Carlsson, M., Rouppe van der Voort, L. H. M., Rutten, R. J., Hansteen, V. H., and Watanabe, H. (2012). Ubiquitous Torsional Motions in Type II Spicules. , 752(1):L12.

de Pontieu, B., McIntosh, S., Hansteen, V. H., and Carlsson, Mats., e. a. (2007). A Tale of Two Spicules: The Impact of Spicules on the Magnetic Chromosphere. , 59:S655.

Eiter, T. and Mannila, H. (1994). Computing discrete fréchet distance.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. 96(34):226–231.

Fedun, V., Shelyag, S., Verth, G., Mathioudakis, M., and Erdélyi, R. (2011). MHD waves generated by high-frequency photospheric vortex motions. *Annales Geophysicae*, 29(6):1029–1035.

Fréchet, M. (1906). Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, 22:1–72.

Gandorfer, A., Grauf, B., Barthol, P., and Riethmüller, T. L., e. a. (2011). The Filter Imager SuFI and the Image Stabilization and Light Distribution System ISLiD of the Sunrise Balloon-Borne Observatory: Instrument Description. , 268(1):35–55.

Gary, S. P., Skoug, R. M., Steinberg, J. T., and Smith, C. W. (2001). Proton temperature anisotropy constraint in the solar wind: ACE observations. , 28(14):2759–2762.

Giagkiozis, I., Fedun, V., Scullion, E., Jess, D. B., and Verth, G. (2018). Vortex Flows in the Solar Atmosphere: Automated Identification and Statistical Analysis. , 869(2):169.

Gonzalez, R. C., Woods, R. E., and Eddins, S. L. (2004). *Digital Image Processing Using Matlab^{tm}*.

Graftieaux, L., Michard, M., and Grosjean, N. (2001). Combining piv, pod and vortex identification algorithms for the study of unsteady turbulent swirling flows. *Measurement Science and technology*, 12(9):1422.

Günther, T., Gross, M., and Theisel, H. (2017). Generic objective vortices for flow visualization. *ACM Transactions on Graphics*, 36:141:1–141:11.

Günther, T. and Theisel, H. (2018). The state of the art in vortex extraction. *Computer Graphics Forum*, 37(6):149–173.

Haller, G. (2015). Lagrangian coherent structures. *Annual Review of Fluid Mechanics*, 47(1):137–162.

Haller, G., Hadjighasem, A., Farazmand, M., and Huhn, F. (2016). Defining coherent vortices objectively from the vorticity. *Journal of Fluid Mechanics*, 795:136–173.

Helman, J. and Hesselink, L. (1989). Representation and display of vector field topology in fluid flow data sets. *Computer*, 22(8):27–36.

Helman, J. and Hesselink, L. (1991). Visualizing vector field topology in fluid flows. *IEEE Computer Graphics and Applications*, 11(3):36–46.

Hunt, J. C., Wray, A. A., and Moin, P. (1988). Eddies, streams, and convergence zones in turbulent flows. *Studying turbulence using numerical simulation databases, 2. Proceedings of the 1988 summer program.*

Hunt, J. C. R. (1987). Vorticity and vortex dynamics in complex turbulent flows. In *Canadian Society for Mechanical Engineering*, volume 11, pages 21–35.

Ichimoto, K., Lites, B., Elmore, D., Suematsu, Y., and Tsuneta, S., e. a. (2008). Polarization Calibration of the Solar Optical Telescope onboard Hinode. , 249(2):233–261.

Iijima, H. and Yokoyama, T. (2017). A Three-dimensional Magnetohydrodynamic Simulation of the Formation of Solar Chromospheric Jets with Twisted Magnetic Field Lines. , 848(1):38.

Jacoutot, L., Kosovichev, A. G., Wray, A. A., and Mansour, N. N. (2008). Numerical simulation of excitation of solar oscillation modes for different turbulent models. *The Astrophysical Journal*, 682(2):1386.

Jeong, J. and Hussain, F. (1995). On the identification of a vortex. *Journal of Fluid Mechanics*, 285:69–94.

Kato, Y. and Wedemeyer, S. (2017). Vortex flows in the solar chromosphere. I. Automatic detection method. , 601:A135.

Kerr, R. A. (2012). Why is the sun's corona so hot?

Kitiashvili, I. N., Abramenko, V. I., Goode, P. R., Kosovichev, A. G., Lele, S. K., Mansour, N. N., Wray, A. A., and Yurchyshyn, V. B. (2012a). Turbulent Kinetic Energy Spectra of Solar Convection from NST Observations and Realistic MHD Simulations. *arXiv e-prints*, page arXiv:1206.5300.

Kitiashvili, I. N., Kosovichev, A. G., Lele, S. K., Mansour, N. N., and Wray, A. A. (2013). Ubiquitous Solar Eruptions Driven by Magnetized Vortex Tubes. , 770(1):37.

Kitiashvili, I. N., Kosovichev, A. G., Mansour, N. N., Lele, S. K., and Wray, A. A. (2012b). Vortex tubes of turbulent solar convection. , 86(1):018403.

Kitiashvili, I. N., Kosovichev, A. G., Mansour, N. N., and Wray, A. A. (2011). Excitation of Acoustic Waves by Vortices in the Quiet Sun. , 727(2):L50.

Kitiashvili, I. N., Kosovichev, A. G., Mansour, N. N., and Wray, A. A. (2012c). Dynamics of Magnetized Vortex Tubes in the Solar Chromosphere. , 751(1):L21.

Kołár, V. (2009). Compressibility Effect in Vortex Identification. *AIAA Journal*, 47(2):473–475.

Kosugi, T., Matsuzaki, K., Sakao, T., and Shimizu, T., e. a. (2007). The Hinode (Solar-B) Mission: An Overview. , 243(1):3–17.

Liu, J., Nelson, C. J., and Erdélyi, R. (2019). Automated Swirl Detection Algorithm (ASDA) and Its Application to Simulation and Observational Data. , 872(1):22.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.

Lugt, H. J. (1979). The dilemma of defining a vortex. In *Recent Developments in Theoretical and Experimental Fluid Mechanics: Compressible and Incompressible Flows*, pages 309–321.

Martínez Pillet, V., Del Toro Iniesta, J. C., Álvarez-Herrero, A., Domingo, V., and Bonet, J. A., e. a. (2011). The Imaging Magnetograph eXperiment (IMaX) for the Sunrise Balloon-Borne Solar Observatory. , 268(1):57–102.

Menelas, B., Ammi, M., Pastur, L., and Bourdot, P. (2009). Haptical exploration of an unsteady flow. pages 232–237.

Moll, R., Cameron, R. H., and Schüssler, M. (2011). Vortices in simulations of solar surface convection. *Astronomy Astrophysics*, 533:A126.

Mumford, S. J. and Erdélyi, R. (2015). Photospheric logarithmic velocity spirals as MHD wave generation mechanisms. , 449(2):1679–1685.

Mumford, S. J., Fedun, V., and Erdélyi, R. (2015). Generation of Magnetohydrodynamic Waves in Low Solar Atmospheric Flux Tubes by Photospheric Motions. , 799(1):6.

Nordlund, A. (1985). Solar Convection. , 100:209.

Nordlund, Å., Stein, R. F., and Asplund, M. (2009). Solar Surface Convection. *Living Reviews in Solar Physics*, 6(1):2.

Okubo, A. (1970). Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences. *Deep Sea Research A*, 17(3):445–454.

Park, S. H., Tsiropoula, G., Kontogiannis, I., Tziotziou, K., Scullion, E., and Doyle, J. G. (2016). First simultaneous SST/CRISP and IRIS observations of a small-scale quiet Sun vortex. , 586:A25.

Porfir'eva, G. A. and Yakunina, G. V. (2016). Chromospheric spicules and solar corona. *Astronomical and Astrophysical Transactions*, 29(4):567–580.

Priest, E. (2014). *Magnetohydrodynamics of the Sun*.

Quintero Noda, C., Schlichenmaier, R., and Bellot Rubio, L. R., e. a. (2022). The European Solar Telescope. , 666:A21.

Rappazzo, A. F., Velli, M., Dahlburg, R. B., and Einaudi, G. (2019). Magnetic Field Line Twisting by Photospheric Vortices: Energy Storage and Release. , 883(2):148.

Rappazzo, A. F. and Velli, M. M. (2010). Coronal Loops Dynamics and Photospheric Forcing Patterns. In *AGU Fall Meeting Abstracts*, volume 2010, pages SM51C–1846.

Rast, M. P., Bello González, N., and Bellot Rubio, Luis, e. a. (2021). Critical Science Plan for the Daniel K. Inouye Solar Telescope (DKIST). , 296(4):70.

Requerey, I. S., Cobo, B. R., Gošić, M., and Bellot Rubio, L. R. (2018). Persistent magnetic vortex flow at a supergranular vertex. , 610:A84.

Riethmüller, T. L., Solanki, S. K., Berdyugina, S. V., Schüssler, M., Martínez Pillet, V., Feller, A., Gandorfer, A., and Hirzberger, J. (2014). Comparison of solar photospheric bright points between Sunrise observations and MHD simulations. , 568:A13.

Robinson, S. K. (1991). Coherent motions in the turbulent boundary layer. *Annual Review of Fluid Mechanics*, 23:601–639.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Scharmer, G. B. (2006). Comments on the optimization of high resolution Fabry-Pérot filtergraphs. , 447(3):1111–1120.

Scharmer, G. B., Bjelksjo, K., Korhonen, T. K., Lindberg, B., and Petterson, B. (2003a). The 1-meter Swedish solar telescope. In Keil, S. L. and Avakyan, S. V., editors, *Innovative Telescopes and Instrumentation for Solar Astrophysics*, volume 4853 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 341–350.

Scharmer, G. B., Bjelksjo, K., Korhonen, T. K., Lindberg, B., and Petterson, B. (2003b). The 1-meter Swedish solar telescope. In Keil, S. L. and Avakyan, S. V., editors, *Innovative Telescopes and Instrumentation for Solar Astrophysics*, volume 4853 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 341–350.

Scharmer, G. B., Narayan, G., Hillberg, T., de la Cruz Rodriguez, J., Löfdahl, M. G., Kiselman, D., Sütterlin, P., van Noort, M., and Lagg, A.

(2008a). CRISP Spectropolarimetric Imaging of Penumbral Fine Structure. , 689(1):L69.

Scharmer, G. B., Narayan, G., Hillberg, T., de la Cruz Rodriguez, J., Löfdahl, M. G., Kiselman, D., Sütterlin, P., van Noort, M., and Lagg, A. (2008b). CRISP Spectropolarimetric Imaging of Penumbral Fine Structure. , 689(1):L69.

Scheuermann, G., Kruger, H., Menzel, M., and Rockwood, A. (1998). Visualizing nonlinear vector field topology. *IEEE Transactions on Visualization and Computer Graphics*, 4(2):109–116.

Schüessler, M. (1984). The interchange instability of small flux tubes. , 140(2):453–458.

Scullion, E., Morgan, H., Lin, H., Fedun, V., and Morton, R. (2022). SULIS: A coronal magnetism explorer for ESA's Voyage 2050. *Experimental Astronomy*.

Shelyag, S., Cally, P. S., Reid, A., and Mathioudakis, M. (2013). Alfvén Waves in Simulations of Solar Photospheric Vortices. , 776(1):L4.

Shelyag, S., Fedun, V., Erdélyi, R., Keenan, F. P., and Mathioudakis, M. (2012). Vortices in the Solar Photosphere. In *Second ATST-EAST Meeting: Magnetic Fields from the Photosphere to the Corona.*, volume 463 of *Astronomical Society of the Pacific Conference Series*, page 107.

Shetye, J., Verwichte, E., Stangalini, M., Judge, P. G., Doyle, J. G., Arber, T., Scullion, E., and Wedemeyer, S. (2019). Multiwavelength High-resolution Observations of Chromospheric Swirls in the Quiet Sun. , 881(1):83.

Silva, S. S. A., Fedun, V., Verth, G., Rempel, E. L., and Shelyag, S. (2020). Solar Vortex Tubes: Vortex Dynamics in the Solar Atmosphere. , 898(2):137.

Silva, S. S. A., Rempel, E. L., Pinheiro Gomes, T. F., Requerey, I. S., and Chian, A. C. L. (2018). Objective Lagrangian Vortex Detection in the Solar Photosphere. , 863(1):L2.

Silva, S. S. A., Verth, G., Rempel, E. L., Shelyag, S., Schiavo, L. A. C. A., and Fedun, V. (2021). Solar Vortex Tubes. II. On the Origin of Magnetic Vortices. , 915(1):24.

Skirvin, S., Verth, G., Juan González-Avilés, J., Shelyag, S., Sharma, R., Guzmán, F., Ballai, I., Scullion, E., Silva, S. S. A., and Fedun, V. (2022). Small-scale solar jet formation and their associated waves and instabilities. *arXiv e-prints*, page arXiv:2205.09598.

Snow, B., Fedun, V., Gent, F. A., Verth, G., and Erdélyi, R. (2018). Magnetic Shocks and Substructures Excited by Torsional Alfvén Wave Interactions in Merging Expanding Flux Tubes. , 857(2):125.

Solanki, S. K., Barthol, P., Danilovic, S., Feller, A., and Gandorfer, A., e. a. (2010). SUNRISE: Instrument, Mission, Data, and First Results. , 723(2):L127–L133.

Steiner, O., Franz, M., Bello González, N., Nutto, C., and Rezaei, R., e. a. (2010). Detection of Vortex Tubes in Solar Granulation from Observations with SUNRISE. , 723(2):L180–L184.

Tavabi, E., Koutchmy, S., and Ajabshirizadeh, A. (2012). solar spicules and jets. In Faurobert, M., Fang, C., and Corbard, T., editors, *EAS Publications Series*, volume 55 of *EAS Publications Series*, pages 71–78.

Truesdell, C. and Noll, W. (1965). *The Non-Linear Field Theories of Mechanics*, pages 1–541. Springer Berlin Heidelberg, Berlin, Heidelberg.

Tziotziou, K., Scullion, E., Shelyag, S., Steiner, O., Khomenko, E., Tsiropoula, G., Canivete Cuissa, J. R., Wedemeyer, S., Kontogiannis, I., Yadav, N., Kitiashvili, I. N., Skirvin, S. J., Dakanalis, I., Kosovichev, A. G., and Fedun, V. (2023). Vortex Motions in the Solar Atmosphere. , 219(1):1.

Vargas Domínguez, S., Palacios, J., Balmaceda, L., Cabello, I., and Domingo, V. (2011). Spatial distribution and statistical properties of small-scale convective vortex-like motions in a quiet-Sun region. , 416(1):148–154.

Wang, Y., Noyes, R. W., Tarbell, T. D., and Title, A. M. (1995). Vorticity and Divergence in the Solar Photosphere. , 447:419.

Wedemeyer, S. and Steiner, O. (2014). On the plasma flow inside magnetic tornadoes on the Sun. , 66:S10.

Wedemeyer-Böhm, S. and Rouppe van der Voort, L. (2009). Small-scale swirl events in the quiet Sun chromosphere. , 507(1):L9–L12.

Wedemeyer-Böhm, S., Scullion, E., Steiner, O., Rouppe van der Voort, L., de La Cruz Rodriguez, J., Fedun, V., and Erdélyi, R. (2012). Magnetic tornadoes as energy channels into the solar corona. , 486(7404):505–508.

Weiss, J. (1991). The dynamics of enstrophy transfer in two-dimensional hydrodynamics. *Physica D Nonlinear Phenomena*, 48(2-3):273–294.

Winebarger, A. R., Warren, H. P., Cirtain, J. W., Kobayashi, K., Korreck, K. E., Golub, L., Kuzin, S., Walsh, R. W., DeForest, C., De Pontieu, B., Title, A. M., and Weber, M. (2012). The Fundamental Structure of Coronal Loops. In *AGU Fall Meeting Abstracts*, volume 2012, pages SH31B–06.

Witze, A. (2017). Eclipse promises to reveal mysteries of sun's corona. *Nature*, 548(7666):146–147.

Wray, A. A., Bensassi, K., Kitiashvili, I. N., Mansour, N. N., and Kosovichev, A. G. (2015). Simulations of Stellar Magnetoconvection using the Radiative MHD Code 'StellarBox'. *arXiv e-prints*, page arXiv:1507.07999.

Yadav, N., Cameron, R. H., and Solanki, S. K. (2020). Simulations Show that Vortex Flows Could Heat the Chromosphere in Solar Plage. , 894(2):L17.

Yadav, N., Cameron, R. H., and Solanki, S. K. (2021). Vortex flow properties in simulations of solar plage region: Evidence for their role in chromospheric heating. , 645:A3.

Yadav, N., Keppens, R., and Popescu Braileanu, B. (2022). 3D MHD wave propagation near a coronal null point: New wave mode decomposition approach. , 660:A21.

Zhou, J., Adrian, R. J., Balachandar, S., and Kendall, T. M. (1999). Mechanisms for generating coherent packets of hairpin vortices in channel flow. *Journal of Fluid Mechanics*, 387(1):353–396.

Zigunov, F., Sellappan, P., and Alvi, F. S. (2020). Dynamics of the slanted cylinder afterbody vortices.

Zirin, H. (1998). Dynamics of solar spicules. In Guyenne, T.-D., editor, *Solar Jets and Coronal Plumes*, volume 421 of *ESA Special Publication*, page 39.