



Accurate Identification of Galaxy Mergers with Stellar Kinematics

R. Nevin¹, L. Blecha², J. Comerford³, J. E. Greene⁴, D. R. Law⁵, D. V. Stark⁶, K. B. Westfall⁷,
J. A. Vazquez-Mata⁸, R. Smethurst⁹, M. Argudo-Fernández¹⁰, J. R. Brownstein¹¹, and N. Drory¹²

¹Center for Astrophysics|Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA; rebecca.nevin@cfa.harvard.edu

²Department of Physics, University of Florida, Gainesville, FL 32611, USA

³Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA

⁴Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

⁵Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

⁶Department of Physics and Astronomy, Haverford College, 370 Lancaster Avenue, Haverford, PA 19041, USA

⁷University of California Observatories, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

⁸Instituto de Astronomía, Universidad Nacional Autónoma de México, A.P. 70-264, 04510, CDMX, México

⁹Oxford Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK

¹⁰Instituto de Física, Pontificia Universidad Católica de Valparaíso, Casilla 4059, Valparaíso, Chile

¹¹Department of Physics and Astronomy, University of Utah, 115 S. 1400 E., Salt Lake City, UT 84112, USA

¹²McDonald Observatory, The University of Texas at Austin, 1 University Station, Austin, TX 78712, USA

Received 2020 October 5; revised 2021 February 1; accepted 2021 February 1; published 2021 May 3

Abstract

To determine the importance of merging galaxies to galaxy evolution, it is necessary to design classification tools that can identify the different types and stages of merging galaxies. Previously, using GADGET-3/SUNRISE simulations of merging galaxies and linear discriminant analysis (LDA), we created an accurate merging galaxy classifier based on imaging predictors. Here, we develop a complementary tool, based on stellar kinematic predictors, derived from the same simulation suite. We design mock stellar velocity and velocity dispersion maps to mimic the specifications of the Mapping Nearby Galaxies at Apache Point (MaNGA) integral field spectroscopy (IFS) survey, and utilize an LDA to create a classification, based on a linear combination of 11 kinematic predictors. The classification varies significantly with mass ratio; the major (minor) merger classifications have a mean statistical accuracy of 80% (70%), a precision of 90% (85%), and a recall of 75% (60%). The major mergers are best identified by predictors that trace global kinematic features, while the minor mergers rely on local features that trace a secondary stellar component. While the kinematic classification is less accurate than the imaging classification, the kinematic predictors are better at identifying post-coalescence mergers. A combined imaging + kinematic classification has the potential to reveal more complete merger samples from imaging and IFS surveys such as MaNGA. We note that since the suite of simulations used to train the classifier covers a limited range of galaxy properties (i.e., the galaxies are of intermediate mass, and disk-dominated), the results may not be applicable to all MaNGA galaxies.

Unified Astronomy Thesaurus concepts: Galaxy evolution (594); Galaxy mergers (608); Galaxy kinematics (602); Stellar kinematics (1608); Galaxy interactions (600)

1. Introduction

Observations of galaxies have revealed that they evolve over cosmic time from smaller, bluer, more irregular star-forming galaxies at higher redshifts, to larger, redder, more elliptical galaxies in the local universe (e.g., Glazebrook et al. 1995; Lilly et al. 1995; Giavalisco et al. 1996). Moreover, the bimodality of galaxy properties such as color, mass, and star formation rate at low redshift implies that galaxies are quenching, or shutting down their star formation, in the local universe as well (e.g., Schawinski et al. 2009; Masters et al. 2010; Weigel et al. 2017). Galaxy evolution, or changes in the size, structures, and star formation properties of lower mass ($\log(M_*/M_\odot) < 10.5$) galaxies is largely driven by the accretion of gas and/or the prevention of this gas from forming stars (Robotham et al. 2014). Many different processes can drive this evolution, from internal processes, dependent on specific galaxy properties, to external processes, which are related to the surroundings of the galaxy. Examples of internal processes include feedback from active galactic nuclei (AGNs; Croton et al. 2006; Fabian 2012; Heckman & Best 2014), star formation driven outflows (Rupke 2018 and references therein), and morphological quenching due to structures such

as bars or stellar bulges (Sheth et al. 2005). External processes include galaxy interactions with a hot intracluster medium that removes or heats gas (Gunn et al. 1972), “cold flow” accretion from the cosmic web (Dekel et al. 2009), and galaxy mergers (Silk & Rees 1998; Di Matteo et al. 2005; Kaviraj 2014).

While the current Λ CDM framework for virial formation in the universe points to the importance of mergers in assembling dark matter halos (White & Rees 1978; White & Frenk 1991; Cole et al. 2008), the relative contribution of mergers to galaxy evolution via processes such as star formation, AGN activity, and/or morphological transformation remains unclear. This disagreement stems largely from the difficulty of building large, unambiguous samples of merging galaxies. Galaxy mergers are inherently difficult to identify; they persist for $\sim$ few Gyr, and exhibit a diversity of identifying characteristics that vary with merger stage, mass ratio, gas fraction, orbital parameters, and other merger initial conditions.

The difficulty of identifying merging galaxies also contributes to uncertainty regarding the merger rate (R_{merg}), which is a key measurement for quantifying the role of mergers in galaxy evolution, and comparing observations to simulations (e.g., López-Sanjuan et al. 2008). The merger rate can either be measured directly from simulations, or empirically, using the

observed merger fraction, and assuming a merger “observability” timescale (Lotz et al. 2011). Both techniques show large scatter between different estimates of merger rate; semi-analytic models and hydrodynamical simulations can result in discrepancies of about an order of magnitude (see Hopkins et al. 2010 and references therein), while observations have also not converged, due to uncertainties with respect to merger timescales, and the completeness of the different methodologies. Recently, however, Mantha et al. (2018) have demonstrated that different empirical estimates of the merger rate can be brought into agreement. This work demonstrates the importance of a careful calibration of the completeness of the merger identification methodology, as well as the observability timescale.

Clean and complete samples of merging galaxies are therefore needed to address the contributions of mergers to evolutionary processes in galaxies, and to reduce systemic uncertainties in the galaxy merger rate. This in turn necessitates a thorough understanding of the limitations and observability timescale of the techniques used to identify merging galaxies. A variety of imaging techniques exist to identify merging galaxies, all of which are susceptible to their own biases. These often rely upon individual imaging tools, or predictors, such as the *Gini* – M_{20} methodology, or the asymmetry of galaxy light in imaging. One approach to overcome these biases is to utilize simulations of merging galaxies to better understand the shortcomings of individual tools, and to characterize the observability timescales of these methods. For example, Lotz et al. (2008, 2010a, 2010b) use simulations of merging galaxies to measure the length of time for which a major merger is observable by the *Gini* – M_{20} and asymmetry metric. They find 0.3–0.5 Gyr observability timescales, meaning that merging galaxies are only visible as mergers using these techniques for a short time during the $\sim$ few Gyr duration of the merger. Another strategy is to combine the predictors to create a single classification tool, which dramatically lengthens the observability timescale by capitalizing on the strengths of individual methods (e.g., Goulding et al. 2018; Snyder et al. 2019).

In Nevin et al. (2019) (henceforth N19), we pursue both of these approaches, and utilize GADGET-3/SUNRISE simulated galaxies to build a merger identification technique. This technique combines seven imaging predictors to create one more accurate and precise classifier, incorporating the strengths of all of these predictors, and lengthening the observability timescale to >2 Gyr. Using this approach to simulate merging galaxies, we achieve a high temporal resolution (relative to cosmological simulations), which enables us to construct a more complete picture of the different stages of a merger. The suite of simulated mergers also provides a known sample of merging and nonmerging galaxies, on the basis of which we can identify the limitations of the identification technique before it is applied.

Recent years have witnessed an increase in the quantity and quality of integral field spectroscopy (IFS) data sets. With these advancements, kinematic predictors provide a promising addition to imaging predictors in the merger identification toolkit. Kinematic predictors are able to directly probe the dynamical histories of galaxies by tracing baryonic and dark matter (Glazebrook 2013). Disturbances in the stellar kinematics are dynamically long-lived, and can identify a merger long after the imaging signatures have faded. For instance, morphological disturbances such as tidal tails can fade on a

~ 500 Myr timescale following final coalescence, and are faint compared to the light of the galaxy (e.g., Hung et al. 2014; Wen & Zheng 2016), whereas kinematic disturbance in the stars of a galaxy can persist for longer (up to $\sim$ Gyr after final coalescence; Hung et al. 2016).

Kinematic predictors may clear up ambiguities in imaging. For instance, some clumpy star-forming galaxies can appear to be mergers, based on imaging, due to their disturbed morphologies (Miralles-Caballero et al. 2011; Petty et al. 2014), yet galaxies with clumpy morphologies may prove to be nonmerging spiral galaxies with clumps of star formation at their centers or in their spiral arms (Alonso-Herrero et al. 2006; Haan et al. 2011). Kinematics have shown promise as an additional tool for determining whether or not a star-forming galaxy is disk-like (e.g., White et al. 2017).

This type of clumpy star-forming galaxy is even more abundant at intermediate and high redshifts, where a higher fraction of galaxies are expected to be actively merging, yet many isolated (nonmerging) galaxies are also inherently clumpy (e.g., Guo et al. 2015). In addition to their clumpy and rapidly evolving morphologies, high redshift galaxies also exhibit distinct kinematic features, such as high velocity dispersions, regardless of whether they are actively merging or isolated (e.g., Law et al. 2012a, 2012b). The decreasing spatial resolution and surface brightness dimming of high redshift galaxies also confound the identification of mergers. Since high redshift galaxies present a host of additional complications, in this work, we focus on local galaxies in order to develop the groundwork for a method that could eventually be extended to the more distant universe.

As with every other merger identification tool, kinematic predictors have their own set of ambiguities and limitations. For instance, in gas-rich mergers, disks are able to survive the merger, and these recently merged galaxies can masquerade as isolated disk galaxies (e.g., Robertson et al. 2006). Hung et al. (2015) found that relying upon kinematics alone to classify a sample of ULIRGs led to many merging galaxies being identified as isolated disks, and provided a false-negative merger identification for up to 50% of ULIRGs. Moreover, the identification technique depends strongly on the merger stage, and the choice of kinematic predictor. Other works have confirmed that some mergers with highly disturbed visual morphology exhibit a distinct lack of disturbance in the stellar kinematics (Bellocchi et al. 2013; Hung et al. 2016). It is therefore important to probe the kinematics of merging galaxies using simulations in order to understand the biases and limitations of these tools before applying them to real galaxies.

There is currently a wealth of work dedicated to the imaging approach to identifying merging galaxies from large surveys. While there are many detailed case studies of the kinematics of individual local mergers (e.g., Dasyra et al. 2006; Piqueras López et al. 2012), there is a lack of detailed statistical-sized kinematic studies of local mergers. Recent years have seen a revolution, with the prevalence of more and more capable IFS surveys, creating opportunities to identify merging galaxies using kinematic signatures. Surveys such as ATLAS-3D (Cappellari et al. 2011), CALIFA (Sánchez et al. 2012), SAMI (Croom et al. 2012), MaNGA (Bundy et al. 2015), and HECTOR (Bryant et al. 2016) offer a promising avenue for the study of the spatially resolved spectral properties of an astounding number of galaxies. Here, we focus on the nearing-completion Mapping Nearby Galaxies at Apache Point

Table 1
Key Simulation Parameters and Matched Isolated Galaxies

Simulation	Mass Ratio	Gas Fraction	Stellar Mass of Primary [$10^{10} M_{\odot}$]	Matched Isolated Galaxies
q0.5_fg0.3	1:2	0.3	3.9	m0.5_fg0.3, m1_fg0.3
q0.333_fg0.3	1:3	0.3	3.9	m1_fg0.3
q0.333_fg0.1	1:3	0.1	4.7	m0.333_fg0.1, m1_fg0.1
q0.2_fg0.3_BT0.2	1:5	0.3	4.2	m1_fg0.3_BT0.2
q0.1_fg0.3_BT0.2	1:10	0.3	4.2	m1_fg0.3_BT0.2

Note. The simulations are named for the mass ratio, gas fraction, and bulge-to-total mass ratio of the merging galaxies. For instance, q0.5_fg0.3 is a 1:2 mass ratio merger, where each progenitor galaxy has a gas fraction of 0.3, and an initial B/T ratio of 0. The stellar mass of the primary (more massive) galaxy is $3.9 \times 10^{10} M_{\odot}$. The matched isolated galaxies are mass-matched to the merging galaxies, and are named based on which merging galaxy they are matched to (i.e., m0.5_fg0.3 is matched to the smaller of the two merging galaxies in the q0.5_fg0.3 merger).

Observatory (MaNGA) survey. MaNGA is an IFS survey of $>10,000$ local galaxies (with a median redshift of $z \sim 0.03$), with a spectral resolution of $R \sim 2000$, and a spatial sampling of 1–2 kpc (Bundy et al. 2015). One of MaNGA’s secondary scientific goals is to help disentangle the evolutionary pathways of galaxies, and to focus on incorporating simulations of merging galaxies with observations. It is thus uniquely well-suited to our project, where the goal is to create a merger classification technique from the kinematics of simulated galaxies, which we will then apply to the kinematics of the $>10,000$ galaxies in the MaNGA survey in order to identify mergers in future work.

This paper is organized as follows: In Section 2 we review the GADGET-3/SUNRISE simulations from N19, describe the process of creating mock stellar kinematic maps from the synthetic spectra of the galaxy merger simulations, introduce the kinematic predictors, and review the linear discriminant analysis (LDA) technique used in N19 and in this work. In Section 3 we describe the results of the LDA classification, including the coefficients of the LDA, the observability timescales, and the accuracy and precision of the method. In Section 4 we describe the LDA coefficients in the context of previous work on mergers, and how the classification changes with mass ratio; we also examine the performance of the kinematic classification technique in the context of other tools and statistical methods. We present our conclusions in Section 5. In this work, we focus on creating kinematic classifications based on simulated galaxies. In future work, we plan to apply these classifications to galaxies in the MaNGA survey. A cosmology with $\Omega_m = 0.3$, $\Omega_{\Lambda} = 0.7$, and $h = 0.7$ is assumed throughout.

2. Methods

In order to construct a merger identification framework using the kinematics of simulated galaxies, we follow a detailed procedure to mimic observations from the MaNGA survey. We introduce the galaxy merger simulations in Section 2.1, and describe the process of preparing mock kinematic maps of the simulated galaxies in Section 2.2. Finally, we introduce the kinematic predictors utilized in our kinematic classification in Section 2.3.

We dedicate several appendices to discussions that are informative, but ancillary to the goals of this paper. We make the deliberate choice to extract the stellar kinematics from the SUNRISE spectra, as opposed to relying directly on particle velocities. We discuss the implications of this choice, and compare the extracted stellar velocity and velocity dispersion

maps to the inherent velocity of the simulation particles, in Appendix A. In Appendix A we also discuss the effects of dust on the simulated observations. In Appendix B, we include further details regarding the addition of noise to the mock spectra. In Appendix C we address AGN contamination, and how we extract stellar kinematics from galaxies hosting AGN.

2.1. GADGET-3/SUNRISE Overview

As in N19, we utilize GADGET-3/SUNRISE simulations of merging galaxies. GADGET-3 (Springel & Hernquist 2003; Springel 2005) is a smoothed particle hydrodynamics (SPH) and N -body code, modeling processes such as radiative heating, radiative cooling, star formation, supernova feedback, and the multi-phase interstellar medium (ISM), using sub-resolution prescriptions. GADGET-3 also includes SMBH accretion, as well as AGN feedback (this is achieved by coupling 5% of the accreted luminosity to the gas as thermal energy). GADGET has been used for many different astrophysical applications, including widespread use in studies of merging galaxies (e.g., Di Matteo et al. 2005; Cox et al. 2006; Hopkins et al. 2006, 2008, 2013a; Robertson et al. 2006; Blecha et al. 2011, 2013b; Hopkins et al. 2013b).

We present the five galaxy merger simulations and their matched isolated simulations in Table 1. The framework for these simulations is established in Blecha et al. (2018), and the simulations themselves are presented in N19. Three of the simulations are major mergers (where the mass ratio, q , of the progenitors is greater than $q = 0.25$, or 1:4; Rodriguez-Gomez et al. 2015; Nevin et al. 2019) and two of the simulations are minor mergers. The major mergers have mass ratios of 1:2, 1:3, and 1:3. We define the gas fraction of these simulations as $f_{\text{gas}} = M_{\text{gas,disk}} / (M_{\text{gas,disk}} + M_{\text{*,disk}})$. The 1:2 and 1:3 mass ratio major mergers have a relatively high gas fraction of 0.3, and one of the 1:3 mass ratio major mergers has a relatively low gas fraction of 0.1.

We verify that the different gas fractions of the simulations (0.1 and 0.3) cover the full range of the gas fractions in the MaNGA galaxies. The mean gas fraction in MaNGA is defined by Barrera-Ballesteros et al. (2018) as

$$\mu_{\text{gas}} = \frac{\sigma_{\text{gas}}}{\sigma_{\text{gas}} + \sigma_{*}},$$

where σ_{gas} is the gas mass density, and σ_{*} is the stellar mass density. The stellar mass density is derived from the stellar template fitting and the gas mass density is derived from CO measurements in the EDGE-CALIFA survey (Bolatto et al. 2017; Barrera-Ballesteros et al. 2020).

Barrera-Ballesteros et al. (2018) find that the mean gas fraction for the MaNGA sample ranges from 0.16 to 0.32. A f_{gas} value of 0.1 is therefore below the mean for MaNGA galaxies, and as such is relatively gas poor, while $f_{\text{gas}} = 0.3$ is at the top of the range of mean values for the sample, and relatively gas-rich.

These simulations are named for their mass ratio and gas fraction; for instance, the gas-rich 1:2 mass ratio major merger is q0.5_fg0.3, the gas-rich 1:3 mass ratio major merger is q0.333_fg0.3, and the gas-poor 1:3 mass ratio major merger is q0.333_fg0.1. All of the major merger progenitors have a bulge-to-total (B/T) mass ratio of 0, meaning that they are a pure disk initially. Both of the progenitor galaxies of the minor mergers have a B/T ratio of 0.2, and both are gas-rich. These simulations are q0.2_fg0.3_BT0.2, which is the 1:5 mass ratio minor merger, and q0.1_fg0.3_BT0.2, the 1:10 mass ratio minor merger.

We build the matched sample of isolated galaxy simulations for each merger simulation based on two sources. Firstly, we use a standalone sample of isolated galaxies, matched for mass and gas fraction to each of the simulations. Some simulations have more than one matched isolated galaxy, but for a case where there is only one isolated galaxy, it is matched to the mass of the larger merging galaxy from the corresponding merger simulation. Secondly, we define snapshots of each simulated merger falling before the first pericentric passage, or >0.5 Gyr after final coalescence, as isolated galaxies. We refer to those isolated galaxies from snapshots occurring prior to the first pericentric passage as “pre-merger” isolated galaxies, and those snapshots occurring >0.5 Gyr after final coalescence as “post-merger” isolated galaxies. This distinction is useful, as the properties of these two populations differ.

We couple GADGET-3 with SUNRISE in order to directly compare the simulated galaxies with observations. SUNRISE is a 3D polychromatic Monte Carlo dust radiative transfer (RT) code (Jonsson 2006; Jonsson et al. 2010), which is used to model a wide range of isolated and merging galaxies (e.g., Narayanan et al. 2010; Hayward et al. 2011, 2014; Blecha et al. 2013a; Snyder et al. 2013). The full details of the SUNRISE prescription are presented in Blecha et al. (2013b, 2018) and N19. Briefly, SUNRISE performs Monte Carlo radiative transfer on a 3D adaptively refined grid to compute the emission from stars, H II regions, and AGN. SUNRISE uses the STARBURST99 stellar population synthesis models (Leitherer et al. 1999) to calculate the age- and metallicity-dependent spectral energy distributions for each star particle. The treatment for dust includes dust self-absorption and thermal re-emission, as well as polycyclic aromatic hydrocarbon absorption and emission. In addition, we include kinematic (Doppler) effects, which requires very high spectral resolution. Ultimately, SUNRISE calculates the emergent, attenuated, resolved UV-to-IR spectra (3300–6990 Å, $\Delta\lambda = 0.3$ Å) for seven isotropically positioned viewing angles.

We utilize a data cube of SUNRISE optical synthetic spectra for the seven isotropically positioned viewpoints from each merger snapshot to produce the mock data cubes. In N19, a “snapshot” is the SUNRISE image; in this work, we use the term “snapshot” to refer to the full data cube from a specific point in time. These snapshots occur at 50–100 Myr intervals during each merger, and we refer to them as early-stage, late-stage, and post-coalescence stage snapshots. We define these stages using the r -band images from N19. The early-stage

mergers occur after the first pericentric passage, and have (viewpoint) average stellar bulge separations of $\Delta x \geq 10$ kpc, late-stage mergers have separations of $1 \text{ kpc} < \Delta x < 10 \text{ kpc}$, and post-coalescence mergers are no longer resolvable with separations $\Delta x \leq 1 \text{ kpc}$ until 0.5 Gyr after final coalescence. With a 50–100 Myr cadence for snapshots, we therefore have 5–10 snapshots for each of these stages. In total, there are ~ 20 snapshots per simulation, and seven viewpoints per snapshot, which amounts to 100–200 observations per merger simulation.

We further discuss the importance of running RT, and incorporating dust attenuation and scattering into the merger snapshots in Appendix A; briefly, the stellar kinematic maps are affected by both the presence of dust, and dust scattering. The implication is that for this type of kinematic analysis, it is important to use velocities derived directly from the RT product (synthetic spectra), as opposed to the original SPH particle velocities.

2.2. Preparing Mock MaNGA Kinematic Maps

To produce stellar kinematics for our sample of simulated galaxies, we use the specifications of MaNGA to create a data cube of spectra, and then we mimic the MaNGA Data Analysis Pipeline (DAP) to extract stellar kinematics for use in our kinematic classification. Examples of finalized “MaNGA-ized” stellar velocity, and stellar velocity dispersion maps, are presented in Figure 1. In this section, we describe how we mimic the specifications of MaNGA. This involves reducing the spatial and spectral resolution of the simulations to create a MaNGA-ized data cube, placing an appropriately sized fiber bundle over each galaxy, and fitting each spaxel with `ppxf` (a penalized pixel fitting method from Cappellari & Emsellem 2004; Cappellari 2017) to obtain the velocity and velocity dispersion of the stars at each spatial position.

SDSS-IV/MaNGA is an IFS survey, targeting a sample of $>10,000$ nearby galaxies, and selected to span a wide range of environments and stellar masses (Gunn et al. 2006; Smee et al. 2013; Bundy et al. 2015; Drory et al. 2015; Law et al. 2015; Blanton et al. 2017). Spectra are obtained using the BOSS spectrograph on the 2.5 m telescope at Apache Point Observatory (Gunn et al. 2006), which has a spectral resolution of $R \sim 2000$. Fibers are bundled into integral field units (IFUs); MaNGA has five different fiber bundles, equipped with 19, 37, 61, 91, and 127 fibers (the largest fiber bundle is known as the “frankenbundle”); each individual fiber has a $2''.0$ diameter with $2''.5$ spacing between fibers (Smee et al. 2013; Drory et al. 2015; Yan et al. 2016a; Wake et al. 2017). These fiber bundles range from $12''.5$ to $32''.5$ in diameter.

MaNGA has a median point-spread function (PSF) of $2''.5$, which roughly corresponds to a spatial resolution of 1–2 kpc. The primary sample of galaxies (which is $2/3$ of the full sample) has coverage out to 1.5 times the r -band effective radius (R_e), and the secondary sample has coverage to $2.5 R_e$ (Yan et al. 2016a; Wake et al. 2017). The redshift range of the MaNGA survey is $0.01 \leq z \leq 0.15$.

The most recent internal MaNGA product launch (MPL-9) includes 8000 unique galaxies, observed and reduced by the Data Reduction Pipeline (DRP; Law et al. 2016). The publicly available version is released as DR-15, which includes 4621 unique galaxies (Aguado et al. 2019). The derived properties (including the stellar kinematics) are produced by the DAP (Belfiore et al. 2019; Westfall et al. 2019) in the format of a single data cube per galaxy (Yan et al. 2016b). The MaNGA

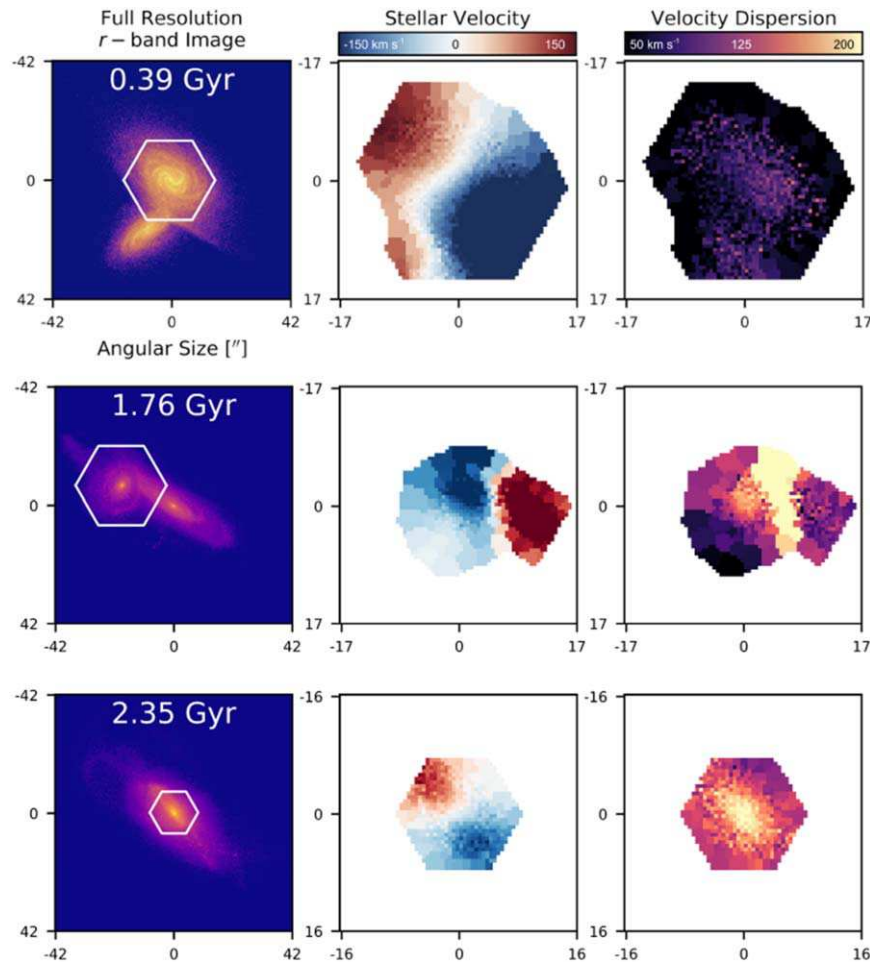


Figure 1. Snapshots of images (left column), stellar velocity maps (middle column), and stellar velocity dispersion maps (right column) from different epochs of the q0.5_fg0.3 simulation. The r -band image is the log-scaled full-resolution simulation prior to the mock-up process, in order to show all of the features of the merger. The colorbar for the middle and right columns is given in km s^{-1} . The spatial position for all panels is given in arcsec, and the stellar velocity and stellar velocity dispersion columns have the same spatial coverage. We include snapshots of an early-stage merger (first row), a late-stage merger (second row), and a post-coalescence merger (third row). The stellar kinematics change over the course of the merger. For instance, the stellar velocity map is distorted due to the superposition of two merging galaxies, while the velocity dispersion map undergoes a global enhancement with time.

team also creates and maintains Marvin, which is a useful tool for visualizing MaNGA data (Cherinka et al. 2019).

To create the mock data cubes, we begin with the SUNRISE synthetic spectra, which we extract at the median redshift of the MaNGA survey ($z = 0.03$). We also use the SUNRISE SDSS r - and g -band images to construct the mock data cubes, as these are essential for certain steps in the process. We follow this procedure (which mirrors the MaNGA DAP whenever possible):

1. Convolve the data cube with the $2''.5$ MaNGA PSF.

Here we use a $2''.5$ Gaussian kernel, which is a good approximation of the effective PSF of the MaNGA data cubes. A model of the effective PSF is automatically computed for each data cube as part of the MaNGA DRP, which convolves a simulated point source with the fiber footprint of a given set of observations, incorporating as-observed details of seeing, transparency, differential atmospheric refraction, dithering, and other instrumental effects (Law et al. 2016). We briefly investigate the difference between using our simplified Gaussian kernel and the effective PSF model provided by the DRP, and find that there are small differences in the maps when the

reconstructed PSF is used, in particular a slight increase in the spread of values in the velocity dispersion maps. This is to be expected, given that the reconstructed PSF does not have a perfectly Gaussian shape. However, the differences in the kinematic maps are minimal, and do not cause significant differences in classification. We define the classification “significantly different” in Section 3.8. We therefore conclude that using the $2''.5$ Gaussian PSF is adequate for this work.

2. Rebin to match the spatial ($0''.5$ spaxels) and spectral sampling ($R \sim 2000$) of MaNGA. The spectral sampling varies as a function of wavelength.
3. Use a mock g -band image, rebinned to the $0''.5$ spatial scale, to mask all spaxels in the data cube that fall below a g -band signal-to-noise ratio (S/N) cutoff value of 1.

Follow the procedure from N19 to convolve, rebin, and introduce noise characteristic of SDSS imaging to the mock g -band images, to match the $0''.5$ spatial binning of the MaNGA cubes. Next, find the average g -band S/N per spaxel, and mask all spaxels that fall below an S/N cutoff value of 1. This procedure directly follows the MaNGA DAP (Westfall et al. 2019), which masks all spaxels using the same g -band S/N cutoff.

4. Use the MaNGA procedure to select which size of fiber bundle to use for each mock data cube, and mask all spaxels that are external to this hexagonal footprint.

We use `statmorph` (Rodríguez-Gomez et al. 2019) to measure the effective radius of the mock r -band images from N19. We then determine the smallest fiber bundle needed to cover each galaxy to $1.5 R_e$ (this is how MaNGA’s primary sample is defined). We select the smallest fiber bundle if the total angular extent of the galaxy ($2 \times 1.5 R_e$) is smaller than $12''.5$, and the largest fiber bundle if the angular size exceeds $32''.5$.

5. We introduce noise to each spaxel to produce a data cube with noise, and a `sqrt(variance)` data cube (from here on, “error data cube”).

We first produce a typical noise spectrum to demonstrate how the noise trends with wavelength for MaNGA observations. We then normalize this noise spectrum, using the g -band S/N value for each spaxel. The end result is a `sqrt(variance)`, or error spectrum, which we use to introduce random noise to each spaxel in the data cube. The noisy spectra, and the accompanying error spectra, are the inputs to `ppxf`. Further details of this process can be found in Appendix B.

To verify that the S/N of the simulated spectra are representative of the MaNGA sample, we use the peak g -band S/N as a comparison statistic, defined as the maximum value of the g -band S/N (per pixel) from a single galaxy observation. The peak g -band S/N ratio of a sample of MaNGA galaxies spanning the full range of sizes, surface brightnesses, and stellar masses of the MaNGA sample ranges from 10–60, with a median of 25. The same statistic for the simulation suite ranges from 10–100, with a median of 30. In Section 3.9, we experiment with changing the S/N of simulated spectra, and investigate how this affects the classification.

Since the MaNGA data cubes oversample the effective PSF, they also contain significance covariance in the errors between adjacent spaxels, such that the S/N ratio of binned spectra does not increase as $\sqrt{N}$. This covariance is irrelevant for the fitting of individual spectra, but we account for it in our Voronoi binning by following the analytic approximation given by Law et al. (2016), as discussed below.

6. Having completed the masking steps, we further exclude regions that are background dominated.

At this stage, we note that the data cubes have “patchy” outskirts, or regions of low S/N data, surrounded by masked regions. The MaNGA data cubes do not have this feature; instead, they exclude regions that could be characterized as “background dominated.” This patchiness does not affect the results of the classification; instead, we choose to correct it for cosmetic purposes. To do this, we mask spaxels where the g -band signal is less than 3σ above the background value, where σ is the standard deviation of the noise given above. This produces the desired effect, whereby the mask has a sharper cutoff, matching the appearance of the MaNGA cubes.

7. Rebin spatially, using a Voronoi binning scheme with a g -band S/N of 10 (Cappellari & Copin 2003).

We create spatial bins with a g -band S/N of 10, reproducing the procedure described in Westfall et al. (2019).

When a Voronoi bin contains more than one spaxel, the new spectrum is the masked average of all constituent spectra, while the error spectrum for that bin is determined by co-adding the error spectra. It is important to account for covariance between neighboring spaxels in our Voronoi bin calculation. In order to avoid the computational cost of calculating the covariance matrix for all spaxels, we instead use the correction given in Law et al. (2016). This correction is an analytic function of the number of spaxels in a bin (N_{bins}):

$$n_{\text{measured}}/n_{\text{no covar}} = 1 + 1.62 \times \log(N_{\text{bins}}),$$

where n_{measured} is the corrected noise level after the correction is applied to the co-added error, where covariance is not considered ($n_{\text{no covar}}$), and N_{bins} is the number of spaxels in a bin.

The final step in the creation of mock kinematic maps is to pass the Voronoi binned spectra through `ppxf` (Cappellari & Emsellem 2004; Cappellari 2017). `ppxf` is a penalized pixel fitting method, which assumes that a galaxy spectrum is a combination of stellar templates, convolved with line-of-sight velocity distribution (LOSVD).

To run `ppxf`, we follow these steps from the DAP:

1. Normalize the flux data so that the mean over all templates is unity.
2. Mask the spectra to match the wavelength range of the MILES–HC library (3600–7400 Å).
3. Mask the emission lines using the DAP module `StellarContinuumBitMask()`.
4. Use the 42 template MILES–HC spectral library to globally fit each data cube.

These templates are first convolved to the spectral resolution of MaNGA.¹³

5. We use the “NZT,” or non-zero template iteration mode to fit all bins with `ppxf`.

In this mode (which is also used in the DAP), we first fit the masked average of all spectra in the data cube, and use this global fit to isolate the subset of templates that are allocated non-zero weights. This template subset is then used to individually fit each bin.

6. Each fit iteration of `ppxf` uses an additive eight-order Legendre polynomial, and a Gaussian line of sight velocity dispersion (LOSVD) with two moments. As in the DAP, due to limited spectral resolution, we do not solve for the higher order moments, h_3 and h_4 (Westfall et al. 2019).

The final product of our MaNGA-izing procedure is the first two moments of the LOSVD, or a stellar velocity map and a stellar dispersion map, both with associated error maps from the fit to the stellar continuum.

2.3. Preparing Kinematic Predictors

Here, we define and describe the predictors extracted from the stellar kinematic maps. The goal is to create a set of kinematic predictors to adequately describe the different types of merger-induced kinematics in the velocity and velocity dispersion maps.

¹³ This is a departure from the DAP. However, as noted in Westfall et al. (2019), there is no mathematical difference between our approach and a later subtraction of the difference in resolution in quadrature from the `ppxf` result.

Table 2
Synthesis of All Kinematic Predictors Measured in This Paper

Predictor Name	Description	Derivation
A	The weighted asymmetry in the position angle, calculated from the Radon profile	$A = \frac{\sum_i \delta \hat{\theta}_i}{2N_{i,j}} w_{i,j}$ where $\hat{\theta}$ is the best-fit kinematic position angle
A₂	The error-weighted asymmetry in the position angle	$A_2 = \sum_i \frac{\delta \hat{\theta}_i}{\sigma_{\delta \hat{\theta},i}}$
ΔPA	The difference between the global kinematic and photometric position angles, measured based on <code>kinemetry</code> and the <i>g</i> -band image	$\Delta PA = \text{PA}_{\text{kin}} - \text{PA}_{\text{img}} $
σ_{asym}	Describes the degree of smoothness of the velocity dispersion map	$\sigma_{\text{asym}} = \left\langle \frac{\sum_{n=1}^5 k_{n,v}/5}{A_{0,v}} \right\rangle_r$ which is the sum of the higher order coefficients from <code>kinemetry</code>
v_{asym}	The deviation of the velocity dispersion map from ordered rotation	$v_{\text{asym}} = \left\langle \frac{\sum_{n=2}^5 k_{n,v}/4}{B_{1,v}} \right\rangle_r$ as above, but excluding the $k_{1,v}$ term
resid	The residual between the best-fit <code>kinemetry</code> model and the velocity map	$\text{resid} = \frac{\sum_{i,j} V_{\text{obs}} - V_{\text{model}} }{N}$
λ_{R_e}	The approximate spin parameter	$\lambda_{R_e} = \frac{\sum_{n=1}^N F_n R_n V_n }{\sum_{n=1}^N F_n R_n \sqrt{V_n^2 + \sigma_n^2}}$
ε	Galaxy ellipticity	Measured using <code>statmorph</code> from the <i>r</i> -band imaging
Δx_v	The spatial distance between the center of the velocity map and the imaging center in kpc	The imaging center is measured from the <i>r</i> -band image, and the kinematic center is based on the Radon Transform
Δx_σ	Same as above, but for the velocity dispersion map	The center of the velocity dispersion map is determined using a low-pass filter
μ_{1,v} and μ_{1,σ}	The mean distribution of the velocity and velocity dispersion maps	The distributions for each map are created by collecting the values of all spaxels
μ_{2,v} and μ_{2,σ}	The variance of the distributions	
 μ_{3,v} and μ_{3,σ} 	The skewness of the distributions	
μ_{4,v} and μ_{4,σ}	The kurtosis of the distributions	

Note. We bold those predictors designated as important. We include a brief description and derivation for each predictor. For further details, see Section 2.3 for those predictors designated as important, and Appendix D for predictors not used in the classification.

To develop this kinematic identification tool, we use stellar kinematics, rather than warm ionized gas kinematics (henceforth, “gas kinematics”). The stellar and gas kinematics trace different physical regions and processes in the merging galaxies. We select the stellar kinematics because they directly trace the assembly history of a galaxy’s stellar population. On the other hand, gas kinematics can be subject to a number of non-gravitational forces. Stellar and gas kinematics diverge in the presence of shocks, inflows, and/or outflows, all of which are processes not limited to merging galaxies. An analysis built on gas kinematics is a compelling direction for future work, but falls beyond the scope of this paper (i.e., see Khim et al. 2020 or Feng et al. 2020). For an analysis comparing the gas kinematics and the stellar kinematics of non-merging and merging galaxies in CALIFA, see Barrera-Ballesteros et al. (2014) and Barrera-Ballesteros et al. (2015), respectively.¹⁴

The kinematic predictors are based on previous work to identify merging galaxies from the stellar kinematics of observed and simulated galaxies. All of these predictors are sensitive to the different orientations, merger stages, mass ratios, and/or gas fractions of merging galaxies. Our goal is to combine them into one LDA classification in order to best identify a variety of different types and epochs of merging galaxies. In total, we extract the following predictors (which

are all introduced in Table 2): *A*, *A₂*, Δ*PA*, *v_{asym}*, *σ_{asym}*, *resid*, *λ_{R_e}*, *ε*, Δ*x_v*, Δ*x_σ*, *μ_{1,v}*, *μ_{1,σ}*, *μ_{2,v}*, *μ_{2,σ}*, *|μ_{3,v}|*, *|μ_{3,σ}|*, *μ_{4,v}*, and *μ_{4,σ}*. We include a brief definition of all predictors in Table 2, but focus the remainder of this section on those kinematic predictors selected by the random forest term selection technique described in Section 3.3: *A₂*, Δ*PA*, *resid*, *λ_{R_e}*, *μ_{1,v}*, *μ_{1,σ}*, *μ_{2,v}*, *μ_{2,σ}*, *|μ_{3,v}|*, *|μ_{3,σ}|*, *μ_{4,v}*, and *μ_{4,σ}*. These terms are the most informative for identifying the merging galaxies, and we discuss them throughout the remainder of this paper. We further describe those kinematic predictors that were not selected in Appendix D.

To define the asymmetry in the kinematic position angle (*A₂*), we utilize the Radon Transform from Stark et al. (2018). We transform the velocity maps into circular coordinates (*ρ*, *θ*), where *ρ* is the distance from the spaxel to the center of the velocity map, which is the kinematic center (defined below), and *θ* is the angle between the positive *x*-axis and the line segment from the kinematic center to the spaxel. The angle *θ* ranges from 0 to 180 in the CCW direction. Positive values of *ρ* are regions of the velocity map above the *x*-axis, and negative values of *ρ* are below the positive *x*-axis.

The Radon Transform is defined as

$$R(\rho, \theta) = \int_0^L v(x, y) dl, \quad (1)$$

where the velocity is summed along line integrals centered on the point (*ρ*, *θ*), and perpendicular to the kinematic center of the

¹⁴ Gas kinematics are not available for many MaNGA galaxies (since many are non-star-forming), but are easier to obtain than stellar kinematics for many high redshift galaxies, and could be a more compelling direction to pursue in this context.

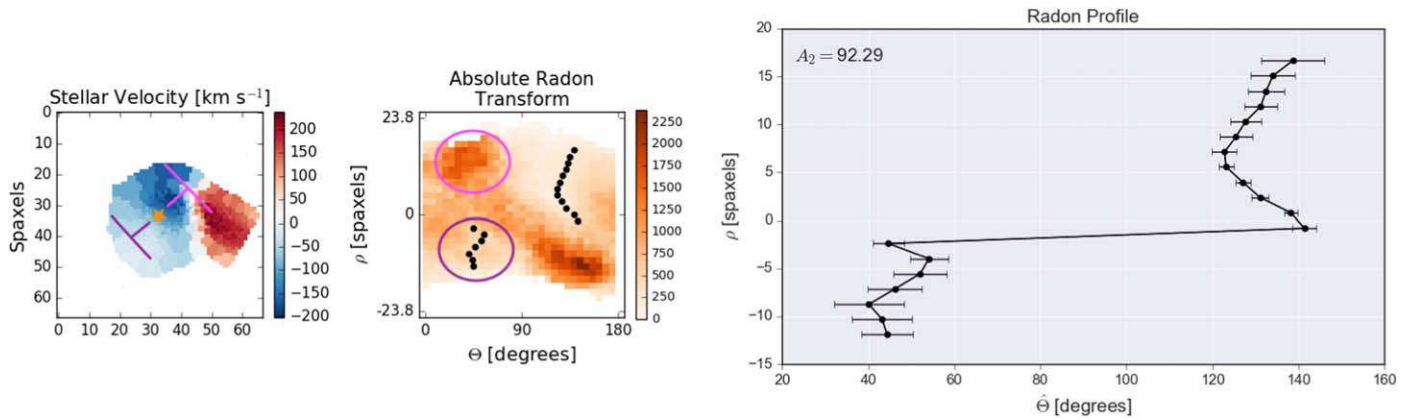


Figure 2. Stellar velocity map (left), bounded Absolute Radon Transform (middle), and Radon profile (right) for a snapshot during the late stages of the q0.5_fg0.3 merger. In this case, the primary galaxy is blueshifted at the center of the velocity map (systemic velocity is $\sim 100 \text{ km s}^{-1}$), and the secondary galaxy approaches from the right, and is redshifted relative to the primary galaxy. To compute the Radon Transform, the velocity field is transformed into θ and ρ coordinates, where $\theta \in [0, 180]$ and $\rho \in [-\infty, +\infty]$, with θ measured CCW from the top of the map. The bounded Absolute Radon Transform is then calculated by creating line integrals over a grid of (ρ, θ) positions, where the line is perpendicular to the kinematic center of the map. It is “bounded” because the line integral is limited to the length, R_e . In the left panel, the kinematic center is a yellow star, and the magenta and purple line segments demonstrate the calculation of the Absolute Radon Transform at $\theta \sim 45$ for positive and negative ρ values, respectively. The magenta and purple regions in the middle panel have large and small values, respectively, demonstrating that the value of the Absolute Radon Transform is smaller in those regions where the spaxel velocities vary less along the line integral. We find the minima (shown in lighter yellow) of R_{AB} at each value of ρ , to measure the Radon profile (right), which is used to calculate the error-weighted asymmetry in the kinematic position angle, A_2 .

galaxy. The Radon Transform is a 2D array, calculated at all values of ρ and θ .

We then calculate the bounded Absolute Radon Transform, R_{AB} , which is integrated over a distance R_e , and denotes the absolute value of the difference between the velocity at each point, and the mean value along the line segment.

We present the bounded Absolute Radon Transform and the Radon profile in Figure 2. The Radon profile is computed by determining the minimum value of θ ($\hat{\theta}$, where the hat operator denotes an estimated value) for each value of ρ from the bounded Absolute Radon Transform. The value of $\hat{\theta}$ traces the direction of maximal rotation in the stellar velocity maps at each radial position.

We follow the procedure of Stark et al. (2018) to determine the galaxy’s kinematic center, which we describe in more detail in Appendix D.

We quantify the asymmetry of the Radon profiles using the kinematic predictor A_2 , defined in Stark et al. (2018):

$$A_2 = \sum_i \frac{\delta\hat{\theta}}{\sigma_{\delta\hat{\theta},i}}, \quad (2)$$

where $\delta\hat{\theta}$ is the absolute magnitude of the difference between the values of θ_i on one side of the Radon profile and the other (same ρ , different sign), $\sigma_{\delta\hat{\theta}}$ is the uncertainty on $\delta\hat{\theta}$, and the expression is summed over the i values of $\hat{\theta}$.

The A_2 predictor incorporates the absolute magnitude of the difference between the measured kinematic PA on one side of the galaxy from the other. We therefore expect that A_2 will be enhanced for merging galaxies, since mergers can cause warps in the stars of a galaxy (e.g., Shapiro et al. 2008).

We use *kinemetry* to measure both ΔPA and residuals (resid) from the LOSVD (Krajinovic et al. 2006). Functionally, *kinemetry* measures the kinematic asymmetry from the line of sight velocity maps by dividing them into a set of nested elliptical rings. The best-fit model at each radius is determined using a ring defined by the kinematic PA, and the flattening factor $q_f = 1 - e$, where e is the ellipticity of the ring in the plane of the sky. These models use a decomposition of the moment

maps into harmonic Fourier coefficients in polar coordinates. For instance, a velocity map, $K(r, \psi)$, can be expanded into a finite number of harmonic frequencies:

$$K(r, \psi) = A_0(r) + \sum_{n=1}^N A_n(r) \sin(n\psi) + B_n(r) \cos(n\psi), \quad (3)$$

where r is the semimajor axis of the ellipse, ψ is the azimuthal angle, $A_0(r)$ is the systemic velocity, N is the number of ellipses fit, and A_n and B_n are the coefficients of the harmonic expansion. The best-fitting ellipses are obtained by minimizing χ^2 for the linear combination of the A_n and B_n coefficients.

An ideal rotating disk can be described using only the B_1 term, which represents the cosine term for the circular velocity of a galaxy’s rotating disk:

$$V(r, \psi) = V_c(r) \sin i \cos \psi, \quad (4)$$

where r is the radius in the plane of the galaxy, ψ is the azimuthal angle, $V_c(r)$ is the circular velocity, and i is the inclination of the galaxy disk.

To determine the best-fit Fourier coefficients, we run *kinemetry* multiple times. We first allow the best-fit kinematic PA and value of q_f to vary for each radius. We define the kinematic position angle (PA_{kin}) to be the median value of the best-fit kinematic PAs. We then allow the value of q_f to vary and determine the median value. After determining the global values for kinematic PA and q_f , we do a final run to determine the values of the higher order kinematic moments and therefore the best-fit disk model. We then compare PA_{kin} to the imaging major axis (PA_{img} , measured using *statmorph* from the r -band imaging) to create the predictor, ΔPA . Since ΔPA traces the recent global misalignments of stars, it should be elevated for those merging galaxies with misaligned stellar disks.

We use the global kinematic position angle from *kinemetry* to measure ΔPA , rather than the median of the kinematic position angles from the Radon Transform. The main motivation for this choice is that *kinemetry* uses an adaptive

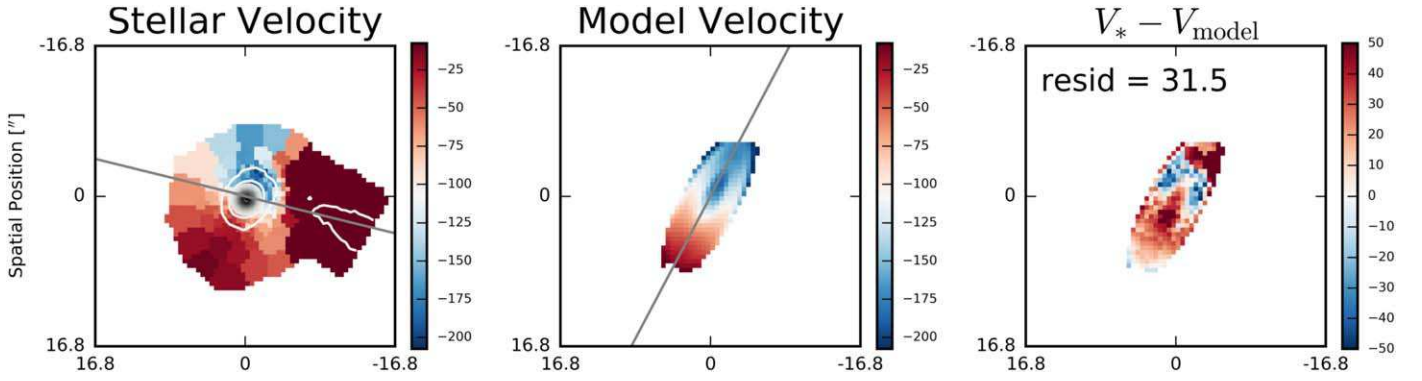


Figure 3. Sample `kinemetry` fit to a snapshot of the q0.5_fg0.3 merger simulation, with observed stellar velocity map (left), best-fit `kinemetry` model (middle), and the model velocity, subtracted from the stellar velocity map (right). Note that this is the same snapshot as that shown in Figure 2. The color bars show the velocity in km s^{-1} . In the left panel we overplot the contours from the r -band imaging, and the imaging position angle. The kinematic position angle (from `kinemetry`) is the straight line in the middle panel. We utilize the normalized residuals as a predictor (right), which we refer to as “resid,” i.e., the sum over all spaxels of the absolute value of the difference of the stellar velocity and the model velocity, normalized by the number of spaxels in the model.

binning scheme; at each step outwards, the ellipses are larger, which gives less weight to the kinematic confusion at the outskirts of the galaxy. The Radon Transform, however, is equally sampled in ρ (see Figure 2), and can therefore be more influenced by the measurement of the kinematic PA at the outskirts of the galaxy. In most cases, the two measurements agree within error, but in cases where the kinematic maps are disturbed, the global kinematic PA from `kinemetry` more closely matches our by-eye assessment of the kinematic PA.

We also extract “resid,” or the `kinemetry` residuals between the best-fit rotating disk and the velocity map. This predictor is defined as

$$\text{resid} = \frac{\sum_{i,j}^N |V_* - V_{\text{model}}(r, \Psi)|}{N}, \quad (5)$$

where V_* is the observed velocity map, $V_{\text{model}}(r, \Psi)$ the circular velocity model from `kinemetry`, and N is the number of fitted spaxels. We include this normalization factor in order to penalize those fits that converge to a very inclined galaxy. For these galaxies, the fit is attempting to avoid fitting disordered kinematics in the exterior regions of the galaxy by fitting a smaller region. We show an example of a simulated galaxy snapshot from the q0.5_fg0.3 simulation fit with `kinemetry` and its velocity residuals in Figure 3.

We measure λ_{R_e} , the approximate spin parameter, from the stellar velocity and velocity dispersion maps, as defined by Emsellem et al. (2007):

$$\lambda_{R_e} = \frac{\sum_{n=1}^N F_n R_n |V_n|}{\sum_{n=1}^N F_n R_n \sqrt{V_n^2 + \sigma_n^2}}, \quad (6)$$

where F_n is the (r -band) flux of a spaxel, R_n is the distance from the kinematic center, V_n is the stellar velocity, and σ_n is the stellar velocity dispersion. We measure λ_{R_e} to the r -band effective radius. Since the fiber bundles are designed to provide coverage of each galaxy to $1.5R_e$, if a secondary nuclei falls toward the outside edge of the hexagonal FOV, it is excluded from the measurement of λ_{R_e} . This effect is more relevant to minor mergers, where the secondary component covers a smaller effective area of the hexagonal FOV.

We measure the ellipticity of a galaxy, ϵ , based on r -band photometry, using `statmorph`. This is distinct from the

ellipticity parameter used by `kinemetry` to fit rotation curves. We do not use ϵ as a kinematic predictor. Instead, we use it to construct the λ_{R_e} - ϵ diagnostic diagram in Section 4.1, where the division between fast and slow rotators is as defined by Cappellari (2016):

$$\lambda_{R_e} = 0.08 + \epsilon/4, \quad (7)$$

where slow rotators fall below this line.

In the λ_{R_e} - ϵ diagram, λ_{R_e} is the more predictive of the two axes; it decreases dramatically for the “slow-rotating” population of galaxies, which are dynamically disordered and dispersion-dominated. We predict that λ_{R_e} will decrease for merging galaxies, since mergers are kinematically disordered, and can contribute to bulge growth, which is associated with enhanced velocity dispersion.

In addition to the kinematic predictors utilized in previous works, we define a new set of predictors, based on the distributions of values in the velocity and velocity dispersion maps. These predictors include $\mu_{1,v}/\mu_{1,\sigma}$, $\mu_{2,v}/\mu_{2,\sigma}$, $|\mu_{3,v}|/|\mu_{3,\sigma}|$, and $\mu_{4,v}/\mu_{4,\sigma}$, which are the standardized moments of the stellar velocity/velocity dispersion maps. These predictors are similar to the formulation given by Sweet et al. (2020), which calculates the moments of PDF(s), where s is the normalized specific angular momentum.

To determine the values of these predictors, we measure the four standardized moments of the distribution; mean (μ_1), variance (standard deviation; μ_2), skewness (μ_3), and kurtosis (μ_4). This produces eight different predictors (four each from the velocity and velocity dispersion distributions). These quantities are different from the higher order moments, h_3 and h_4 , which are typically measured by `ppxf`. We show an example of these predictors, measured from a velocity dispersion map, in Figure 4.

We expect to see an offset in the mean velocity ($\mu_{1,v}$) from the systemic for merging systems, and an enhanced mean velocity dispersion ($\mu_{1,\sigma}$). The spread in the velocity distribution ($\mu_{2,v}$) and the dispersion of the velocity dispersion distribution ($\mu_{2,\sigma}$) could identify superpositions of dynamically distinct stellar components. This could include a secondary merging galaxy, or features such as a stellar bulge.

The higher order moments could be useful for identifying subtler features of mergers, beyond bulk shifts in $\mu_{1,v}$, for example. The skewness of a distribution is sensitive to the tails; we take the absolute value to treat positive and negative skew

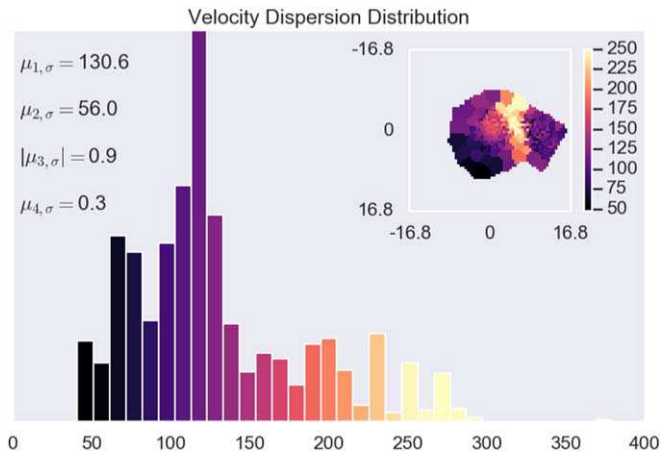


Figure 4. Distribution of the velocity dispersion values (in km s^{-1}) taken from each spaxel in the velocity dispersion map (inset, velocity dispersion bar is given in km s^{-1} , and the spatial axis is given in arcsec). This snapshot is also showcased in Figures 2 and 3. We also include the measured values of the mean ($\mu_{1,\sigma}$), dispersion ($\mu_{2,\sigma}$), skew ($|\mu_{3,\sigma}|$), and kurtosis ($\mu_{4,\sigma}$) of this distribution. A distribution with a larger skew is asymmetric about the mean. A distribution with a positive kurtosis has a high degree of peakedness, relative to a normal distribution.

identically. A skewed velocity or velocity dispersion distribution ($|\mu_{3,V}|$ and $|\mu_{3,\sigma}|$) could have a faint secondary source in the field of view, where the distribution is actually a combination of two galaxy rotation curves. Kurtosis measures how peaked a distribution is, relative to normal; a flatter distribution has a negative kurtosis, and a more peaked distribution has a higher peak. A smoothly rotating velocity dispersion field has a normally shaped distribution, whereas a disturbed field may have a negative (flatter) kurtosis ($\mu_{4,\sigma}$). On the other hand, post-coalescence mergers with recent bulge growth could have a positive kurtosis in the velocity dispersion distribution.

To summarize, we extract the following kinematic predictors: A_2 , ΔPA , v_{asym} , σ_{asym} , resid , λ_{R_e} , ϵ , Δx_V , Δx_σ , $\mu_{1,V}$, $\mu_{1,\sigma}$, $\mu_{2,V}$, $\mu_{2,\sigma}$, $|\mu_{3,V}|$, $|\mu_{3,\sigma}|$, $\mu_{4,V}$, and $\mu_{4,\sigma}$. We then use the techniques described in the following sections (Sections 3.1–3.3) to select the most informative of these predictors. We ultimately use the following predictors in the LDA classification: A_2 , ΔPA , resid , λ_{R_e} , $\mu_{1,V}$, $\mu_{1,\sigma}$, $\mu_{2,V}$, $\mu_{2,\sigma}$, $|\mu_{3,V}|$, $|\mu_{3,\sigma}|$, $\mu_{4,V}$, and $\mu_{4,\sigma}$.

3. Results

Having created mock MaNGA data cubes for the five simulations of merging galaxies (and matched isolated galaxies), we extract the kinematic predictors introduced in Section 2.3. We then prepare the input data, select those predictors that are most informative, and create and assess the classification itself.

In Section 3.1, we describe the LDA technique. We then provide an overview of our process for preparing and examining the data, in the context of the assumptions made by the LDA in Section 3.2. Prior to running the LDA classification, we perform an initial term selection, using a random forest regressor, which we describe in Section 3.3. We present the classification results in Section 3.4, and measure performance statistics in Section 3.5. We present the LDA observability time in Section 3.6. Next, in Section 3.7, we explore some failure modes of the classification. Finally, we analyze how the classification changes with redshift and

decreasing signal-to-noise (S/N) in Section 3.9. Further details of the classification are discussed in the appendices, where we analyze possible biases in classification in Appendix E.

3.1. Linear Discriminant Analysis

The classification in this work relies on an LDA technique that separates nonmerging galaxies from merging galaxies, based on a combination of input predictors (for a review of LDA, see James et al. 2013). This approach was first presented in N19 for imaging predictors; here, we use this approach for kinematic predictors.

LDA is one of many statistical learning tools for performing classification tasks. Using pre-defined features (predictors) as inputs, LDA solves for the hyperplane in multi-dimensional predictor space that maximizes the separation between different classes of objects (i.e., mergers and nonmergers). The solution is a linear combination of the input predictors; the classification is therefore relatively easy to interpret, since its complexity is low. Recent works have employed other techniques to identify merging galaxies, such as random forest regressors (e.g., Goulding et al. 2018; Snyder et al. 2019), and convolution neural networks (CNNs; e.g., Bottrell et al. 2019). These techniques have various advantages and disadvantages, based on the data set at hand, and the goals of the work. Since we aim to optimize the interpretability of the method, we select LDA over, e.g., a CNN. CNNs may increase the number of correct classifications, but they achieve this using complex nonlinear features, which are not easily interpreted.

In this work, we have made several important changes to the technique employed in N19. As such, we review the relevant details from the LDA in N19, prior to discussing these changes. Relevant details of the LDA technique from N19 include:

1. All predictors are linearly standardized prior to the LDA technique, meaning that predictors with large numerical values (such as A_2) do not have an outsized effect on the analysis.
2. We utilize priors on the relative fraction of merging and nonmerging galaxies in nature versus that in the simulations. This accounts for the presence of a greater number of merging galaxy snapshots (relative to nonmerging snapshots) for each simulation. We use the same priors as given in N19; $f_{\text{merg}} = 0.1$ for major mergers, and $f_{\text{merg}} = 0.3$ for minor mergers. These priors are based on the fraction of nonmerging and merging galaxies, taken from both observation and simulations (e.g., Conselice et al. 2009; López-Sanjuan et al. 2009; Shi et al. 2009; Lotz et al. 2011; Rodriguez-Gomez et al. 2015).
3. We include interaction terms to explore correlations between predictors.
4. In order to select which coefficients are necessary for the classification, we use a forward stepwise selection technique, which orders and includes only the most important terms. This technique adds additional terms to LD1 only if they improve the F1 statistic, which is defined in Section 3.5. It also protects against the unnecessary addition of terms by finding the minimum number of terms that produce an F1 statistic, consistent with the maximum (within 1σ errors, where σ is the standard deviation on the F1 statistic, measured from each k -fold cross-validation set).

5. We use k -fold cross-validation to obtain 1σ errors on the predictor coefficients. At each step of the forward stepwise selection process, we divide the sample into k subsets. We then train the LDA on the first $k-1$ subsamples, and test on the remaining subsample, which functions as the “cross-validation” set. We repeat these steps k times for all combinations of subsamples, and the variation in predictor coefficient values from the cross-validation subsamples is the 1σ error.

For complete details, including the full mathematical formulation for LDA, see N19.

We make several changes to the above technique, motivated by the additional challenges arising from the kinematic data:

1. Due to the number of predictor terms in this work, instead of including all of the kinematic predictors in the final classification, we first utilize the RFR technique as a selection technique to eliminate uninformative predictors from our analysis (see Section 3.3).
2. We adjust the model optimization statistic. In N19, we minimized the number of misclassifications, in order to both select predictors, and determine their coefficients. Here, we utilize the F1 statistic defined in Section 3.5 instead; this does a better job of balancing the number of false negatives and false positives in each classification.
3. We also adjust the k -fold cross-validation. Rather than using $k=10$, we find that a value of $k=5$ improves the performance of the LDA, creating a training set, constituting 80% of the sample, and a cross-validation set, constituting 20% of the sample (as opposed to the 90%/10% divide in N19). We then train the LDA on nine of the subsamples, and test on the tenth sample. We repeat this procedure ten times, and the mean number of misclassifications across all ten test samples allows us to determine which set of input predictors to select

We use the LDA both as a term selector, and to determine the coefficients and standard errors for each selected predictor. In order to directly compare imaging classification to kinematic classification, we utilize the same snapshots from all simulations, and we rerun the imaging analysis using all the same methods as for the kinematic classification.

3.2. Data Preparation and LDA Assumptions

Prior to term selection and classification, we examine the distributions of predictor values. We screen for outliers and examine the data in the context of the assumptions made by the LDA. The goal is to gain an understanding of the properties of the data, prior to classification.

Firstly, we remove outliers by transforming the distribution of each predictor into log space. We define data points falling more than 5σ above or below the mean of the distribution for each predictor as outliers. This combination of log transformation and 5σ cutoff allows us to identify outliers caused by errors in the creation of the mock maps, and not simply related to very disturbed kinematics. There are ~ 4 outliers per simulation, amounting to 4 out of 100 or 200 data points.

Secondly, we check the input data for significant violations of the LDA’s assumptions. LDA operates under the assumptions that the predictors are normally distributed (multivariate normality), the covariance among the merging and nonmerging

classes is equal (homoscedasticity), and the predictors are not strongly correlated with one another (multicollinearity).

Here, we test these three assumptions by closely examining the data. We carry out the same statistical tests described in N19 to test for normality, homoscedasticity, and multicollinearity. We find that the data violates all three assumptions. In addition, we plan to introduce interaction terms into the LDA classification, which will further increase the multicollinearity.

As discussed in N19, LDA has proven to be robust to violations of multivariate normality and homoscedasticity (Duda et al. 2001; Li et al. 2006). To ensure that the LDA technique used in this work is robust to these violations, we directly compare the LDA results to those of a logistic regression. The logistic regression and the LDA produce similar results, both in terms of the relative importance of the predictors for each simulation, and in the performance of the method. This is an indication that the LDA is converging, even though it nominally violates several assumptions.

3.3. Random Forest Regressor Term Selection

In N19, we used a forward stepwise selection technique within the LDA to select informative predictors. Here, motivated by both an increase in the number of initial terms,¹⁵ and a decrease in the predictive power of these terms, we modify the term selection procedure. We introduce a random forest regressor (RFR) into the methodology to select a subset of predictors for each simulation, which will then be presented to the LDA classifier.

An RFR (Ho 1995) is an ensemble learning technique that aggregates the result of many individual decision trees, run in parallel. We specifically utilize the `scikit-learn` implementation of RFR (Pedregosa et al. 2011). In an RFR, the number of features that can be used to split at each node of the decision tree is limited to a percentage of the total number of features, ensuring that the ensemble model does not rely too heavily on any one feature. This means that the RFR is able to combine all potentially predictive variables in a fair way. It is also able to incorporate nonlinear features to capture some higher order interaction terms. In practice, we find that the RFR is an efficient method of initially identifying useful features in the data set from an extensive list of kinematic predictors.¹⁶

In order to select the informative terms from the RFR, we include an additional predictor. This predictor is assigned a random number for each galaxy snapshot, and therefore shows no significant difference between nonmerging and merging galaxies. We use this technique to eliminate all terms with a feature importance lower than the random term for all simulations. In this step, we eliminate the v_{asym} , σ_{asym} , A , Δx_v , and Δx_σ predictors. Next, for each individual simulation, we additionally eliminate predictors with an importance lower than the random value prior to initiating the LDA classification. The terms eliminated in this step vary from simulation to simulation.

¹⁵ This is partially due to a dearth of historically utilized kinematic predictors, so we initially introduce many more terms, so as to determine which are informative.

¹⁶ We do not use it as the primary classification technique, because its features can be highly nonlinear, and more opaque to interpretation. Moreover, this technique is designed to directly complement the LDA technique in N19 for the purpose of comparison.

Table 3Final LD1 Predictor Coefficients ($\hat{w}$) with 1σ Confidence Intervals after Term Selection for the First Four Most Important Terms, and the Intercept ($\hat{w}_0$) for All Simulations

Simulation	$\hat{w}$				$\hat{w}_0$
	1	2	3	4	
All Major	$-6.76 \pm 0.45 \lambda_{Re}$	$4.99 \pm 0.6 \mu_{3,\sigma} $	$4.54 \pm 0.36 \mu_{1,\sigma} * \lambda_{Re}$	$-4.44 \pm 0.51 \mu_{1,\sigma} * \mu_{3,\sigma} $	-1.21 ± 0.07
All Minor	$-4.99 \pm 0.74 \mu_{2,\sigma}$	$-4.97 \pm 0.59 \mu_{2,\sigma} * \mu_{4,V}$	$3.47 \pm 0.62 \mu_{4,V} * \mu_{4,\sigma}$	$2.44 \pm 0.38 \mu_{4,V}$	-0.76 ± 0.04
q0.5_fg0.3	$-7.15 \pm 0.78 \mu_{1,\sigma} * \mu_{3,\sigma} $	$-6.7 \pm 0.63 \mu_{1,\sigma} * \mu_{2,\sigma}$	$6.65 \pm 0.53 \mu_{2,\sigma}$	$5.75 \pm 0.2 \mu_{1,\sigma}$	-2.57 ± 0.05
q0.333_fg0.3	$8.27 \pm 0.35 \mu_{1,\sigma}$	$-7.84 \pm 0.71 \mu_{1,\sigma} * \mu_{2,\sigma}$	$5.92 \pm 0.52 \mu_{2,\sigma}$	$5.21 \pm 0.73 \mu_{3,\sigma} $	-0.77 ± 0.18
q0.333_fg0.1	$-7.78 \pm 0.91 \mu_{1,\sigma} * \mu_{2,\sigma}$	$7.09 \pm 0.59 \mu_{2,\sigma}$	$5.97 \pm 0.61 \mu_{1,\sigma}$	...	-0.26 ± 0.28
q0.2_fg0.3_BT0.2	$-6.51 \pm 1.09 \mu_{1,\sigma}$	$-6.2 \pm 0.93 \mu_{2,\sigma} * \lambda_{Re}$	$-5.75 \pm 1.65 A_2$	$5.5 \pm 0.67 \mu_{1,\sigma} * \lambda_{Re}$	-0.79 ± 0.05
q0.1_fg0.3_BT0.2	$25.06 \pm 5.11 \mu_{1,\sigma} * \mu_{4,V}$	$16.02 \pm 3.19 \mu_{1,\sigma}$	$-12.88 \pm 2.51 \mu_{4,V}$	$6.8 \pm 0.99 \mu_{4,\sigma}$	-1.06 ± 0.07

3.4. Classification Results

Having utilized RFR term selection to narrow the number of kinematic terms down to 11 (ΔPA , resid , λ_{Re} , A_2 , $\mu_{1,V}$, $\mu_{2,V}$, $\mu_{2,\sigma}$, $|\mu_{3,V}|$, $|\mu_{3,\sigma}|$, $\mu_{4,V}$, and $\mu_{4,\sigma}$), we then run the LDA classification for each individual simulation. We also combine the three major mergers into a combined major merger classification, and the two minor mergers into a combined minor merger classification. We run the LDA with interaction terms; the result is a linear combination of selected predictors and coefficients which is unique for each simulation. We present the term coefficients and standard errors for the four most important terms and the intercept term in Table 3. Finally, we briefly discuss the main results of the LDA classification for each simulation, which we will examine in more detail in Section 4.

We first introduce the mechanics of the classification. LD1, which is the first linear discriminant axis, is formed from the linear combination of coefficients, multiplied by the standardized predictors and an intercept term:

$$\text{LD1} = C * X + B$$

where C is the matrix of coefficients, X is the standardized values of the selected predictors, and B is the intercept term.

LD1 is the hyperplane that best separates the populations of merging and nonmerging galaxies for each simulation. We use the result from the major merger classification as an illustrative example of how to interpret the LDA results. The LD1 for the major merger combined run (truncated after seven terms) is

$$\begin{aligned} \text{LD1}_{\text{all major}} = & -6.8\lambda_{Re} + 5.0|\mu_{3,\sigma}| + 4.5\mu_{1,\sigma} * \lambda_{Re} \\ & - 4.4\mu_{1,\sigma} * |\mu_{3,\sigma}| - 1.0\mu_{1,\sigma} * \text{resid} \\ & + 1.7\mu_{1,V} * \lambda_{Re} + 1.7\mu_{4,\sigma} * \mu_{4,V} \\ & + \dots - 1.2, \end{aligned} \quad (8)$$

where LD1 is a linear combination of all selected terms, which comprise a coefficient (positive or negative), followed by the standardized value of a predictor. The last term is the intercept term.

The higher the value of LD1, the more likely the galaxy is to be classified as merging. We calculate the values of LD1 for all snapshots from the major and minor combined simulations in Figure 5. The horizontal line at an LD1 value of zero is the decision boundary, corresponding to a p_{merg} value of 0.5; all galaxies with an LD1 value greater than zero would be classified as merging, using a threshold value of 0.5.¹⁷ We find

that the classification is better able to separate the merging and nonmerging classes for the major merger simulations; this is reflected in Figure 5.

There are important nuances to the interpretation of the selected predictors and their coefficients in Equation (8), because the interaction terms complicate the analysis. For instance, in Equation (8), the first selected predictor is λ_{Re} , which has a negative coefficient. Ignoring the rest of the equation, this means that if the λ_{Re} value is large, then the probability that a given galaxy is merging will decrease. However, there are other λ_{Re} terms in the equation that are coupled with other predictors in interaction terms. This means that tweaking the value of λ_{Re} will not linearly change the value of LD1. While the interaction terms complicate the analysis, they are an integral part of the classification. Many of the most important terms for LD1 in Table 3 are interaction terms, and including them significantly improves the performance of the LDA. As we discuss in more detail in Section 4.4, these terms are able to capture the non-monotonic movement of mergers through predictor parameter space.

While it may be difficult to untangle many of the contributing terms, we can use Table 2 to determine which predictors are most prevalent, and therefore informative, for each simulation. For instance, the $\mu_{2,\sigma}$ and $\mu_{1,\sigma}$ predictors are selected as either primary or interaction terms for all simulations. They are therefore universally useful kinematic predictors (for a full discussion of why these terms are important, see Section 4.2.2). The selected predictors from the q0.333_fg0.1 simulation are similar to those of the q0.333_fg0.3 simulation. These two simulations are matched for mass ratio, but not for gas fraction. However, the difference between a gas fraction of 0.1 and 0.3 is insignificant, so we hesitate to form any conclusions regarding the impact of gas fraction on stellar kinematics. On the other hand, the minor merger simulations differ from the major merger simulations in the selected predictors. We find that λ_{Re} and $|\mu_{3,\sigma}|$ are important for the major mergers, whereas the minor mergers are more reliant on certain higher order terms, such as $\mu_{4,V}$ and $\mu_{4,\sigma}$. We explore the implications of these findings, in terms of the physical nature of the kinematics of mergers, in Section 4.

3.5. Performance Statistics and Hyperparameter Tuning

Here, we define and measure the accuracy, precision, recall, and F1 statistic of the simulations (for a review, see Fawcett 2006). We present these results using a confusion matrix for the major and minor combined simulations in Figure 6, which shows the relative fraction of known mergers and nonmergers in the cross-validation samples, classified by

¹⁷ This decision boundary can be moved, either before the creation of the LDA, or after, to be more or less tolerant of false negatives and false positives.

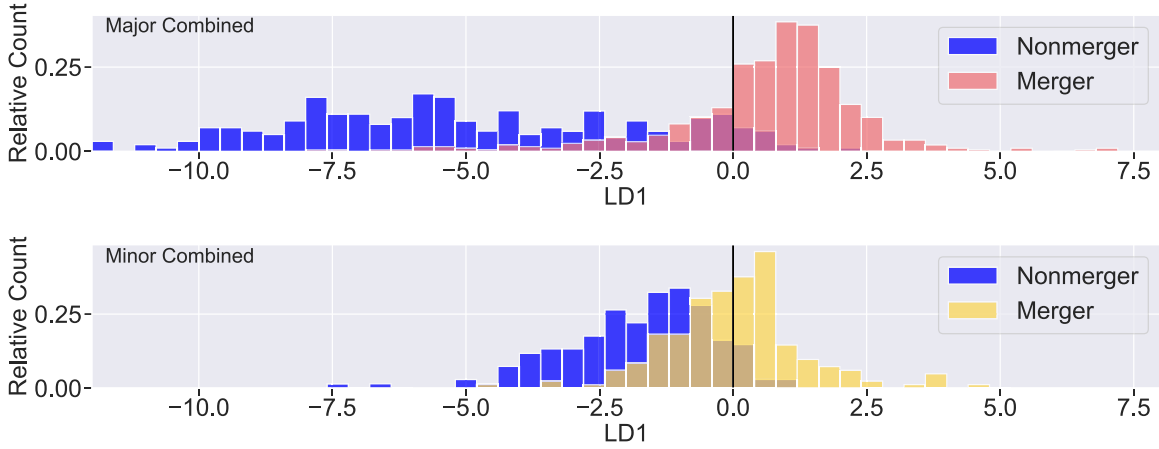


Figure 5. Histograms of LD1 for populations of merging and nonmerging galaxies, based on the combined major merger (top) simulation and the combined minor merger (bottom) simulation. The blue nonmerging samples include both standalone isolated galaxies, and pre- and post-merger isolated galaxies. The nonmerging galaxies in the top and bottom plots span different ranges in LD1 as they are composed of different samples of nonmerging galaxies, and because the selected linear combination of predictors is different for the major and minor merger combined simulations. The vertical black line is the decision boundary; it lies at the midway point between the mean of the nonmerger and merger populations. If the LD1 value of a galaxy falls above this line, the galaxy is more likely to be a merger.

the LDA as merging and nonmerging. These quantities are derived by taking the mean of the performance statistics measured on each of the cross-validation samples. We quantify the accuracy, precision, recall, and F1 score for all simulations in Table 4.

The accuracy for a given simulation is defined as the number of correct classifications of mergers as mergers (true positives), and the number of correct classifications of nonmergers as nonmergers (true negatives), divided by the number of total classifications:

$$A = \frac{TP + TN}{TP + TN + FP + FN}, \quad (9)$$

where FP is the number of false positives, or nonmerging galaxies classified as mergers, and FN is the number of false negatives, or mergers classified as nonmerging. A classifier has a higher accuracy when it is able to increase the number of true classifications relative to false classifications.

Precision is defined as the number of true positive classifications over the total number of positive classifications:

$$P = \frac{TP}{TP + FP}. \quad (10)$$

A precise classifier maximizes the fraction of true positive classifications relative to false positives. Precision is also known as the “positive predictive value.” In this work, we seek to eliminate false positives from the sample, i.e., nonmerging galaxies, incorrectly classified as mergers.

Recall is defined as the number of true positive classifications over the total number of known mergers:

$$R = \frac{TP}{TP + FN}. \quad (11)$$

A classifier with high recall is also known as “complete,” because it correctly identifies the majority of mergers as such.

Finally, we measure the F1 score, or the F1 statistic, which is the harmonic mean of recall and precision:

$$F1 = \frac{2P * R}{P + R}. \quad (12)$$

F1 ranges in value from 0 to 1, and is strongly penalized if either precision (P) or recall (R) is small. We maximize the F1 statistic within the LDA during cross-validation, in order to select the predictor terms used in the classification.

Figure 6 presents the number of true negatives, false positives, false negatives, and true positives (left to right, top to bottom) for the combined major and minor merger simulations. It also quantifies the accuracy, precision, recall, and F1 score. The major merger classification performs better, with accuracy/precision/recall values of 0.81/0.95/0.76, respectively, whereas the minor mergers have values of 0.69/0.87/0.51, respectively. The imbalance between precision and recall is due to the priors utilized in the classification (we use the priors from N19, where $f_{\text{merg}} = 0.1$ for the major mergers, and 0.3 for the minor mergers). We have designed the classification with these strong priors so that when it is applied to galaxy surveys (where there are fewer mergers), the classifier will be more balanced. As a result, the classifier produces more false negatives than false positives when tested on the training set.

We experiment with adjusting the performance statistics of the classification, a process also known as “hyperparameter tuning.” It is possible to increase the number of false positives while decreasing the number of false negatives, either by adjusting the decision boundary (i.e., the threshold of p_{merg}), or by changing the priors. This could be a direction to pursue in future work, if we find that we are no longer tolerant of false negatives, or if we wish to adjust the priors on f_{merg} . As a test, we adjust the priors so that $f_{\text{merg}} = 0.5$, finding that it produces a similar classification, lower precision, and higher recall for the major merger classifications, and results in slightly different selected predictors and higher performance statistics for the minor mergers. While the classification with adjusted priors performs better on the training and cross-validation data sets, we find that it is not a fair representation of the fraction of

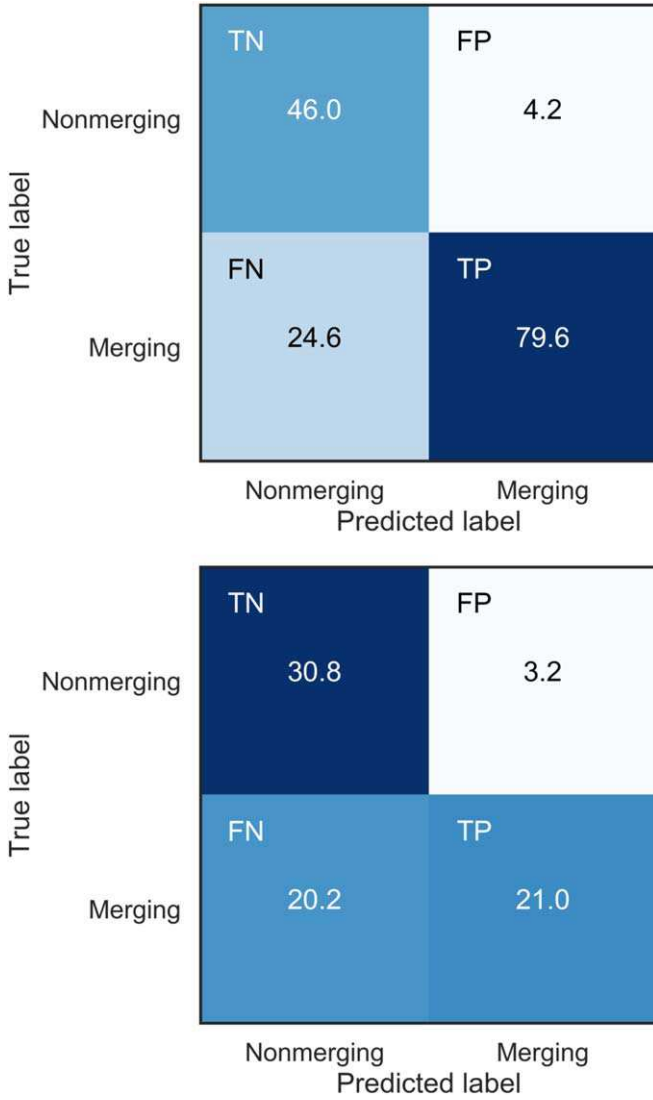


Figure 6. Confusion matrices, showing the number of true negatives (upper left quadrant), false positives (upper right), false negatives (lower left), and true positives (lower right) for the major merger (top) and minor merger (bottom) combined simulations. These matrices show the mean number of galaxy snapshots in each category for the five ($k = 5$) different CV samples.

Table 4
Accuracy, Precision, Recall, and F1 Score for All LDA Runs

Simulation	Accuracy	Precision	Recall	F1
All Major	0.81	0.95	0.76	0.84
All Minor	0.69	0.87	0.51	0.64
q0.5_fg0.3	0.81	0.92	0.60	0.73
q0.333_fg0.3	0.80	0.90	0.79	0.84
q0.333_fg0.1	0.80	0.93	0.80	0.86
q0.2_fg0.3_BT0.2	0.73	0.83	0.62	0.71
q0.1_fg0.3_BT0.2	0.81	0.83	0.69	0.75

Note. We define these statistics in Equations (9)–(12), respectively. The recall value is much lower than precision in all cases, as there is a much higher fraction of false negatives, or mergers that are missed by this method, yet there is a low value of contaminants, or false positives. The performance statistics of the major merger classifications are $\sim 10\%$ higher than those for minor merger classifications.

Table 5

LDA Observability Time, Total Merger Time, and Observability Fraction (LDA Observability Time/Total Merger Time) for Each Simulation

Simulation	LDA Observability Time (Gyr)	Total Merger Time (Gyr)	Observability Fraction
q0.5_fg0.3	0.9	2.2	0.4
q0.333_fg0.3	2.2	2.6	0.8
q0.333_fg0.1	2.4	2.8	0.9
q0.2_fg0.3_BT0.2	3.0	3.5	0.9
q0.1_fg0.3_BT0.2	6.6	9.2	0.7

merging galaxies in nature, so we present the original classification, where $f_{\text{merg}} = 0.1$ and 0.3 , in this work.

Overall, we find that the kinematic classifications generally score lower for all performance statistics than the imaging classifications. This is true for all viewpoints (see Section 3.8 for a discussion of the effect of viewpoint on kinematic classification). For instance, the accuracy/precision/recall/F1 values for the combined major merger run with the imaging predictors are $0.88/0.98/0.84/0.90$. For the combined minor merger run with imaging predictors, the values are $0.80/0.89/0.72/0.80$. Another result is that kinematic minor merger classifications generally score $\sim 10\%$ lower across all statistics than major mergers. We discuss the implications of the performance of kinematic predictors in comparison to imaging predictors, together with the performance of major versus minor merger classifications, in Section 4.8.

3.6. Observability Timescale

The LDA observability timescale is defined as the sum of all consecutive snapshots where the viewpoint-averaged LD1 value for a given snapshot is greater than zero. We present the observability timescales for all simulations in Table 5, along with the total merger duration for each simulation, and the fraction observability, or the fraction of a merger’s duration observable via the LDA technique. We exclude the combined major and minor mergers from this table, since they are built from mergers that progress at different rates.

All of the simulations have a relatively long timescale of observability (2–6 Gyr). The exception is q0.5_fg0.3, where the observability timescale is 0.9 Gyr, due to a decline in the LD1 values for a handful of snapshots in the late stages of the merger. We present a visualization of how the mean values of LD1 change throughout the lifetime of each merger in Figure 7. Here, and throughout the remainder of this paper, we show the q0.5_fg0.3 and q0.2_fg0.3_BT0.2 simulations as examples of major and minor mergers, respectively. This shows the viewpoint-averaged value of LD1 for each snapshot, as well as the 1σ confidence interval on this value, and for the total range. We also plot the decision boundary for each simulation, which falls at an LD1 value of zero (horizontal line). The minor mergers do not fall significantly above this line; even though the viewpoint-averaged LD1 values for the q0.2_fg0.3_BT0.2 simulation fall above the decision boundary, they overlap the decision boundary to 1σ confidence at almost all points in time. This means that not all viewpoints are significantly above this boundary. On the other hand, the major merger simulations are significantly above this boundary for the majority of their duration. While this is not shown in Figure 7, the q0.5_fg0.3 simulation is an outlier with regard to LDA observability time.

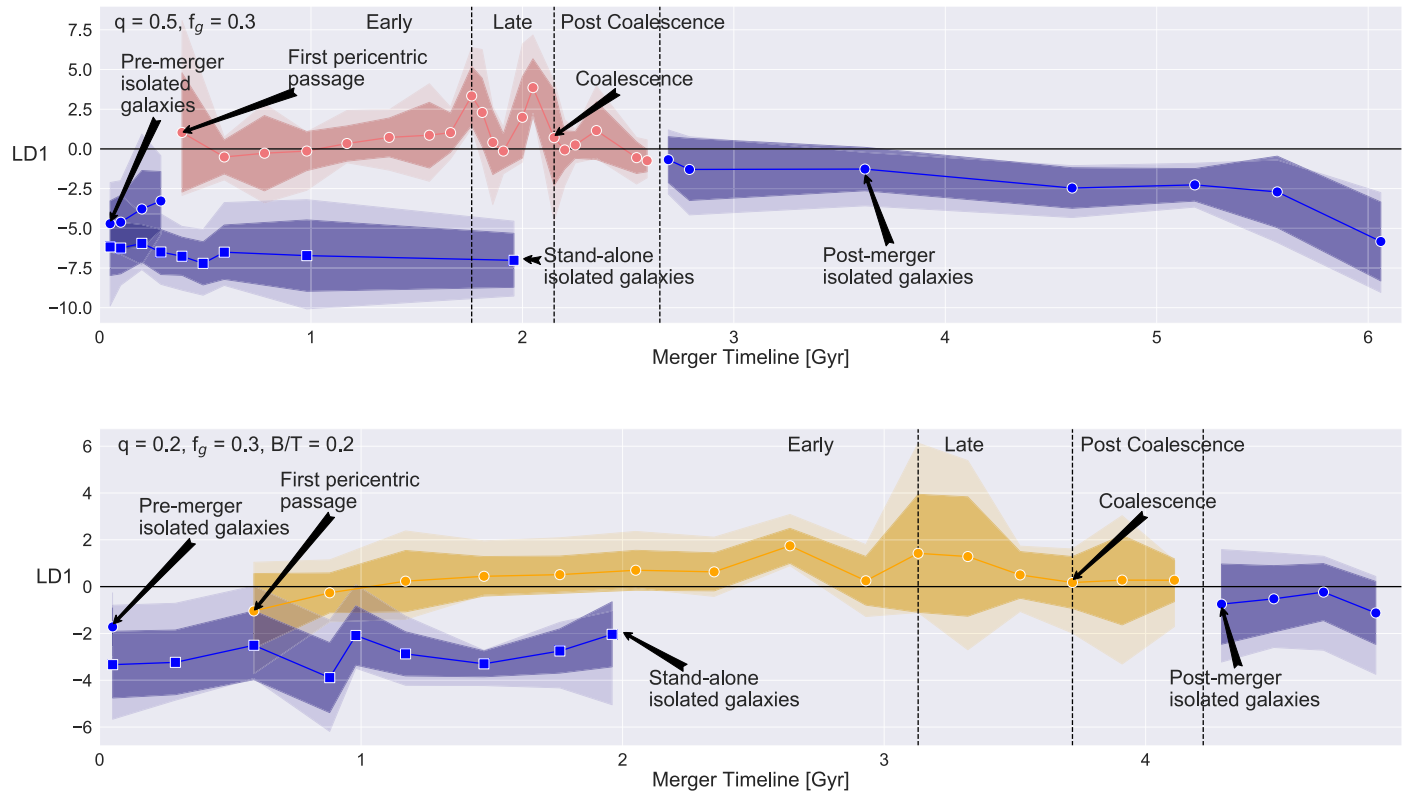


Figure 7. LD1 sensitivity with time for the $q0.5_fg0.3$ (top) and $q0.2_fg0.3_BT0.2$ simulations (bottom). These two simulations are chosen because they are representative of the major and minor merger simulations, respectively. The points represent the viewpoint-averaged value of LD1 for each snapshot in time, along with shaded 1σ confidence intervals (darker shade), based on the scatter of the LD1 values for each snapshot. We also include the full range of values for each snapshot (lighter shade). We divide each plot into the early, late, and post-coalescence stages of the merger. The blue lines and shaded 1σ confidence intervals are associated with isolated galaxies for each simulation. This includes pre- and post-merger isolated galaxies (circles), and standalone isolated galaxies (squares). The horizontal black line is the decision boundary, which marks the divide between merging and nonmerging galaxies, or $p_{\text{merg}} = 0.5$. This figure demonstrates that the major mergers have little to no overlap with isolated galaxies, which produces a more accurate and complete classification (see Section 3.5). The LD1 sensitivity plots for all simulations will be available in an interactive figure.

Figure 7 also demonstrates the sensitivity of the LDA classification to the merger stage. For instance, there are a number of false negatives at the early stages of mergers, where the galaxies are more disk-like. Another key finding is that the sensitivity of the technique gradually decays during the post-coalescence and post-merger stages. We discuss the implications of the different observability timescales, and variations with time, in more depth in Sections 4.5, 4.6, and 4.7.

3.7. Where and Why Does the LDA Fail?

Here, we summarize the factors most likely to lead to false classifications (false positives and false negatives) for the different simulations. Our goal is to identify the primary failure modes of the classification system, and assess whether or not it is making reasonable choices. In other words, we should be concerned if our *by-eye* classification disagrees with the majority of the false classifications.

We present a visual version of a confusion matrix for the $q0.5_fg0.3$ and $q0.2_fg0.3_BT0.2$ classifications in Figures 8 and 9, respectively. These simulations are representative of the results for major and minor mergers, respectively. We generate example velocity and velocity dispersion maps for each classification category (in rows, top to bottom: TP, TN, FN, FP) by combining the results of each iteration of the k -fold cross-validation, then randomly selecting example snapshots from each category.

In Figure 8, following a by-eye examination, it makes sense that many of the false negatives and false positives are incorrectly classified in the $q0.5_fg0.3$ simulation. The false negatives (third row) exhibit orderly rotation with relatively low velocity dispersions. These resemble the pre-merger isolated population shown in the true negatives row. The majority of the false positives (fourth row) are post-merger snapshots with kinematic disturbances.

Incorrect classifications of major mergers can mainly be attributed to two factors. Firstly, the false negatives are due to a lack of disturbed features, meaning that it is difficult to correctly classify many of these snapshots as mergers. Despite these limitations, the classification correctly identifies the majority of early-stage snapshots as mergers, meaning that it is out-performing the by-eye assessment in many cases. Secondly, kinematic disturbances induced by the merger persist into the post-merger stages, producing a number of false positives. These kinematic features are very similar to the features in the post-coalescence stages, so it makes sense that these are commonly classified as false positives. As we discuss in more detail in Section 4.6, our definition of the “end” of the merger (the dividing line between post-coalescence and post-merger stages) is somewhat arbitrarily defined, and results in a number of misclassifications from these two stages.

In Figure 9 we find that it is more challenging to correctly classify the nonmerging and merging galaxies in the $q0.2_fg0.3_BT0.2$ simulation using a by-eye assessment. For

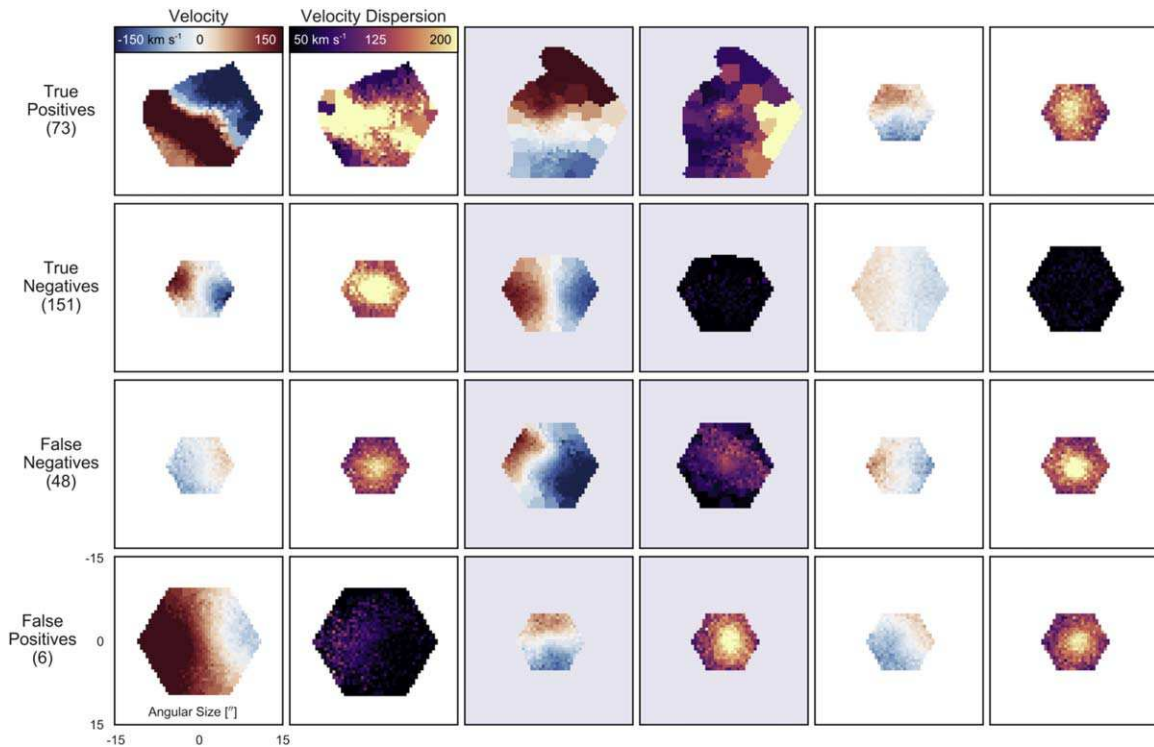


Figure 8. Correct and incorrect classifications from the cross-validation sets for the q0.5_fg0.3 simulation, which is representative of the major merger simulations. The correct classifications include true positives (first row) and true negatives (second row) and the incorrect classifications include false negatives (third row) and false positives (fourth row). We include the number of (non-repeated) galaxies in each category, together with three examples per row of galaxies from the cross-validation sample. The velocity and velocity dispersion maps for each example galaxy cover two consecutive panels, represented by alternating white and gray backgrounds.

instance, the false negatives are very visually similar to the true negatives in terms of their kinematic features. The same is true for the false positives, which are similar to the example true positives. The exception is a number of obvious merger snapshots (we show an example of one such snapshot in the upper middle pair of panels) where the kinematics are dramatically affected. However, these disturbances are short-lived, so that the majority of the merging snapshots appear as per the example in the upper right corner of the diagram. This diagram illustrates the crux of the problem for minor mergers. While the LDA is able to pick up on a number of subtle features (such as stellar bulge enhancements), it ultimately struggles with a number of limitations, related to the lack of identifiability of all stages of a merger as such. These challenges contribute to minor merger classification having lower performance statistics as compared to major merger classification.

Overall, the LDA is not misclassifying obvious (by-eye) mergers or nonmergers. The lack of identifiability of mergers/nonmergers given their kinematic maps is therefore the largest challenge for this technique. Other work highlights this same challenge with kinematic predictors; Hung et al. (2016) find that a significant fraction of merging galaxy kinematics remain indistinguishable by-eye relative to the nonmerging kinematics. This indicates something fundamental about galaxy kinematics: namely, that we are not missing obvious features and instead that merging galaxies are often indistinguishable from nonmergers.

3.8. The Role of Viewing Angle in the Classification

It is well known that many kinematic predictors, such as λ_{R_e} , are correlated with galaxy inclination (e.g., Cappellari et al. 2007; Emsellem et al. 2011; Harborne et al. 2019). In this

section, we examine how the viewing angle, which is a proxy for inclination, affects kinematic predictors, and ultimately, the LDA.

As described in Section 2.1, there are seven isotropically distributed viewpoints (0–6) for each snapshot. Critically, the inclinations are not an exact match between standalone isolated galaxies and merging galaxies. For instance, viewpoints 0 and 4 are the most face-on viewpoints for the merging galaxies, but viewpoints 4, 5, and 6 are the most face-on for the standalone isolated galaxies.

We first explore how inclination affects the λ_{R_e} predictor in Figure 10, where λ_{R_e} increases as the galaxy inclination increases. For instance, viewing angles 0 and 4 are the most face-on, and also have the lowest values of λ_{R_e} . When the 1σ error bars are taken into consideration, the difference in λ_{R_e} values is marginally significant. This is fully consistent with the results from Emsellem et al. (2011), who predict that the measured λ_{R_e} and ϵ values of an axisymmetric rotating oblate spheroid vary, relative to the viewing angle (see Figure 3 of Emsellem et al. 2011). These error bars represent the standard deviation in λ_{R_e} values for all of the different moments in time of the merger. We observe a larger difference in the λ_{R_e} values as a function of time (Section 4.2.1) than as a function of viewpoint.

We next investigate how the LDA classification changes as a function of viewpoint. To visualize this, we plot the distribution of LD1 values in Figure 11. We include the histograms of the LD1 value for nonmerging (blue) and merging (red and orange) galaxies in both the q0.5_fg0.3 simulation (left panel), and the q0.2_fg0.3_BT0.2 simulation (right panel). We then overplot the mean and standard

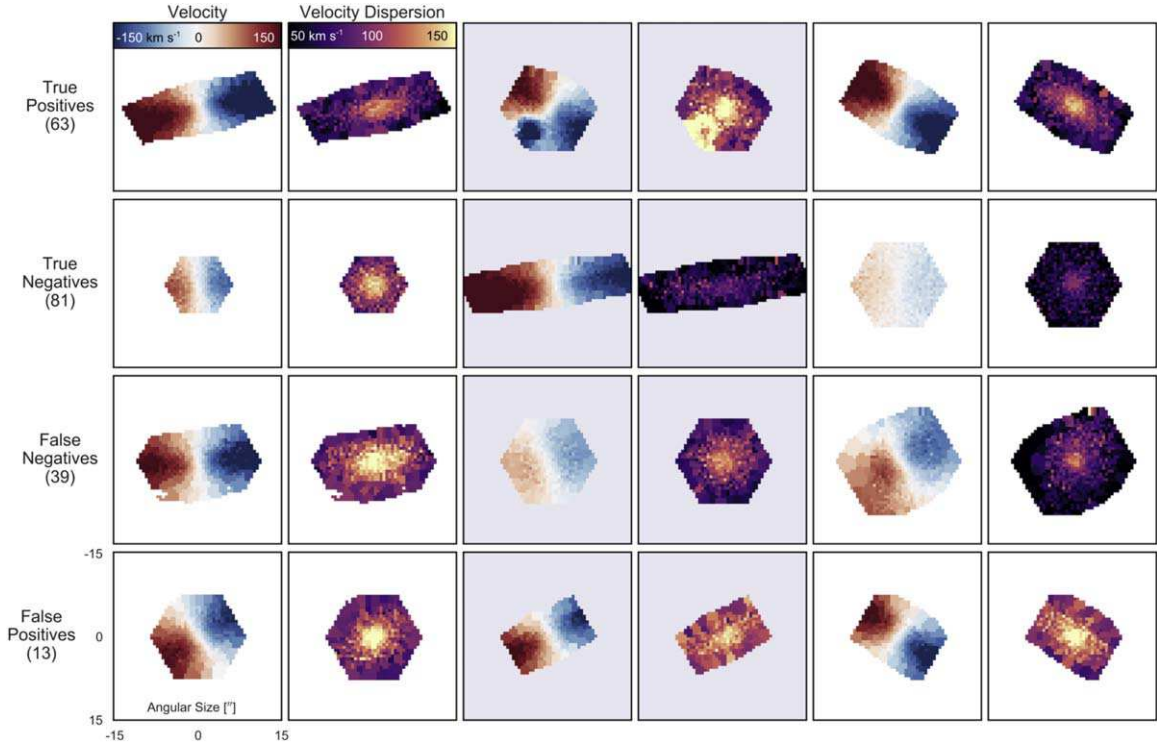


Figure 9. As Figure 8, but for the q0.2_fg0.3_BT0.2 simulation, which is representative of the minor merger simulations.

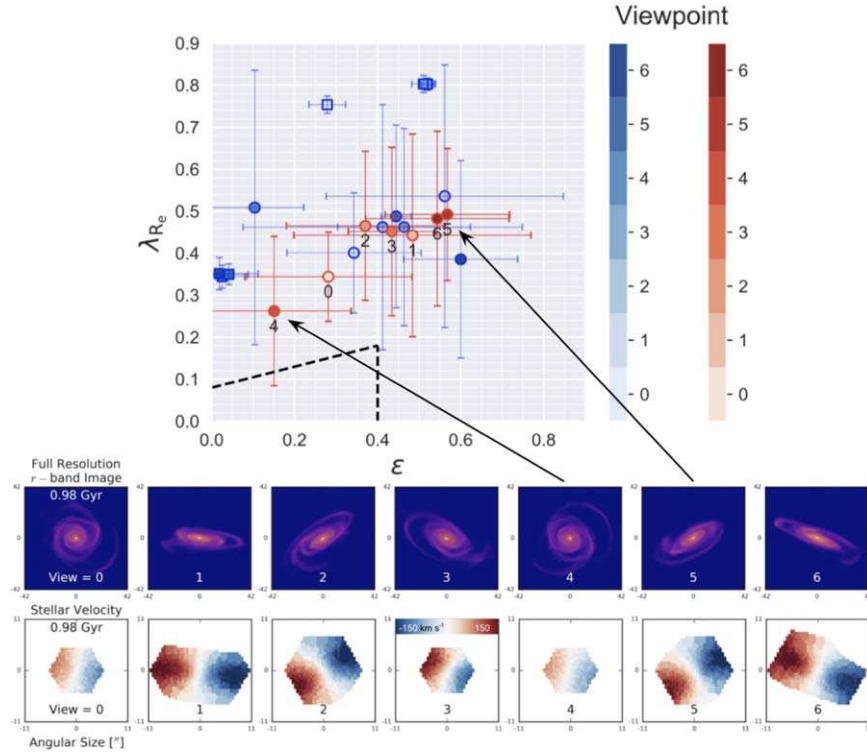


Figure 10. Distribution of the mean values of λ_{Re} and ϵ as a function of viewpoint (top) and full-resolution r -band images (middle) and stellar velocity maps (bottom) for all of the different viewpoints from a snapshot in time for the q0.5_fg0.3 simulation. The more face-on viewpoints (i.e., 4), tend toward lower values of λ_{Re} and ϵ , while the more edge-on viewpoints (i.e., 5) tend to have a larger λ_{Re} value. We also include error bars to demonstrate the standard deviation of the spread at each viewpoint for all of the different moments in time in this simulation. While there is a relationship between inclination and λ_{Re} , the trend is borderline significant. This is consistent with the trend of varying λ_{Re} and ϵ values with viewing angle, as discussed by Emsellem et al. (2011).

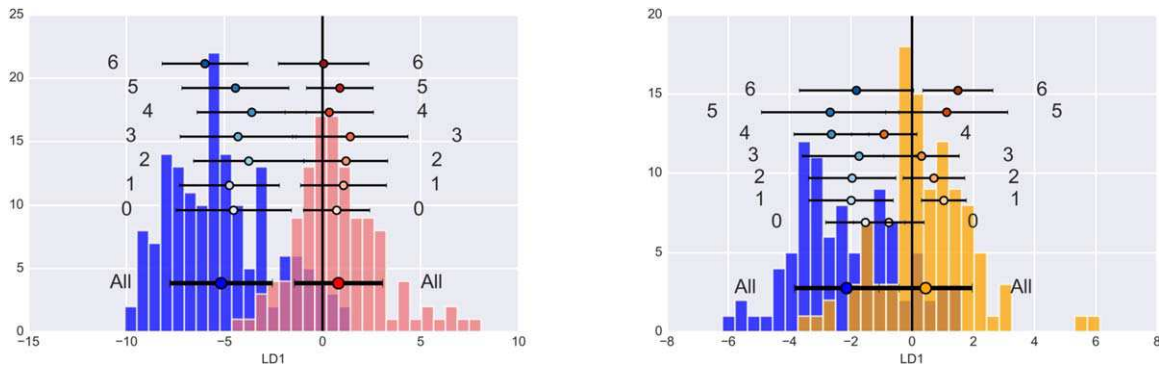


Figure 11. As given in Figure 5, histograms of the LD1 value for the q0.5_fg0.3 simulation (left) and the q0.2_fg0.3_BT0.2 simulation (right). We also overplot the time-averaged LD1 values for each viewpoint, together with error bars to demonstrate the 1σ variation among these values for all moments in time. There is less variation as a function of viewpoint than as a function of time (shown in Figure 7).

deviation of the LD1 values for all snapshots of each specific viewpoint. Focusing on the mean LD1 values for the merging sample, we can determine whether the LDA shows variation as a function of viewpoint.

Focusing first on the left panel of this figure, which refers to the q0.5_fg0.3 simulation, the means for both nonmerging and merging galaxies are not significantly different when considered as a function of viewing angle. In fact, the mean LD1 values are more similar than the variation we observe in LD1 as a function of time in Figure 7. The implication is that the major mergers are LDAs that are fairly robust to changes in viewing angle. For the q0.2_fg0.3_BT0.2 simulation, we observe slightly more variation in the LD1 distribution as a function of viewpoint, and the most face-on viewpoints (0 and 4) have lower LD1 values, which result in more false-negative detections at these viewpoints.

To further quantify changes in the LDA as a function of viewpoint, we iteratively drop the merging galaxies at each viewpoint from the analysis, and rerun the classification for the q0.5_fg0.3 and q0.2_fg0.3_BT0.2 simulations. If the classification changes, this could indicate that a given viewing angle and/or inclination is significantly more or less accurate than the other viewpoints, which would point to inclination itself being the primary driver of this difference.

From here on, we determine that the LDA is “significantly different” from the fiducial run if either of the following criteria are met: firstly, the majority of the selected predictors in the top four selected terms must change, or secondly, the performance statistics in Table 4 must change by more than 10% on average. This quantification of a significantly different classification applies to this section, where we explore the role of different viewing angles, and also to Section 3.9, where we experiment with changes in the data reduction (i.e., changing the S/N or redshift of the simulated galaxies).

When we rerun the LDA classification for the q0.5_fg0.3 simulation iteratively without each viewpoint, the LDAs are not significantly different from the fiducial run. This confirms our findings in Figure 11. For the q0.2_fg0.3_BT0.2 classification, we find that when viewpoints 2, 5, and 6 are absent, the classification is significantly different, with lower performance statistics, and different selected predictors. Our interpretation is that the minor mergers are best identified when the secondary nuclei are within the field of view, which occurs most often in viewpoints 2, 5, and 6. Therefore, the significant changes to the classification as a function of viewpoint both in Figure 11, and in the rerun of the LDA without these viewpoints, can be

attributed to the chance positioning of the secondary galaxy as a function of viewpoint. This means that inclination-related effects on the intrinsic kinematic properties of the primary galaxy are not primarily responsible for the differences in the q0.2_fg0.3_BT0.2 LDA as a function of viewpoint. As a final note, in Appendix E in our discussion of between-class biases, we introduce inclination itself as a predictor in the LDA. We ultimately determine that ϵ , which we use as a proxy for inclination, is not an important predictor. This further supports the finding that changes in the kinematic predictors purely due to inclination effects do not bias the LDA classification itself.

To conclude, we have determined that while the kinematic predictors themselves can vary as a function of viewing angle and/or galaxy inclination, the LDA classification is only sensitive to viewpoint in the sense of the relative positioning of the secondary nucleus relative to the line-of-sight vector.

3.9. Limitations of the Technique in z and S/N

As with the imaging identification technique in N19, the kinematic technique is sensitive to both S/N and resolution, meaning that as the S/N decreases in the spectra and/or as the redshift of the galaxy increases, the technique will undergo significant changes. We test this sensitivity by decreasing the S/N, and by moving the mock galaxies to a higher redshift.

To test the sensitivity of the classification to decreased S/N, we decrease the average S/N of q0.5_fg0.3 simulation by factors of 1.5 and 2, respectively. In Figure 12, we compare a snapshot with S/N that has been decreased by a factor of 2 to the same snapshot from the fiducial run. When the S/N is decreased by a factor of 2, the classification is significantly different. While many of the predictors remain the same, the performance statistics decrease overall, and there is an increase in the number of false negatives during the early stages of the merger. When the S/N is decreased, the Voronoi bins increase in size in the exterior regions of the galaxy. This obscures the large-scale kinematic features, which lowers the performance of the classification. We therefore predict that MaNGA galaxies with low S/N are more likely to be misclassified.

We plan to explore two possible approaches to classifying MaNGA galaxies with different S/N ratios. One option is to implement an S/N cut when we apply the fiducial classification to the MaNGA galaxies. However, MaNGA galaxies with lower S/N need not be excluded from classification. Instead, another option is to use the classification and completeness correction from the lower S/N ratio LDA run, so as to classify these galaxies separately. We predict that we should be able to

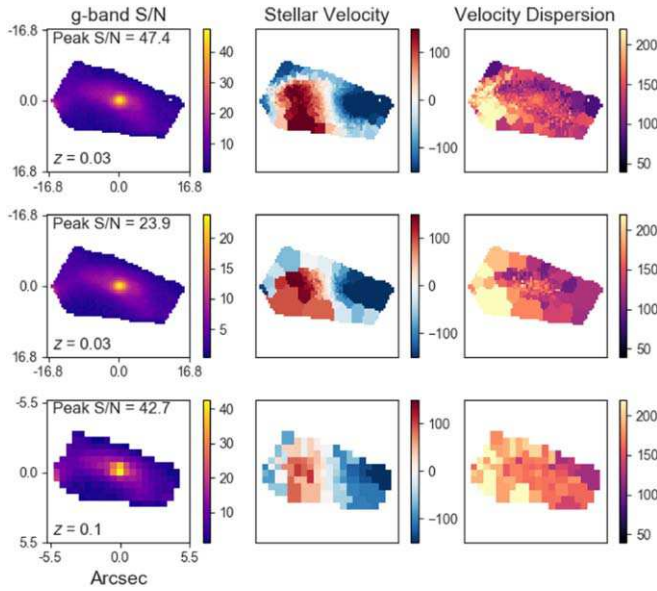


Figure 12. g-band S/N (left), stellar velocity (middle), and velocity dispersion (right) maps from a snapshot of the q0.5_fg0.3 simulation. We have decreased the S/N by a factor of 2 (second row) and redshifted the galaxy to $z = 0.1$ (third row) to demonstrate the point at which the classification begins to change. The classification has a higher failure rate when the S/N is decreased by a factor of 2, mostly due to the sparsity of the Voronoi bins. Moreover, when the spaxel size is increased to mimic a galaxy that is redshifted to $z = 0.1$, the classification begins to change, as this is the point at which the larger-scale kinematic features are distorted by the large spaxel size.

classify the majority of the MaNGA sample, since the fiducial (non-decreased) S/N of the simulation suite is representative of the MaNGA sample.

For the purposes of comparison, in Figure 13 we present a sample of MaNGA galaxies across a range of surface brightness, redshift, and stellar mass. The approximate stellar mass is taken from the NSA catalog, and is estimated using the kcorrect code (Blanton & Roweis 2007). The simulated galaxies, whose stellar masses are $\log M_* \sim 10.6$, are intermediate-mass galaxies, relative to the full MaNGA sample.

We also experiment with increasing the S/N by a factor of 2. The classification undergoes minimal changes, with a slight increase in the performance statistics. When the S/N is higher, we predict that the current classification will be better able to determine whether a galaxy is merging. Since the selected predictors do not significantly change, we can conclude that the fiducial run is successfully identifying all relevant kinematic features.

The mock galaxies are placed at a redshift of $z = 0.03$, which is the median redshift of galaxies observed by MaNGA. In order to understand the limitations of identification over the full range of redshifts for the MaNGA survey ($0.01 < z < 0.15$), we experiment with increasing the redshift of the mock galaxies. To do this, we increase the spaxel size from $0''.5$ to $1''.0$, and $1''.5$, and we increase the PSF size to $5''.0$ and $7''.5$. This mimics the effects of moving the simulated galaxies to a redshift of $z = 0.07$ and $z = 0.1$, respectively. When we artificially redshift the galaxies, we do not introduce cosmological dimming, meaning that the galaxies have the same S/N as the sample at $z = 0.03$. This is because we want to understand the effects of the apparent size of galaxies independently of S/N effects.

The classification does not change significantly when the galaxies are placed at a redshift of $z = 0.07$. At $z = 0.1$, the

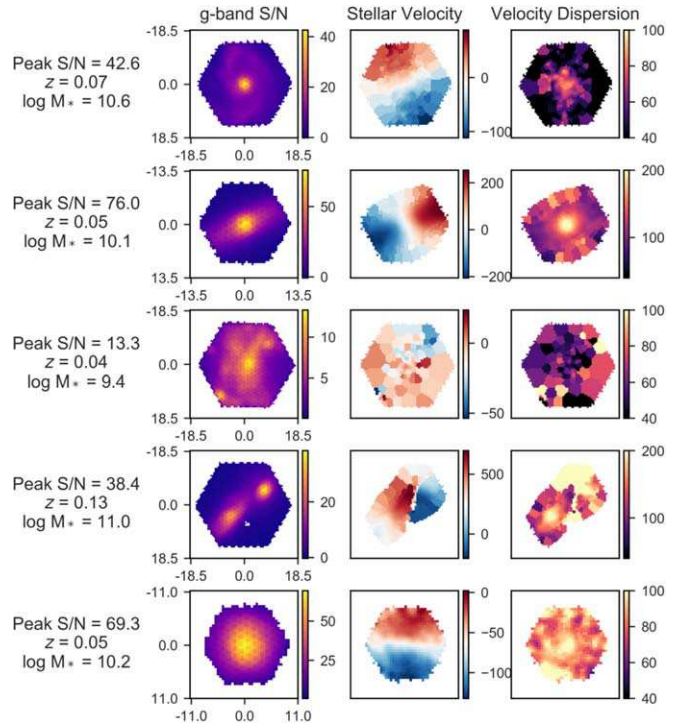


Figure 13. As Figure 12, but for a sample of MaNGA galaxies across a range of surface brightness, redshift, and stellar mass. At low S/N (i.e., first and third rows) the velocity maps have large Voronoi bins, and some kinematic predictors will be difficult to measure. As discussed in the text, the higher redshift galaxies (first and fourth row) in MaNGA also tend to be larger and more massive.

classification is significantly different; the number of misclassifications increases, and the selected terms are different. However, these results are based on a galaxy–galaxy merger, where each galaxy has a stellar mass on the order of $10^{10} M_\odot$. While it is a valid conclusion that the technique will struggle with an intermediate-mass galaxy at $z = 0.1$, the MaNGA sample does not tend to include this type of galaxy. Instead, MaNGA is designed to maintain roughly uniform coverage in $\log M_*$ and radial coverage, meaning that higher mass galaxies ($> 10^{11} M_\odot$), which are more luminous and have larger angular sizes (i.e., the fourth row of Figure 13), are observed primarily at higher redshift, somewhat alleviating this concern (Bundy et al. 2015; Wake et al. 2017).

3.10. Limitations of the Technique in Stellar Mass and B/T

The simulation suite of merging galaxies is limited in stellar mass and B/T ratio. The mergers can all be characterized as intermediate-mass disk-dominated galaxies, spanning a range of $(3.9\text{--}4.7) \times 10^{10} M_\odot$ in stellar mass, and 0–0.2 in initial B/T ratio. These limitations are of particular importance, given that many of the leading kinematic predictors (λ_{R_s} , $\mu_{1,v}$, $\mu_{1,\sigma}$, $\mu_{2,v}$, and $\mu_{2,\sigma}$) are related to the intrinsic kinematic properties of galaxies. For instance, $\mu_{1,\sigma}$ is a proxy for stellar mass, so we are skeptical if this classification can reliably be applied to galaxies that differ in properties, i.e., bulge-dominated, elliptical galaxies.

One possible approach to circumvent these concerns is to remove these predictors from the classification. We rerun the LDA for all simulations, without the λ_{R_s} , $\mu_{1,v}$, $\mu_{1,\sigma}$, $\mu_{2,v}$, and $\mu_{2,\sigma}$ predictors, and find that performance significantly

decreases. Specifically, the accuracy, recall, and F1 score of the major mergers decrease by 20%–50%. This is unsurprising, given that the leading predictors presented in Table 2 include all of the predictors that are tied to the intrinsic kinematic properties of galaxies (λ_{R_c} , $\mu_{1,V}$, $\mu_{1,\sigma}$, $\mu_{2,V}$, and $\mu_{2,\sigma}$). Interestingly, when the intrinsic kinematic predictors are excluded, the performance of the minor merger simulations is not significantly affected. In fact, the performance of the major merger simulations is comparable to, or worse than, that of the minor merger simulations. This highlights the fact that major mergers undergo a more dramatic global transformation during the merging process; this is reflected in the intrinsic kinematic properties of the remnant galaxy.

Since removing these predictors significantly decreases the performance of the classification, we choose to include all predictors in this work, and to attach the following caveat to this paper: Since this analysis focuses on kinematic predictors that are sensitive to intrinsic galaxy properties, we advise against applying this classification to all galaxy types in MaNGA. In Section 4.9, we discuss possible strategies for carefully applying the classification to MaNGA galaxies.

4. Discussion

In the discussion portion of this paper, we consider the implications of kinematic LDA classifications for merging galaxies. We focus on the individual LD1 coefficients in Section 4.1, where we examine why some of the kinematic predictors that have been useful in the past are not informative in relation to this technique. We then examine the most important kinematic predictors in Section 4.2. In Section 4.3 we explore the impact of mass ratio on the stellar kinematics of mergers. We consider the physical meaning of the interaction terms, and their importance with regard to classification, in Section 4.4. We examine the observability timescale of the kinematic LDA technique, and how the observability of a merger varies with time, in Section 4.5. We specifically focus on the definition of the “end” of a merger in Section 4.6, and the kinematics of the merger remnants in Section 4.7. In Section 4.8 we compare the performance of the imaging classifications to the kinematic classifications. Finally, we end with a note on applying this technique to MaNGA IFS observations in Section 4.9.

4.1. Why Are Some Traditionally Utilized Kinematic Predictors Not Useful in This Classification?

Some of the kinematic predictors traditionally utilized to identify merging galaxies are uninformative in this analysis. In this case, “uninformative” means that the predictor is discarded during the RFR term selection steps, or that it has a small relative coefficient value in the LDA. The uninformative predictors are ΔPA , v_{asym} , σ_{asym} , Δx_V , and Δx_σ .

4.1.1. The Misalignment between Kinematic PA and Imaging PA (ΔPA) Is Most Sensitive to Galaxy Inclination

A small fraction of the most dramatic mergers have significantly disturbed stellar kinematic maps. However, this does not translate to a global kinematic PA that is misaligned from the PA in imaging, owing to two factors: First, many of the warps seen in stellar kinematic disks are symmetric, which can produce a global kinematic PA that is not misaligned. Second, the PA for both kinematics and imaging is not well

determined during the most disturbed stages of a merger. This contributes to random deviations around a low ΔPA value.

4.1.2. The Asymmetry in the Velocity and Velocity Dispersion Maps (v_{asym} and σ_{asym}) Is Only Sensitive to the Most Disturbed Times in the Major Mergers

Previous works based on the gas kinematics of simulated and observed mergers have found that merging galaxies have enhanced values for both v_{asym} and σ_{asym} (e.g., Shapiro et al. 2008; Bellocchi et al. 2012; Hung et al. 2016; Bloom et al. 2017). These studies define a threshold value in K_{asym} to identify merging galaxies, where $K_{\text{asym}} = \sqrt{v_{\text{asym}}^2 + \sigma_{\text{asym}}^2}$. For instance, Bellocchi et al. (2012) studied local luminous infrared galaxies (LIRGs), finding a threshold value of $K_{\text{asym}} > 0.135$. Hung et al. (2016) defined a K_{asym} threshold of 0.15 for the galaxies in their work, calculated from the velocities of gas particles for simulated SUNRISE mergers.

The σ_{asym} and v_{asym} predictors are ultimately unimportant in our analysis. We present the viewpoint-averaged values of σ_{asym} - v_{asym} in Figure 14 for the q0.2_fg0.3_BT0.2, and the q0.5_fg0.3 simulations. We include the K_{asym} diagnostic lines used to identify the mergers from Bellocchi et al. (2012) and Hung et al. (2016). The top panel of the plot demonstrates that there is minimal time evolution in the predictor values for minor mergers. The predictor values for major mergers are only slightly enhanced, falling above the diagnostic line of Hung et al. (2016) for a few snapshots during the late stages of a merger. While Figure 14 presents the predictor values in log space, the standardized predictor values of σ_{asym} and v_{asym} used to construct the classification also have minimal separation between the merging and nonmerging populations. The insensitivity of any of the mergers to an enhancement in the values of these predictors results in their exclusion during the RFR selection step.

Ultimately, the v_{asym} and σ_{asym} predictors are unimportant in our work, because v_{asym} is only elevated at specific points in time during the late phases of merging. They are more useful in studies such as that of Bellocchi et al. (2012), who focused on LIRGs, which are prototypical gas-rich major mergers, or that of Hung et al. (2016), where the simulated galaxies include gas-rich major mergers. It is also of interest to note that Hung et al. (2016) found that the mergers in their sample exceeded the K_{asym} value only during the “strong interaction,” or late stage of merging.

4.1.3. The Offset between the Center of the Velocity and Velocity Dispersion Maps and the Imaging Center (Δx_V and Δx_σ) Are Not Very Sensitive to Mergers

We design the Δx_V and the Δx_σ statistics to identify galaxies exhibiting offsets in their kinematic centers. We find that these values are most elevated during the late stages of a merger, where there are two visible nuclei. However, since the kinematic maps are disk throughout the merger, and not dramatically disturbed at most stages, these two statistics are not noticeably elevated for the duration of the merger, and are therefore relatively unimportant.

Statistics such as these have been used in the past for galaxies such as NGC 4473, which is a “double sigma (2σ)” galaxy, meaning that it has two peaks in its 2D stellar velocity dispersion map, which are aligned with the photometric major axis of the galaxy (Krajnović et al. 2011). This type of velocity

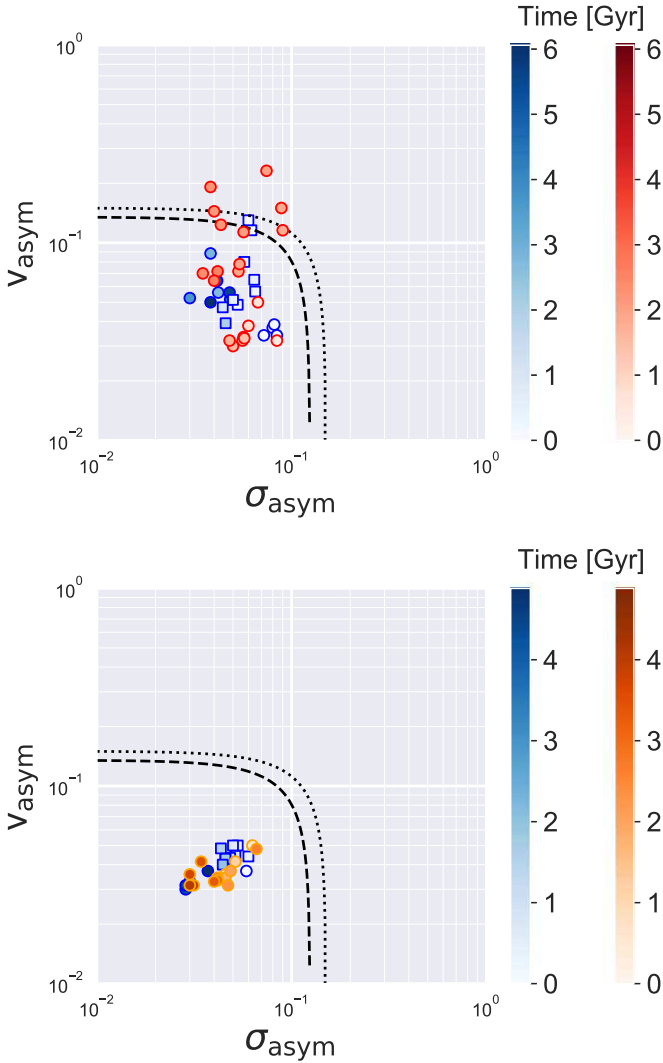


Figure 14. Time evolution of merging (red and orange) and matched nonmerging (blue) galaxies for the q0.5_fg0.3 (top, red) and q0.2_fg0.3_BT0.2 (bottom, orange) simulations on the $v_{\text{asy}} - \sigma_{\text{asy}}$ diagram. The blue squares indicate the matched isolated sample of galaxies, while the blue circles denote pre- and post-mergers. Here, we show the pre-standardized predictor values. The time begins at zero, and progresses in units of Gyr. The $K_{\text{asy}} = 0.15$ (dotted) and 0.135 (dashed) threshold lines are taken from Hung et al. (2016) and Bellocchi et al. (2012), respectively, where galaxies above the diagnostic lines are classified as merging.

dispersion map is rare in observations (e.g., Krajnović et al. 2011), and is associated with the co-addition of a counter-rotating stellar disk, produced via retrograde 1:1 mass ratio mergers in simulations (e.g., Jesseit et al. 2007; Bois et al. 2011). Therefore, these statistics may not be as useful for identifying the more typical types of mergers, which are often more unequal in mass ratio, and do not occur under idealized conditions.

4.2. What Can We Learn from the Most Important LDA Predictors about the Kinematics of Stars in Mergers?

Here, we examine the most important predictors in the LDA for all simulations, and make connections between these predictors and the dynamical evolution of stars during a merger. We split the discussion by predictor, and for brevity, we focus only on the leading predictors presented in Table 3. The top predictors include λ_{R_e} , $\mu_{1,\sigma}$, $\mu_{2,\sigma}$, $|\mu_{3,\sigma}|$, and $\mu_{4,V}$.

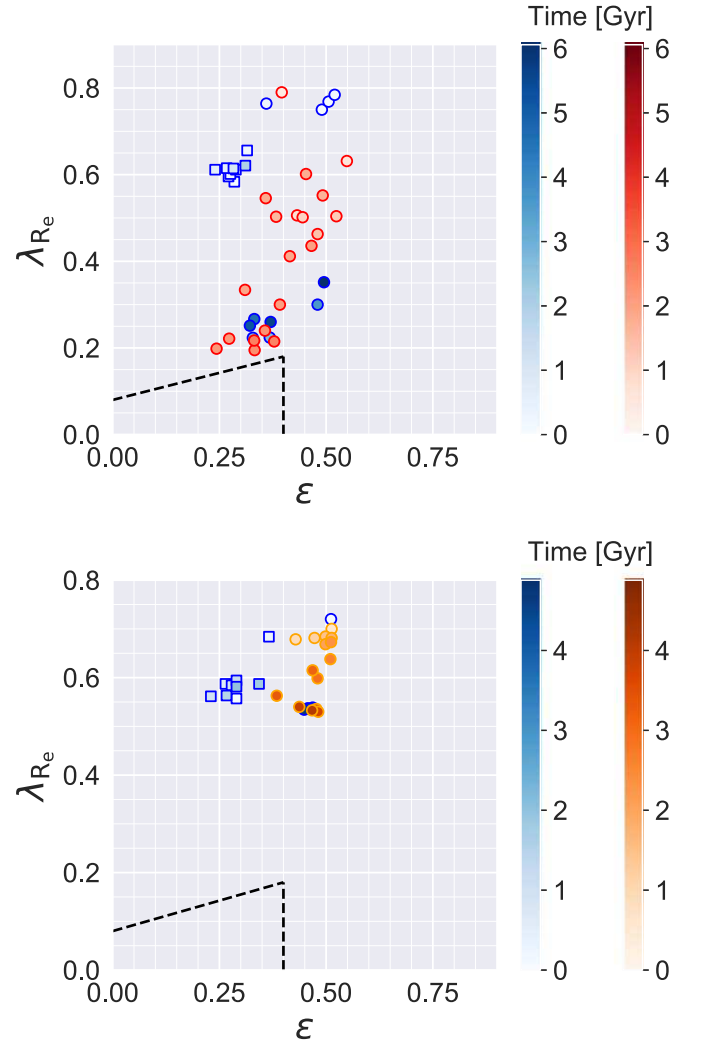


Figure 15. As Figure 14, but for the time evolution of merging (red and orange) and nonmerging (blue) galaxies for the q0.5_fg0.3 (top, red) and q0.2_fg0.3_BT0.2 (bottom, orange) simulations, on a $\lambda_{R_e} - \epsilon$ diagram. As the merger progresses (red points), the galaxies evolve toward decreased values of λ_{R_e} , corresponding to increasing levels of disorder in the kinematic maps. Slow rotators, defined by Cappellari (2016), fall below the dashed line on these plots.

4.2.1. The Approximate Spin Parameter Tracks a Global “Slowdown” in the Velocity Maps of the Major Mergers

The approximate spin parameter (λ_{R_e}) is a key predictor for all merger simulations, and is especially important for the major mergers. The angular momentum of merging galaxies is therefore significantly different from that of the nonmerging population; this effect is more apparent in major mergers. In this section, we examine how λ_{R_e} changes with time for the various simulations, and how this compares to previous works.

Firstly, we directly examine the pre-standardized values of λ_{R_e} and ϵ for the q0.2_fg0.3_BT0.2 and q0.5_fg0.3 simulations in Figure 15. In this diagram, we indicate the “slow rotator” territory, located in the lower left corner. The $\lambda_{R_e} - \epsilon$ diagram is often used to kinematically distinguish the slow rotator population of early-type galaxies from the fast-rotating population. This fast-slow rotator distinction probes the evolutionary histories of galaxies, based on disk assembly (see Cappellari 2016 for a review).

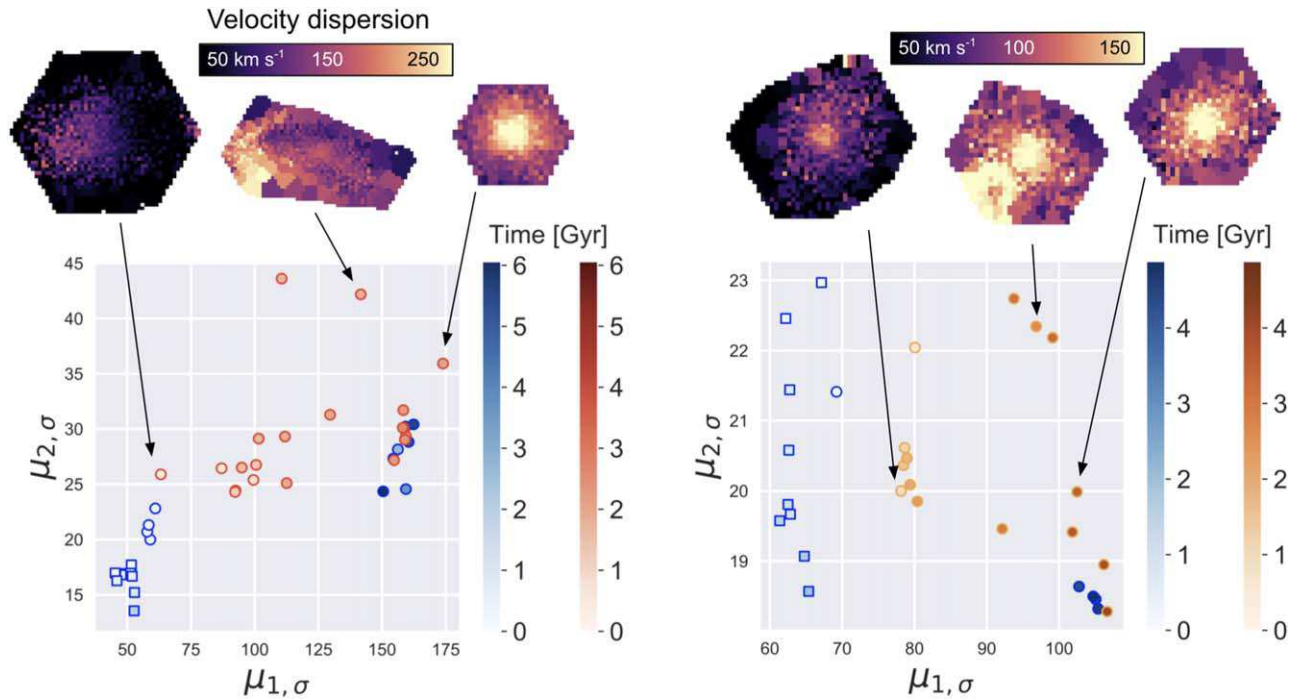


Figure 16. Time evolution of the mean values of $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ for the q0.5_fg0.3 (left plot) and q0.2_fg0.3_BT0.2 simulations (right plot). We indicate the time evolution of the merging galaxies with red and orange points, and of the nonmerging galaxies with blue points. The standalone isolated galaxies are squares, and the pre- and post-merger isolated galaxies are circles. In addition, above each plot, we include representative velocity dispersion maps from key snapshots. We find that the major mergers (left plot) tend to show a consistent evolution in $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ with time; both values increase as the stellar bulge component is built. We include three example velocity dispersion maps (above each plot), for the early (left), late (middle), and post-coalescence (right) stages of a merger, respectively. The late and post-coalescence snapshots have elevated values of $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$; during the late stage, the area between the two nuclei has an enhanced velocity dispersion (middle), and during the post-coalescence stage (right), the center of the galaxy has a larger velocity dispersion value. On the other hand, minor mergers (right plot) show an increase in $\mu_{1,\sigma}$ with time, but there is no significant change to $\mu_{2,\sigma}$ with time. While both types of mergers contribute to a stellar bulge component, the change to the dispersion maps of the major mergers is more dramatic and global.

Much recent work has focused both on examining the observed populations of fast and slow rotators, and on making predictions as to how merging galaxies move through this territory. For instance, Naab et al. (2014) utilized cosmological merger tree simulations to show that major mergers significantly affect the angular momentum content of a galaxy; they can either spin up or spin down the remnant. In our case, all of the simulated galaxies begin with a λ_{R_e} value of ~ 0.7 . For the major mergers, the λ_{R_e} value dramatically decreases, to the boundary of the slow-rotator region. This confirms that major mergers can dramatically affect the kinematic properties of the remnant, kinematically transforming the galaxy from one that can be described as disk-dominated to one that is still rotating, but is dispersion-dominated.

4.2.2. The Mean and Dispersion of the Velocity Dispersion Distribution ($\mu_{1,\sigma}$ and $\mu_{2,\sigma}$) Track the Growth of a Stellar Bulge Component

The overall importance of the $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ predictors reflects the fact that the velocity dispersion map is informative for identifying mergers. We examine how these two predictors evolve with time during a merger in Figure 16, for the q0.5_fg0.3 and q0.2_fg0.3_BT0.2 mergers. Here, we present the average value for all of the viewpoints of a given snapshot. We also include representative velocity dispersion maps for a handful of informative snapshots.

For all merger simulations, the $\mu_{1,\sigma}$ value increases throughout a merger, tracing the assembly of a stellar bulge

component. This increase is more dramatic for the major mergers, which increase to a $\mu_{1,\sigma}$ value of $\sim 200 \text{ km s}^{-1}$. Even for minor mergers, the merger incites growth in the central velocity dispersion with time. This enhancement is still present 0.5 Gyr after coalescence, such that isolated post-coalescence stages are mixed with the merger snapshots along the $\mu_{1,\sigma}$ axis in Figure 16. The signatures of bulge growth are therefore dynamically long-lived, as opposed to imaging features that fade quickly with time following a merger (i.e., in N19, the imaging predictors fade within 0.5 Gyr of final coalescence).

The $\mu_{2,\sigma}$ predictor serves different roles in major versus minor mergers, as reflected in the differing evolution of $\mu_{2,\sigma}$ values with time. An increase of $\mu_{2,\sigma}$ with time for major mergers traces the presence of two kinematic components by capturing the “bridge” of higher velocity dispersion values between two merging galaxies. This is formed by two overlapping counter-rotating features. Moreover, post-coalescence, major mergers exhibit more significant bulge growth, which is reflected both in the enhancement in $\mu_{1,\sigma}$, and in an increase in $\mu_{2,\sigma}$, since the entire distribution is broadened in this process.

Minor mergers show less change in the value of $\mu_{2,\sigma}$ with time. While $\mu_{2,\sigma}$ is still informative (since it increases during the late stages of the merger to track the bridge of higher velocity dispersion), it does not continue to increase into the post-coalescence stages. This could indicate that a smaller fraction of stars are involved in the buildup of the stellar bulge in the case of minor mergers.

4.2.3. The Skewness of the Velocity Dispersion Distribution ($|\mu_{3,\sigma}|$) Identifies Secondary Kinematic Components

The skewness in the velocity dispersion distribution, $|\mu_{3,\sigma}|$, is an important predictor for major mergers; it is sensitive to secondary kinematic components, and to disturbances in the velocity dispersion maps. For instance, the stellar bulge region has a higher dispersion, which manifests itself as a small wing on the velocity dispersion distribution (the main contribution to this distribution is the disk rotation component). In this case, the skewness predictor identifies similar features to the $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ predictors.

The $|\mu_{3,\sigma}|$ predictor is also important for identifying early-stage mergers. These snapshots tend to have low values of $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$. Since they have undergone their first pericentric passage, they exhibit a slight enhancement in the velocity dispersion map in the area of the primary galaxy, which is perturbed by the merger. An example of such a snapshot with this type of velocity dispersion enhancement is the leftmost galaxy in the q0.5_fg0.3 panel in Figure 16. This galaxy is classified as merging by this classification, where the most important predictor leading to this decision is $|\mu_{3,\sigma}|$.

4.2.4. The Kurtosis of the Velocity Distribution ($\mu_{4,v}$) Identifies the Superposition of Two Merging Galaxies

The kurtosis of the velocity distribution, $\mu_{4,v}$, is important for both of the individual minor mergers, as well as the combined minor merger classification. This predictor is sensitive to perturbations in the velocity field, specifically in cases where there are high velocities in the velocity distribution (due to the superposition of two merging nuclei), the kurtosis becomes more negative, due to the flattening of the distribution. The $\mu_{4,v}$ predictor is significant because it is able to track smaller changes in the velocity distribution, as opposed to global disruptions, as in the case of the major mergers. It is a significant predictor for minor mergers, since velocity dispersion distributions do not change dramatically during a minor merger. Instead, the LDA must rely upon the extreme velocities caused by the superposition of a secondary nuclei.

4.3. The Classification Changes with Mass Ratio

Past studies have investigated how the properties of simulated mergers affect the kinematic predictors. Hung et al. (2016) investigated this phenomenon in relation to a set of simulated merging galaxies with mass ratios of 1:1 and 1:4. They found that the merger signatures in kinematics are most apparent for the 1:1 major merger, where they may be visible for up to twice as long as for the 1:4 major merger. While there are significant differences between the work in this paper and that of Hung et al. (2016) (i.e., we perform a full RT and create mock IFS maps, whereas Hung et al. (2016) use the GADGET-3 particle velocities to create velocity maps), we also find that the classification differs significantly with mass ratio.

The major and minor merger classifications are different in several ways. The minor merger classifications have performance statistics that are $\sim 10\%$ – 30% lower. The minor mergers are also more unstable in the terms that are selected, meaning that the coefficient values fluctuate slightly when the classification is rerun. The LDA for the minor mergers therefore uses a small sample of key snapshots to create the classification. If these snapshots are excluded from the training

set, and instead fall in the CV set, then the classification is slightly different. The overall effect is that the minor mergers are less stable, and many of the selected terms have similar coefficient values, making it difficult to assess which are the most important.

Another difference between the major and minor merger classifications is that they are composed of different predictors. While some predictors, such as $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ are important for all classifications, λ_{R_e} is more important for major mergers and $\mu_{4,v}$ is more important for minor mergers. As we discuss in Section 4.2, both major and minor mergers demonstrate bulge growth, which leads to an enhancement in $\mu_{1,\sigma}$, but the change is more apparent in major mergers. The global kinematic properties of major mergers are more significantly impacted; this includes the λ_{R_e} predictor, which traces a global slowdown in the velocity maps of the major mergers. On the other hand, the minor mergers are most sensitive to smaller-scale changes in the kinematic maps; this can be traced by predictors such as $\mu_{4,v}$, which track the superposition of the secondary stellar nuclei.

4.4. The Predictors Evolve Non-monotonically with Time; the LDA Incorporates This Behavior with Interaction Terms

Many of the kinematic predictors in this analysis evolve with time throughout the merger. In most cases, this evolution is non-monotonic, meaning that merging galaxies evolve back and forth in predictor space as a function of merger time. The LDA technique accounts for this behavior using interaction terms.

An example of an interaction term in action is the $\mu_{\sigma} * \sigma_{\sigma}$ term, which has a negative coefficient for several of the major merger simulations. This means that if the value of $\mu_{1,\sigma}$ is relatively large, then $\mu_{2,\sigma}$ must be relatively small in order for the merger probability to increase. The opposite is also true: if $\mu_{1,\sigma}$ is relatively small, then $\mu_{2,\sigma}$ must increase. To be clear, “relatively large” and “relatively small” refer to the standardized values of these predictors, which are measured relative to the distribution of values for the entire merger. As such, if a term is relatively small, the predictor value will be negative. For instance, if $\mu_{1,\sigma}$ is large and $\mu_{2,\sigma}$ is small, then the term becomes: (– coefficient) * (positive) * (negative) = positive value = increase in LD1.

Consider the $\mu_{\sigma} - \sigma_{\sigma}$ diagram for the q0.5_fg0.3 merger, presented in the left panel of Figure 16. The $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ predictors are correlated, so they occupy the diagonal space of this diagram. This correlation and the interaction term function so that the center of the $\mu_{\sigma} - \sigma_{\sigma}$ diagram is the “merger territory.” This picture is somewhat complicated by the fact that there are other terms in the LDA, but if the $\mu_{\sigma} * \sigma_{\sigma}$ interaction term has a large coefficient, then this interpretation generally applies.

The interaction terms account for the non-monotonic evolution of the predictor values with time. When we create a classifier with only the primary predictors, it is fundamentally different from the LDA with interaction terms. The linear classifier is inaccurate, classifying many of the isolated post-merger snapshots as mergers. When the LDA is forced to generalize to one direction of movement across these diagrams in the monotonic case, it loses key information.

When the interaction terms are included, the LDA becomes sensitive to the values of the other predictors, and is therefore able to create “if-then” cases for the predictor space. For

example, if a galaxy is in the later stages of merging (characterized by significant bulge growth and a large value of $\mu_{1,\sigma}$), then to avoid ambiguity with post-merger isolated galaxies, the classification would like to see a relatively negative standardized value of $\mu_{2,\sigma}$ in order to classify the galaxy as merging. On the other hand, if the galaxy has a relatively small value of $\mu_{1,\sigma}$, then $\mu_{2,\sigma}$ should be large in order for the galaxy to be classified as merging. An example of this type of galaxy can be seen in the early stages of a merger, where the bulge growth has not yet begun, so $\mu_{1,\sigma}$ is small. However, the galaxy has undergone its first pericentric passage, and is experiencing an enhancement in velocity dispersion, which leads to an increase in $\mu_{2,\sigma}$. An example of this type of galaxy is shown in the left panel of Figure 16.

We conclude that it is critical to include the interaction terms in the LDA classification. Not only do they improve the performance of the technique, but they are physically motivated by the non-monotonic evolution in kinematic predictors over the course of a merger lifetime.

4.5. The Observability of a Merger Varies with Time; Mergers Are Missed during the Early Stages when the Kinematics Are Disk-like

In Section 3.6, we present the observability timescales for the various simulated mergers. We conclude that the kinematic LDA technique lengthens the observability timescale of the simulated mergers as compared to that measured from the individual predictors. Here, we focus specifically on how the observability of a merger changes with the merger stage, and we refer the reader to Figure 7 for a useful visualization of the mean value of LD1 with time in all of the simulated mergers.

When we examine the observability of the mergers in terms of the predefined early-, late-, and post-coalescence stages, we notice several stage-related differences. Firstly, during the early stages of merging, some simulations show a larger standard deviation in the LD1 values. For all simulations, the value of LD1 tends to fall below the decision boundary during these early stages. Hung et al. (2015) found that using kinematic predictors to identify mergers results in a significant fraction of false negatives from epochs where the merger is indistinguishable from a rotating disk. We also find that a significant fraction of false negatives occur during the early stages, where the rotation is indeed disk-like.

During the late stage of the merger, the minor mergers do not show much variation from one snapshot to the next; the LD1 values are relatively flat. On the other hand, the major mergers, in particular q0.5_fg0.3, show variation between the late-stage snapshots. For q0.5_fg0.3, this factor contributes to the relatively short observability timescale. These changes in LD1 values are significantly greater than the variance due to different viewpoints. The kinematic features are therefore seen to change significantly and rapidly with time for some of the major mergers during their late stages. We also find that most simulations have relatively high LD1 values during the late stage of a merger. The late stage of a merger is therefore characterized by short-lived dramatic kinematic features. This is consistent with Hung et al. (2015), who found that kinematic tracers of mergers tend to be most informative during the late stage of the merger, which is when the imaging predictors are also most useful.

As the merger progresses into the post-coalescence and post-merger isolated stages, we find that the LD1 value is more

stable, which is characteristic of longer-lived kinematic features. The LD1 value does not significantly decline during the post-merger stages. We focus on the kinematics of the post-coalescence and post-merger stages in Sections 4.6 and 4.7.

4.6. When Does a Merger End? The Kinematic Disturbances due to Mergers Are Long-lived

In N19, we define the end of the merger as 0.5 Gyr after final coalescence. This cutoff was selected so that the imaging predictors, and therefore the value of LD1, would decay smoothly until the end of the merger. The imaging technique is therefore very accurate and precise during the transition from a post-coalescence merger to an isolated post-merger galaxy. In contrast, in this work, the kinematic predictors, and therefore the LD1 value, remain elevated during the period of post-merger isolated snapshots. There are visually apparent warps in the velocity maps, and the velocity dispersion maps have elevated values of $\mu_{1,\sigma}$ and $\mu_{2,\sigma}$ for post-merger isolated galaxies. Hung et al. (2016) also found a persistence of kinematic merger signatures up to ~ 2 Gyr following coalescence. We find that the kinematic disturbances fade 2–2.5 Gyr after coalescence, meaning that in order to improve the classification, we would need to significantly extend the post-coalescence phase.

Rather than changing the definition of the end of a merger, a more relevant task might be to define the merger more specifically by stage. For instance, if it is a priority to distinguish post-coalescence from post-merger isolated galaxies (post-coalescence snapshots occur immediately following coalescence until 0.5 Gyr after coalescence, and post-merger snapshots occur >0.5 Gyr after coalescence), we should rely on the imaging predictors, which do a better job in this specific case. In our future work we plan to combine the imaging and kinematic tools, and address this question more directly. One potential path forward could be to create separate classifications to target different stages of a merger. This could provide a more flexible definition of merging galaxies, and allow other users of the technique to target stages of interest in a merger.

4.7. The Kinematics of the Post-merger Stages Track the Growth of a Stellar Bulge and a Kinematically Decoupled Core for the 1:2 Mass Ratio Merger

Here, we focus on the kinematics of the post-coalescence and post-merger stages. During these stages we observe the buildup of a central component in the stellar velocity dispersion maps. The change is more dramatic for major mergers, and can best be explained by tracing the growth of a stellar bulge component. Hopkins et al. (2009) investigated the effect of mass ratio on the merger remnant, finding that the fraction of the primary stellar disk that is relaxed into the bulge is directly proportional to the mass ratio of the merger. This supports the hypothesis that a stellar bulge is built in the post-merger stages, since we would expect major mergers to contribute a larger fraction of stars to the bulge component.

The q0.5_fg0.3 merger has unique kinematic features in its stellar velocity map during the post-merger stages, so we focus the remainder of our discussion on this merger. We present several post-coalescence and post-merger snapshots from the q0.5_fg0.3 merger in Figure 17. The stellar velocity maps are particularly intriguing, as they have a distinct central

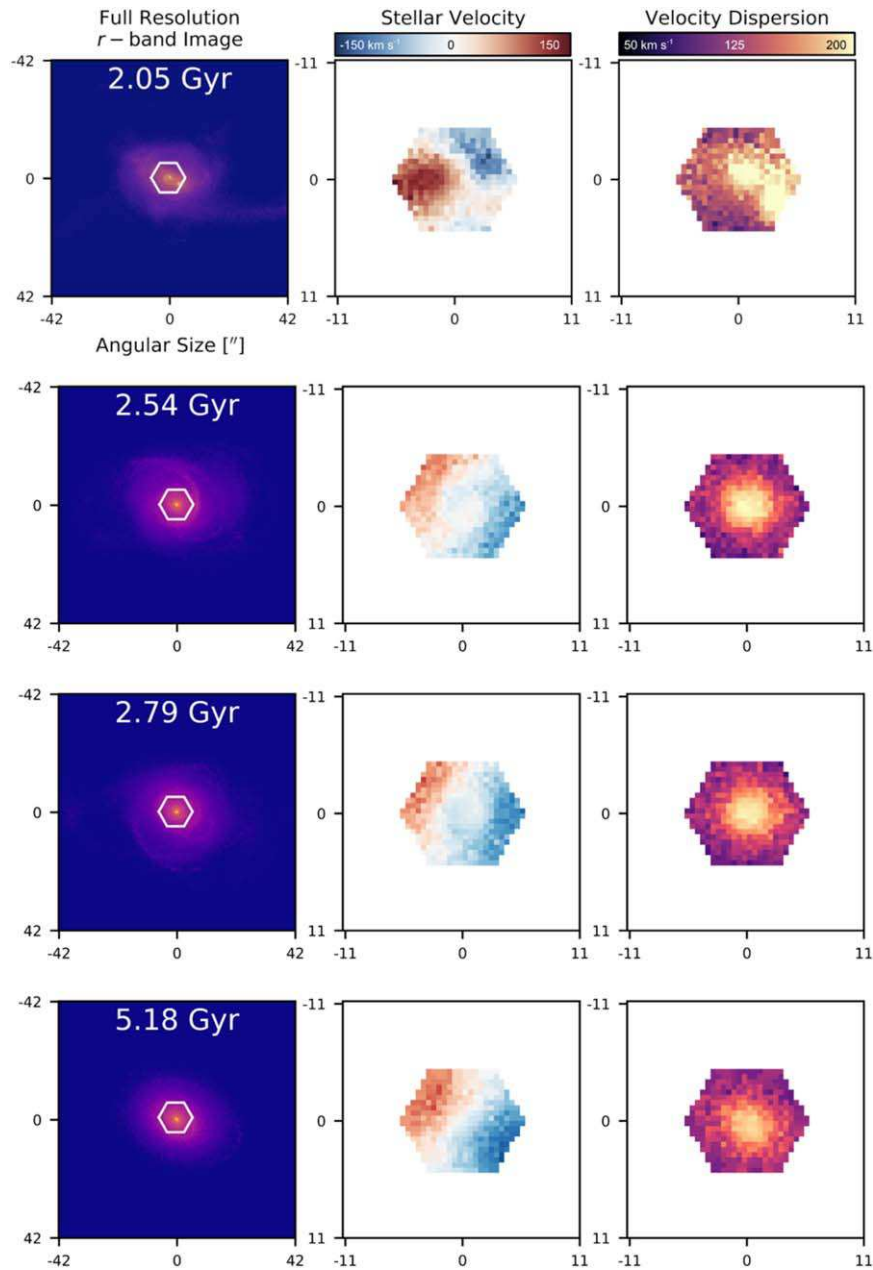


Figure 17. Evolution of the long-lived kinematically decoupled feature in the r -band image (left), stellar velocity map (middle), and stellar velocity dispersion map (right) of the $q0.5_fg0.3$ merger. The top panel depicts the last snapshot before coalescence. At 2.54 Gyr the galaxy is in the post-coalescence stage, and at 2.79 and 5.18 Gyr the galaxy is in the post-merger isolated stages. The central kinematically decoupled component appears in the velocity map at around 2.54 Gyr, i.e., ~ 0.4 Gyr after coalescence, and does not disappear until ~ 3 Gyr after coalescence.

component that appears in the post-coalescence phase, and persists into the post-merger phase. This central feature in the stellar velocity maps is spatially coincident with the bulge-like feature in the stellar velocity dispersion maps. It is not fully counter-rotating, but is misaligned from the main stellar disk. We therefore hypothesize that we have discovered a decoupled kinematic component.

Previous theoretical works have predicted that major mergers with mass ratios of 1:1 or 1:2 may produce this type of intriguing kinematic component (e.g., Bendo & Barnes 2000; Jesseit et al. 2007; Crocker et al. 2009). For instance, Bendo & Barnes (2000) and Jesseit et al. (2007) found that equal mass simulated mergers display a far wider range of kinematic features, including counter-rotating cores and global misalignments,

whereas unequal mass mergers tend to have disk-like kinematics. The merger remnants with complex kinematics are intriguing, because these decoupled central components have also been discovered in observational studies. For example, the ATLAS^{3D} survey found that a significant fraction of slow-rotating ETGs have decoupled kinematic components (e.g., Emsellem et al. 2011).

Our finding of a decoupled kinematic core in the remnant of the 1:2 ($q = 0.5$) mass ratio merger supports these past findings, i.e., that mergers with a large mass ratio can produce dramatic central features in the kinematic maps. Furthermore, this result suggests that selecting galaxies with kinematically decoupled cores would produce a sample of post-coalescence major mergers with a mass ratio $q \gtrsim 0.5$.

4.8. The Kinematic LDA Is Not as Good at Identifying Merging Galaxies as the Imaging Technique

The kinematic LDA technique gives rise to a significant number of false negatives, which drive down the accuracy and recall of the technique. This can be partially ascribed to the chosen priors, which skew the classification toward minimizing false positives (see Section 3.5 for a full discussion). As we discuss in Section 3.7, this is also due to the lack of identifiability of certain snapshots as mergers, meaning that they are indistinguishable from nonmerging disks and/or post-merger remnants, both visually, and based on their kinematic predictors.

The imaging LDA performs better for all runs, with improvements in all performance statistics. This means that fewer galaxies in total will be correctly classified by the kinematic technique, due to the shortcomings listed above. While the major merger combined run is slightly improved when performed using imaging predictors (this improves the accuracy/recall/F1 score by 9%/11%/7%), the minor merger combined run experiences significant improvement (16%/41%/25%). The kinematic minor merger combined classification therefore scores lower across all performance statistics, relative to both the imaging minor merger classification and the kinematic major merger classification. This reflects the specific inability of the kinematic classification to identify minor mergers.

The imaging LDA also has longer observability times on average, ranging from observability times of 2.2–2.8 Gyr for major mergers, and 3.5–9.2 Gyr for minor mergers. The kinematic LDA has observability times of 0.9–2.4 Gyr for the major mergers, and 3.0–6.6 Gyr for the minor mergers. As such, while the observability times for the kinematic LDA significantly improve upon the observability times for individual kinematic predictors, the mergers are still observable for slightly longer timescales with the imaging LDA.

This has important implications for the relative capabilities of imaging versus kinematic predictors in identifying merging galaxies. The stellar kinematics of mergers take longer to exhibit disturbance, and remain disturbed for a longer time following a merger. The imaging predictors are better contained to the duration of the merger, and better able to identify all of the different merger stages. Hung et al. (2016) recommended using kinematic predictors in combination with imaging predictors, due to the frequency of false negatives in their investigation of kinematically identified merging galaxies. We support this conclusion, finding that the kinematic predictors have failure modes that may be ameliorated by incorporating imaging predictors.

However, there are some advantages of kinematic predictors relative to imaging predictors. In Sections 4.6 and 4.7, we find that the kinematic predictors are particularly useful for identifying post-coalescence and post-merger stages, because the kinematic disturbances persist long after the imaging predictors fade. The inference is that while imaging predictors are more informative overall, kinematic predictors are powerful in certain domains, and provide additional information. If forced to select between imaging and kinematic classification methods, the imaging approach is better. However, the best overall approach is to combine the two techniques into one imaging + kinematic classification. We plan to discuss this topic further in future work.

4.9. Applying the Technique to MaNGA IFS in Future Work

In Section 3.10, we discuss the implications of creating a kinematic classification from a suite of simulations with a narrow range in stellar mass ($3.9 \times 10^{10} < M_{\odot} < 4.7 \times 10^{10}$), and in an initial B/T ratio (0–0.2). Many of the kinematic predictors, specifically the most important predictors for the major merger simulations, also probe the intrinsic properties of galaxies. In other words, while these predictors are useful in identifying major mergers, they also change as a function of stellar mass and morphology. The implication is that a classification created from disk-dominated intermediate-mass mergers may not extrapolate well to the MaNGA sample, which spans a wider range in stellar mass ($10^8 < M_{\odot} < 10^{11}$) and morphology.

At present, it is unclear whether this will be a concern only for extreme cases, i.e., the most massive bulge-dominated ETGs, or if it will also cause concern in systems with a mix of rotation and dispersion support in their kinematics. We have preliminarily investigated the distribution of bulge- versus disk-dominated galaxies in MaNGA; while Wang et al. (2020) have found that MaNGA galaxies are predominantly disk-dominated, Graham et al. (2018) found that a significant fraction of MaNGA galaxies (across all masses) are bulge-dominated.

Since the MaNGA sample includes a diversity of different galaxy types, we plan to tread carefully when we apply the classification. One option may be to select those MaNGA galaxies with high values of λ_{R_e} to include in the classification; in this way, we could exclude bulge-dominated galaxies. Another option could be to de-emphasize certain domains of predictor space in the classification, or to remove those kinematic predictors that are most sensitive to intrinsic galaxy properties altogether. The details of this approach will be developed in future work, as they fall beyond the scope of this paper, which focuses mostly on the creation of the merger identification technique.

5. Conclusions

In this work, we have built on the standalone imaging merger classifier in N19 to create a parallel LDA classifier, utilizing kinematic predictors to identify merging galaxies. To produce the classification, we used SUNRISE synthetic spectra from GADGET-3 simulated merging galaxies to create mock “MaNGA-ized” data cubes. We convolved and rebinned the synthetic spectra to the spatial and spectral resolution of MaNGA, introduced noise, and implemented the Voronoi binning scheme used for the MaNGA data cubes. With `ppxf`, we extracted stellar velocity and stellar velocity dispersion maps for each data cube.

We then measured a number of kinematic predictors from the velocity and velocity dispersion maps. We used a random forest regressor (RFR), followed by the LDA classifier to select the most informative kinematic predictors and to carry out the classification. The selected predictors are: the difference between the kinematic PA and the imaging PA (ΔPA), the kinemetry residuals (`resid`), the approximate spin parameter (λ_{R_e}), the asymmetry in the Radon profile (A_2), and the moments of the velocity and velocity dispersion distributions ($\mu_{1,v}$, $\mu_{1,\sigma}$, $\mu_{2,v}$, $\mu_{2,\sigma}$, $|\mu_{3,v}|$, $|\mu_{3,\sigma}|$, $\mu_{4,v}$, and $\mu_{4,\sigma}$). We then ran the LDA as a classifier for all simulations individually, as well as for the combined major merger simulation and the combined minor merger simulation.

We first used the LDA classification as an agnostic approach to determine the most useful kinematic predictors for identifying different types of mergers. Our main conclusions are:

1. Many kinematic predictors used in previous work to identify mergers are not as useful as those in this work (i.e., the deviation of the velocity and velocity dispersion maps from ordered rotation v_{asym} and σ_{asym} , respectively, and ΔPA). These predictors are sensitive to specific stages of equal mass ratio mergers, and are not as sensitive to the full range of merger parameters and stages used in this simulation suite (Section 4.1).
2. The mean and variance of the values in the velocity dispersion maps ($\mu_{1,\sigma}$ and $\mu_{2,\sigma}$, respectively) are the most useful predictors for identifying mergers across all simulations, because they are sensitive to the growth of a stellar bulge component during mergers (Section 4.2).
3. The selected predictors differ as a function of mass ratio. The major mergers exhibit large-scale kinematic changes (i.e., a global slowdown of the rotation), and therefore rely more on predictors such as λ_{R_e} . The minor mergers are identified using predictors such as $\mu_{4,V}$, which trace the superposition of a secondary stellar nucleus (Section 4.3).

We have also examined the performance of the LDA classification, measured using the four performance statistics (accuracy, precision, recall, and F1 score), as well as the observability timescale. Our main findings are:

1. The LDA performance significantly improves when the interaction terms are included. These terms are able to account for the non-monotonic evolution of the kinematic predictors with time (Section 4.4).
2. By combining many different kinematic predictors, we create a classification where the observability timescale is a large fraction of the overall merger time (40%–90%). This corresponds to mergers that are observable for 0.9–6.6 Gyr, and represents an improvement on the observability timescale as compared to any of the individual kinematic predictors (Section 4.5).
3. The sensitivity of the LDA technique varies with epoch during mergers. We find that there are more missed mergers (i.e., false negatives) during the early stage of a merger, where the stellar kinematics are disk-like. The mergers are at their most detectable during the late and post-coalescence stages (Section 4.5).
4. The kinematic predictors (and the LD1 value) are long-lived, and remain elevated for ~ 2 Gyr following final coalescence. The stellar kinematics of the post-coalescence and post-merger epochs capture the formation of a stellar bulge component (Section 4.6).
5. For the (major, gas-rich) q0.5_fg0.3 merger, a kinematically decoupled component is visible in the stellar velocity maps (Section 4.7).
6. The imaging classification performs better than the kinematic classification, and the improvement is larger ($\sim 15\%$ increase in accuracy, recall, and F1 score) for minor mergers. The kinematic LDA can be improved by adding imaging predictors (Section 4.8).
7. The kinematic predictors add unique information about merging galaxies to the toolkit; for instance, the kinematic classification is better at identifying post-coalescence and post-merger galaxies, relative to the imaging classification (Section 4.8).

8. The kinematic classification is created from a suite of simulations that are limited in their scope (i.e., the simulated galaxies are disk-dominated, and span a range of $(3.9\text{--}4.7) \times 10^{10} M_{\odot}$ in stellar mass). We conclude that the results may not be applied to all MaNGA galaxies (which have a range of morphologies, and an approximately flat stellar mass distribution of $10^8 < M_{\odot} < 10^{11}$). We plan to further address this concern in future work (Section 4.9).

In R. Nevin et al. (2021, in preparation) we will combine the kinematic classification with the imaging classification presented in N19, and apply the classifier to MaNGA galaxies. At this point, we will release the python tools for implementing these classifications. These tools are designed to be adaptable to the specifications of other imaging and/or IFS surveys, with the goal of applying the classification to other IFS surveys—e.g., SAMI, CALIFA, HECTOR.

In R. Nevin et al. (2021, in preparation) we will further investigate whether various kinematic parameters enhance the existing imaging classifier, and why, and revisit the hyperparameter tuning to determine the optimal location of the decision boundary. We also plan to investigate the possibility of splitting the classification by merger stage. Our scientific goals include identifying how the star formation histories, metallicities, and AGN activity change for these different stages, as well as for different mass ratios of merging galaxies.

R. N. and J. M. C. are supported by NSF AST-1714503.

L. B. acknowledges support from NSF award #1715413.

J.A.V.-M. acknowledges support from the CONACyT postdoctoral fellowship program.

We thank the anonymous referee for their thorough and thoughtful comments, which have improved the quality and clarity of this paper.

This work utilized the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation, grant number ACI-1548562. We specifically utilized Comet and Oasis through the XSEDE allocation for “An Imaging and Kinematic Approach for Improved Galaxy Merger Identifications” (TG-AST130041). We would also like to acknowledge the help of Martin Kandes, who assisted with the optimization of the LDA tool.

The authors acknowledge University of Florida Research Computing for the provision of computational resources and support contributing to the research results reported in this publication. The website for HiPerGator (the well-named supercomputer) is: <http://researchcomputing.ufl.edu>

This research made use of Marvin, a core Python package and web framework for MaNGA data, developed by Brian Cherinka, José Sánchez-Gallego, Brett Andrews, and Joel Brownstein (Cherinka et al. 2019).

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration, including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, the Harvard-Smithsonian Center

for Astrophysics, Instituto de Astrofísica de Canarias, Johns Hopkins University, the Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), the National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Software: astropy (The Astropy Collaboration et al. 2013, 2018), matplotlib (Hunter 2007), mangadap (Belfiore et al. 2019; Westfall et al. 2019), numpy (van der Walt et al. 2011), openmpi (Gabriel et al. 2004), scikit-learn (Pedregosa et al. 2011), scipy (Virtanen et al. 2020), seaborn (Waskom et al. 2014), sdss-marvin (Cherinka et al. 2019), pandas (McKinney 2010).

Appendix A

The Importance of Radiative Transfer and Dust Scattering and Absorption

A central goal of this paper is to design a process to produce synthetic kinematic maps from simulated galaxies that are as close to full realism as possible. However, we acknowledge that running full radiative transfer (RT) and fitting the absorption lines of the resultant data cube is both complex and computationally expensive. Bottrell et al. (2019) explore the concept of observational realism, and the steps necessary to produce a CNN classifier for merging galaxies capable of achieving high accuracy. They find that the RT step is unnecessary, and that it can be avoided in order to achieve significant gains in computational time. Since the goal of our approach is to produce observations that closely mimic MaNGA kinematic maps, we carry out a full RT (“full”) means that we include the effects of dust scattering and absorption; in addition, we apply observational realism, as detailed in Section 2.2. Here, we focus on the importance of radiative transfer, and the proper treatment of dust.

Firstly, we explore the differences between stellar kinematics derived from the GADGET-3 particles (no RT) and those derived from the SUNRISE spectra (RT), without dust scattering and absorption. From here on, GADGET-3 derived kinematics are the “particle” kinematics, the SUNRISE spectra without dust effects are the “NONSCATTER” spectra, and the SUNRISE spectra incorporating dust scattering and absorption are the “SCATTER” spectra.¹⁸ An important caveat to our analysis of the differences between these kinematic maps relates to differences in observational effects. We apply all observational effects to the NONSCATTER and SCATTER

spectra, convolving, rebinning, and incorporating noise, but we create the particle kinematic maps without noise effects. Moreover, the particle kinematics are mass-weighted, so the convolution kernel we apply to the particle velocity and velocity dispersion maps is a rough approximation of the convolution of a flux-weighted quantity (i.e., the NONSCATTER and SCATTER spectra). We apply this convolution kernel as a final step, once the kinematic maps are generated.

In Figure 18 we present the velocity maps (top row) and the velocity dispersion maps (bottom row) produced at various stages of realism for a snapshot from the q0.5_fg0.3 merger. We begin with the kinematic maps derived from the mass-weighted particle velocities (second column), then the kinematic maps from the NONSCATTER spectra (third column), which are produced using RT, and finally the kinematic maps from the SCATTER spectra (fourth column), which also incorporates the effects of dust scattering and absorption. To a first order approximation, we find that the velocity and velocity dispersion maps are similar enough to verify that our methodology for creating stellar velocity and velocity dispersion maps from mock SUNRISE spectra is not failing. Specifically, significant differences between the SUNRISE-derived maps and the particle-derived maps could indicate bugs in SUNRISE, failures in the steps that produce mock spectral data cubes, or problems with the ppxf absorption line fitting that produces the velocity and velocity dispersion maps.

We present the difference maps between the particle and NONSCATTER, and SCATTER and NONSCATTER maps in Figure 19. We first examine the differences between the particle kinematic maps and the NONSCATTER kinematics maps, as given the first and third panels of this figure, for the stellar velocity and velocity dispersion maps, respectively. Radiative transfer, observational effects, and the fact that the particle maps are derived from mass-weighted velocities can cause differences between these maps. For instance, the NONSCATTER maps have larger bins toward the exterior of the galaxy and in addition, show more variation on a spaxel-to-spaxel basis, due to noise. The NONSCATTER maps have significantly different velocity and velocity dispersion values, particularly at the center of the galaxy. There is a faster rotating feature at the center of the particle velocity map, which manifests as a rotating feature in the leftmost panel of Figure 19. Moreover, the NONSCATTER-derived dispersion is significantly higher at this location, which is apparent as an over-subtraction in the third panel. Since the central feature is slower-rotating, and has a higher velocity dispersion, it is most likely to be related to beam-smearing. The differences in the implementation of convolution in the particle maps, which are convolved as a last step, and the NONSCATTER maps, which are convolved prior to deriving the velocity and velocity dispersion values, could contribute to the differences in these central regions.

We next evaluate the role of dust in the kinematic maps by comparing the NONSCATTER kinematic maps to the SCATTER version in the second and fourth panels of Figure 19, for the stellar velocity and stellar velocity dispersion maps, respectively. The effects of dust on the spectra of galaxies are especially important in the context of merging galaxies. For instance, nearly all ultra-luminous infrared galaxies (ULIRGs, $L_{\text{IR}} > 10^{12} L_{\odot}$) are dust-dominated major mergers, with buried starburst or AGNs (e.g., Veilleux et al. 2002; Tacconi et al. 2008). While the galaxies in this simulation suite are not ULIRGs, they still showcase kinematic effects, due to the presence of dust. For instance, the NONSCATTER maps tend to underestimate the velocity

¹⁸ When we initially began this investigation, we found that there was a small error in the public version of SUNRISE that affected the shape of computed emission and absorption lines (C. Hayward & R. Simons 2021, private communication). This error in SUNRISE occurs when kinematic shifts are applied to emission and absorption lines during the dust scattering stage, and therefore applies only to SCATTER spectra generated from dust RT with kinematics. We were able to locate and fix the bug prior to running the production-run SUNRISE simulations for this paper.

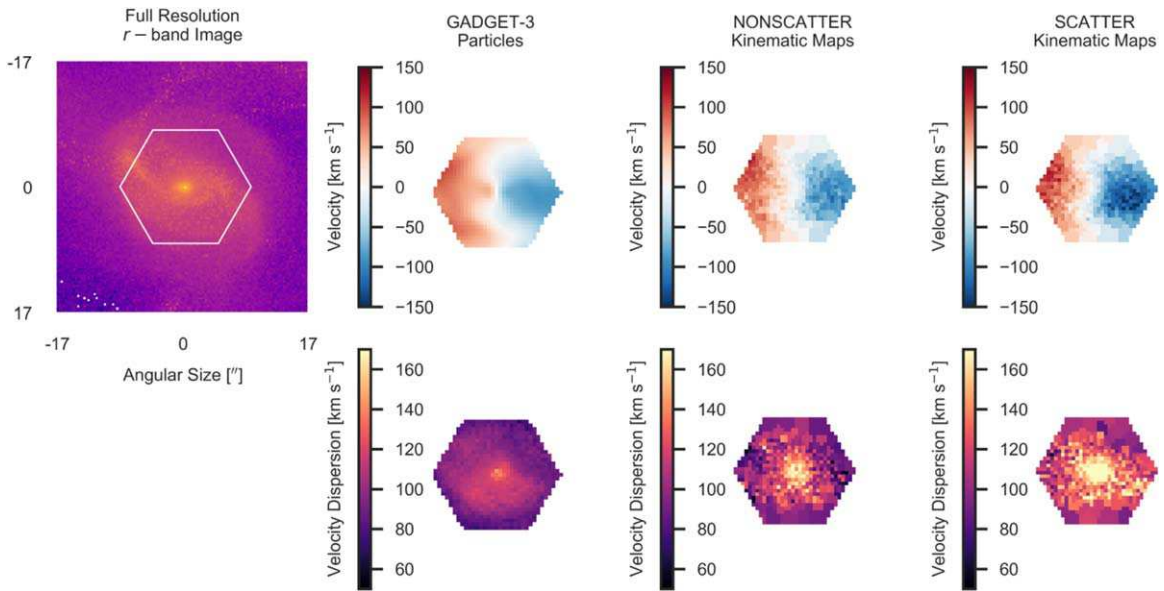


Figure 18. A comparison of the stellar velocity (top row) and velocity dispersion maps (bottom row) for an example galaxy snapshot from the q0.5_fg0.3 merger. We include the r -band image for this snapshot (without observational effects, top left panel). The GADGET-3 particle velocity and velocity dispersion map are shown in the middle left column, the kinematic maps derived from the NONSCATTER SUNRISE spectra in the middle right column, and the kinematic maps from the SCATTER SUNRISE spectra in the right column. The NONSCATTER and SCATTER kinematic maps have undergone all steps to achieve mock MaNGA observations, while the GADGET-3 derived maps have only been convolved and rebinned. Additionally, the GADGET-3 maps are a mass-weighted quantity, while all others are flux-weighted. The differences between the maps can therefore be ascribed to these effects, in addition to the effects of radiative transfer and the proper treatment of dust. The NONSCATTER and SCATTER velocity and velocity dispersion maps have significantly larger values, which is especially apparent in the center of the velocity dispersion maps. The SCATTER maps have larger values at all locations relative to the NONSCATTER maps; this points to the importance of the proper treatment of dust when creating mock stellar kinematic maps.

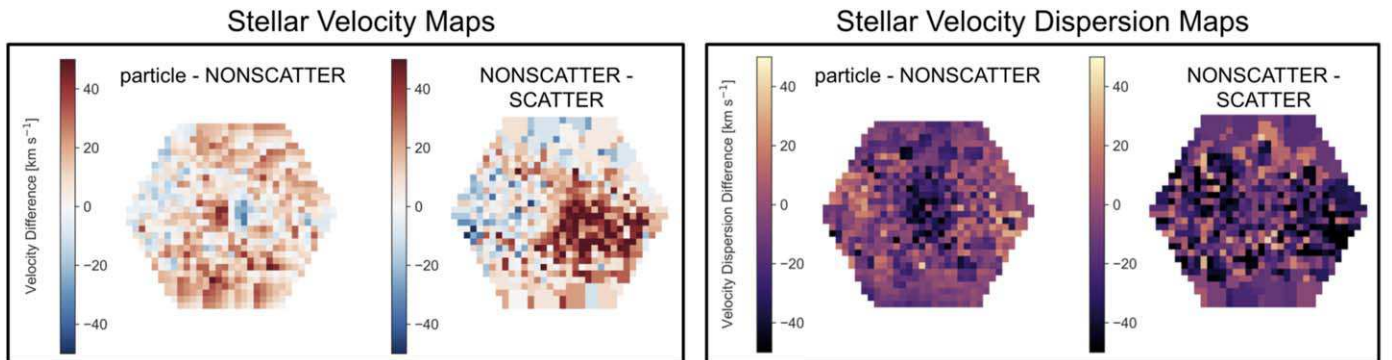


Figure 19. Difference maps for the stellar velocity (first and second panels) and stellar velocity dispersion maps (third and fourth panels). We show the difference between the particle and NONSCATTER maps, and the SCATTER and NONSCATTER maps. The particle velocity map has a faster rotating core (first panel) and a lower central velocity dispersion (third panel) relative to the NONSCATTER maps, which are both characteristic beam-smearing effects. The SCATTER maps are faster rotating, and have a larger velocity dispersion across the galaxy disk, which could be due to dust preferentially obscuring dynamically young stars.

dispersion relative to that of the SCATTER maps. This is apparent in the fourth panel of Figure 19, which has a significant number of spaxels with negative values, indicating that the SCATTER dispersion map has larger values. In addition, the difference is apparent for the majority of spaxels in the galaxy, which could indicate that dust is important on the galaxy scale. The NONSCATTER and SCATTER velocity maps also differ. The stars in the disk have higher rotational velocities for the SCATTER extension, meaning that dust may be obscuring the slower-rotating dynamically young population. We run the classification with both SCATTER and NONSCATTER extensions, finding that both the individual predictors and the classification change significantly (in terms of both the selected predictors and their coefficients); we therefore conclude that the effects of dust are important in terms of stellar kinematics.

Previous works have investigated the effects of dust in SUNRISE spectra in depth. Stickley and Canalizo (2016) compared measurements of the stellar velocity dispersion ($m\sigma_*$, from the GADGET-3 particle velocities) to the SCATTER version of the SUNRISE spectra ($f\sigma_*$), finding that the offset between the two measurements can be as large as 20%–30% for isolated galaxies, and as large as 100% for extreme cases, such as actively merging systems with disturbed morphologies. They conclude that when comparing observations to simulations, it is important to measure flux-weighted kinematics, and to incorporate dust. The dust acts to preferentially obscure younger stars, elevating measurements of $f\sigma_*$ relative to $m\sigma_*$. Stickley and Canalizo (2016) also found that the distribution of the dust is more important than the total attenuation due to dust. For the case of merging galaxies, this

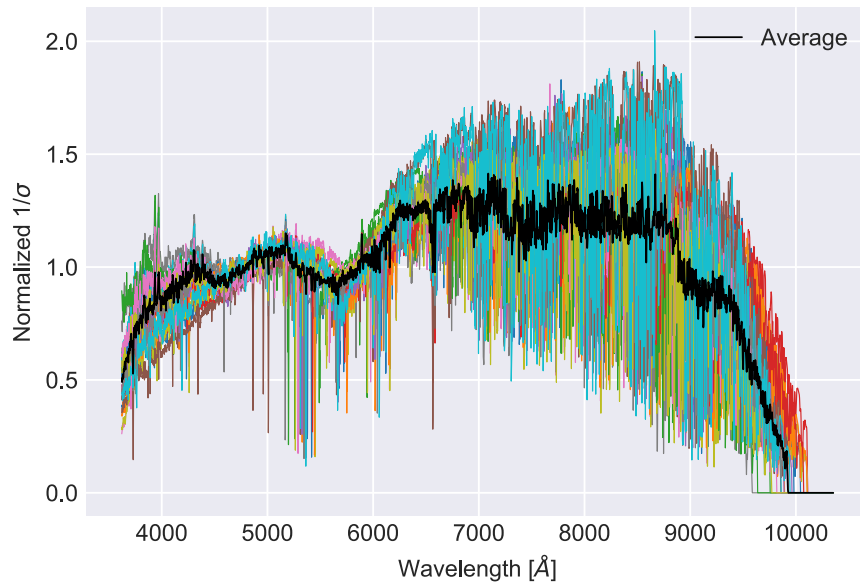


Figure 20. Trend of inverse noise ($1/\sigma$, or $\sqrt{\text{ivar}}$) with wavelength for the central spaxel of 20 MaNGA galaxies (color lines). The normalized median trend is black. The largest variation occurs from galaxy to galaxy, as opposed to spaxel-to-spaxel; as such, we take the average $1/\sigma$ spectrum (black) to use as our characteristic $1/\sigma$, which we scale to the flux of each simulated spaxel to produce an error spectrum.

message is particularly relevant; the effects of dust on merger kinematics cannot be ignored.

We conclude with two points: (1) The kinematic maps derived from particle velocities are similar to the `SUNRISE`-derived kinematic maps, indicating that the technique for creating mock stellar kinematic maps is not failing, and (2) The kinematic maps have important differences, occurring over the course of RT, the proper treatment of dust, and the subsequent mock observational steps. While we do not fully propagate the particle kinematic maps through the LDA, we find that they are significantly different in appearance from the RT’ed maps, and that the derived kinematic predictors are also affected. This confirms that in order to apply the classification to observed galaxies, we need to train the classification using galaxies that have undergone RT, are flux-weighted, and incorporate observational effects. It is beyond the scope of this paper to conduct a full analysis as regards which steps of the RT and subsequent `ppxf` derivation of the kinematic maps are of greatest importance in creating realistic galaxies, and why.

Appendix B Creating a Mock Noise Spectrum

In order to create realistic mock data cubes, we also create a mock noise data cube for each snapshot. For an IFS data cube, the typical noise trends with wavelength, and so we first measure the median inverse noise ($1/\sigma$, or $\sqrt{\text{ivar}}$) trend for the central spaxel of 20 MaNGA galaxies (Figure 20). These 20 galaxies are composed of four randomly selected galaxies from each of the five fiber bundles.¹⁹ We test how the $1/\sigma$ spectrum changes for the location of the spaxel relative to the center of the galaxy, finding that there is more variation (in the shape of

the $1/\sigma$ spectrum) between different galaxies than that found in terms of location in the galaxy. We therefore compute the median $1/\sigma$ spectrum using the central spaxel for all 20 MaNGA galaxies.

To compute the noise spectrum for each spaxel, we multiply the normalized $1/\sigma$ trend with wavelength by the g -band S/N of that spaxel, so as to determine how S/N trends with wavelength. We then divide the mock spectra for each spaxel by the S/N trend with wavelength for that spaxel to get the error (or σ) spectrum. Finally, we multiply the error spectrum by a random normal Gaussian, with a mean of zero and a standard deviation of one. This is the “realization” of the noise for that spaxel. We add the realization of the noise to the spectrum, and use the error spectrum as an input to `ppxf`.

Appendix C AGN Contamination

The framework for the simulation suite was originally created in Blecha et al. (2018) to determine the timeline of the activation and fueling of AGNs during mergers. It therefore contains broad line AGNs, which can complicate the fitting of stellar kinematics. While all of the simulations include AGNs, the `q0.5_fg0.3` simulation hosts the brightest AGN, so we focus on this simulation to describe the effects of a BL AGN on the stellar kinematics. The AGN in the `q0.5_fg0.3` simulation achieves a maximum bolometric luminosity of $L_{\text{AGN}} = 10^{46.3} \text{ erg s}^{-1}$, representing 90% of the total luminosity. The duty cycle is short for moderate and high luminosity AGNs; the AGN dominates the spectra for $<50 \text{ Myr}$. Therefore, even in the `q0.5_fg0.3` simulation, the AGN luminosity is subdominant for the majority of snapshots, even during the late stage of the merger.

Most MaNGA galaxies do not host AGNs, so the MaNGA DAP is not equipped to mask broad emission lines, and therefore fails to fit the stellar absorption lines. Bright AGNs in MaNGA galaxies are often masked by the DAP fitting procedure, leaving a hole at the center of the galaxy in the stellar kinematic maps. We demonstrate this type of failure in the middle row of Figure 21 for the snapshot of the `q0.5_fg0.3`

¹⁹ To verify whether the randomly selected galaxies constitute a fair representation of the MaNGA sample, we compare the average ivar spectrum in Figure 20 with the average ivar spectra, computed as a function of stellar mass. We find that the shape of the ivar spectrum varies little below a stellar mass of $\log M_* \sim 11$, and that the average spectrum presented in Figure 20 is representative of the ivar spectra of the majority of MaNGA galaxies below this mass cutoff.

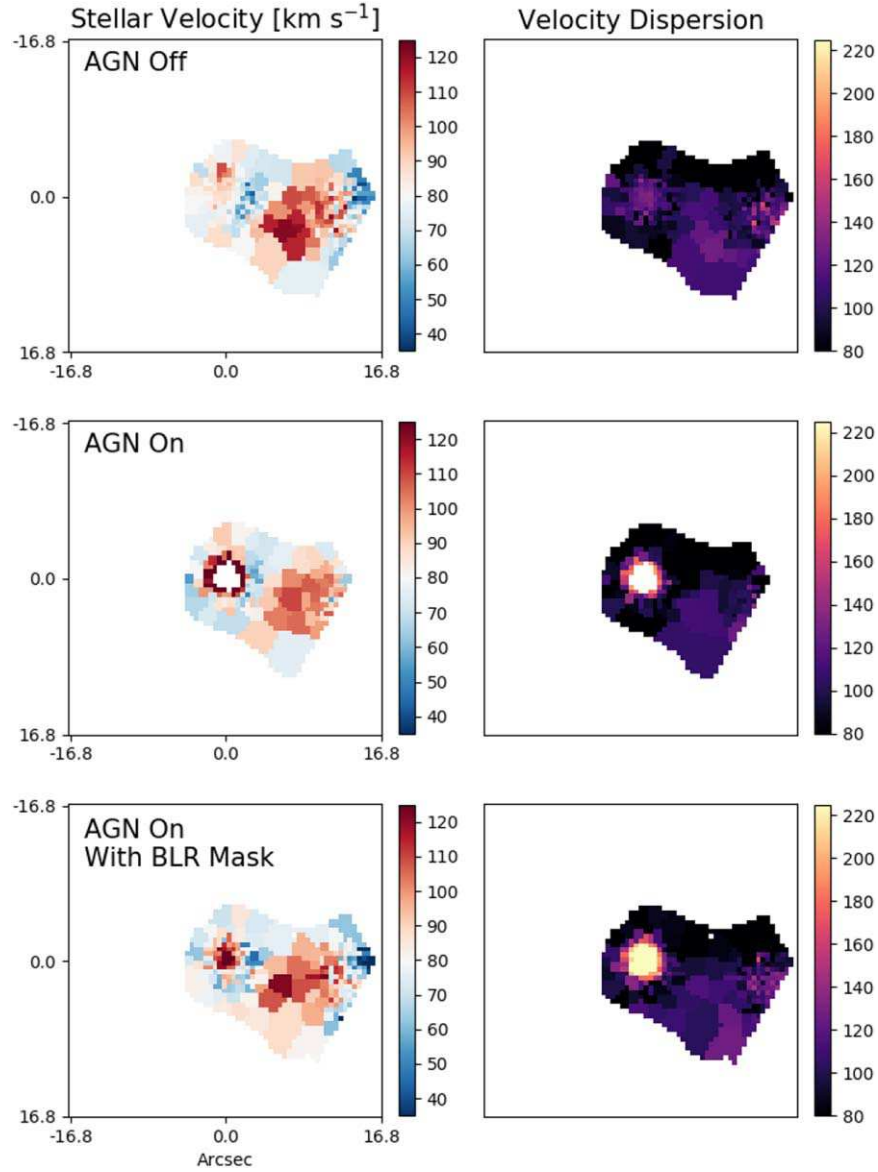


Figure 21. Stellar velocity (left) and stellar velocity dispersion (right) maps for a snapshot of the `q0.5_fg0.3` simulation with an AGN. This specific viewpoint and snapshot has the largest contribution from AGN emission, and therefore showcases the most extreme example of the effects of contamination in the stellar kinematic maps by AGN emission. We present the kinematics from the snapshot with the AGN turned off in the top row, followed by the AGN turned on in the middle row, and the AGN turned on with our modified masking procedure in the bottom row. When the AGN light is not removed, and no emission line mask is used (middle row), `ppxf` is unable to accurately fit the absorption lines in areas contaminated by AGN light, for both velocity and velocity dispersion. These snapshots are not reliable for the extraction of the kinematic predictors. When we mask the emission lines (bottom row), the kinematic predictors are slightly affected. However, this effect is only significant for a few snapshots, so the LDA classification is unaffected by the slight difference between the masked emission lines and the run with the AGN emission turned off.

simulation with the largest contribution from AGN emission. When the AGN continuum is strong, and the emission lines are broad, which is the case for 3–4 (out of 20 total) snapshots in the `q0.5_fg0.3` simulation, the DAP produces velocity and velocity dispersion values of 1000 km s^{-1} , which is the artificial convergence of `ppxf` on the maximum allowed values. These values are then masked.

Since our eventual scientific goals include investigating AGN fueling in mergers, we aim to include galaxies with broad line AGN in the analysis. We therefore explore a variety of options for removing the AGN contamination. One approach to remove AGN contamination is to determine the PSF of the AGN, subtract this from the data cube, and then fit the stellar kinematics. A different approach is to modify `ppxf` to fit

broader emission lines. We find that determining the AGN PSF is nontrivial, and can lead to over-subtraction of continuum light, and as such we pursue the alternate approach, modifying `ppxf` to fit broad emission lines. This approach could also be applied to MaNGA, to fit the stellar kinematics of broad line AGNs.

We modify our `ppxf` fitting procedure to include a step where we use the “ELPEXT” emission line mask to mask all of the emission lines. We experiment with fitting these emission line masks to galaxies with no broad lines, and we find that the returned velocity maps match those maps without an emission line mask, so we use this procedure for all mock data cubes, rather than only those dominated by AGN light. We rerun the `q0.5_fg0.3` simulation with the AGN turned off in order to

verify the success of our approach, and we present the results in Figure 21. We “turn the AGN off” by removing the AGN contribution to the spectrum and re-running the RT (we still include dust absorption and scattering). The GADGET-3 output is unchanged, meaning that we do not remove BH accretion and feedback from the simulated galaxy. The modified version of `ppxf` produces similar velocity and velocity dispersion maps to the simulation with the AGN turned off. In Figure 21, we demonstrate that for a few snapshots where the AGN emission is strongest, the velocity dispersion is elevated at the location of the AGN. A full multiwavelength subtraction of the PSF would be necessary to totally eliminate this effect.

The kinematic predictors measured from the maps where the AGN is off are consistent in most cases with those measured with the modified emission line mask, indicating that the classification does not change significantly, with or without AGN emission. The exception is those kinematic predictors that trace the properties of the velocity dispersion map, such as $\mu_{1,\sigma}$, $\mu_{2,\sigma}$, and $|\mu_{3,\sigma}|$. These predictors all increase, because the velocity dispersion distribution has higher values from the AGN region, resulting in a larger mean, variance, and skew in the velocity dispersion distribution. The λ_{R_e} value also increases in the presence of a bright face-on AGN, because λ_{R_e} is a flux-weighted measurement, far more sensitive to the enhanced velocities in the region of the BL AGN. These predictors are significantly elevated in the presence of an AGN only in the case where the AGN is very bright and face-on. This occurs for ~ 3 snapshots out of 20 total snapshots, and the AGN is only face-on for one of the seven viewpoints, amounting to 3/140 data points. We show one example of dominant AGN emission in Figure 21. These snapshots are already classified as merging when the AGN is off, so while their LD1 value is slightly increased in the presence of an AGN, their classification is unchanged. We can conclude that the AGN alone does not determine the classification, and that it plays a small role for only a few snapshots. In future work, we plan to examine galaxies with significant AGN emission more closely (this will be a small fraction of the MaNGA sample), to determine whether the effects of AGN emission are capable of affecting the classification.

In N19, we did not consider AGN contamination, meaning that the imaging side of the technique is developed with the AGNs present. Since we did not remove the AGNs from the N19 analysis, they may be enhancing the r -band concentration of the gas-rich major mergers. However, the most important predictors in relation to gas-rich major mergers are asymmetry-based, indicating that the classification relies more on the lower surface brightness features of the major mergers, and is not dominated by the central AGN light.

Appendix D Predictors Not Selected by the RFR

Here, we continue the description of kinematic predictors discussed in Section 2.3, i.e., those not selected by the LDA, and therefore not introduced previously. In other words, here we describe predictors that are not as important for identifying mergers: A , v_{asym} , σ_{asym} , Δx_V , and Δx_σ .

In Section 2.3 we define the Radon Transform and Radon profile. Here, we describe how we determine the kinematic center of the galaxy using these as tools, and define the A predictor. We follow the procedure of Stark et al. (2018) in

determining the galaxy’s kinematic center. We use the photometric center as the initial input, but find that the photometric center is not always the same as the kinematic center. Since an incorrect kinematic center can cause variations in the calculation of the Radon Transform, we first determine the kinematic center using the Radon profile, and then extract the kinematic predictors using this kinematic center. The kinematic center is the location where the weighted kinematic asymmetry (A) is minimized, i.e., the asymmetry in the estimated $\hat{\theta}$ values, measured from the bounded Absolute Radon Transform. The A predictor is defined as

$$A = \frac{\sum |\hat{\theta} - \hat{\theta}_{\text{flip}}|}{2N_{i,j}} w_{i,j}, \quad (\text{D1})$$

where $\hat{\theta}$ is reversed to make $\hat{\theta}_{\text{flip}}$, $N_{i,j}$ is the number of values in the $\hat{\theta}$ array at the current “center,” and $w_{i,j}$ is a weight factor

$$w_{i,j} = \frac{N_{0,0}}{N_{i,j}}, \quad (\text{D2})$$

where $N_{0,0}$ is the number of values at the photometric center.

We iteratively measure A in a 3×3 spaxel grid, centered on the photometric center, and select the spaxel with the lowest value as the kinematic center, while taking the minimized A value as our A predictor for each snapshot. If the spaxel with the lowest value of A is consistent, within errors, with the photometric center, we select the photometric center as the kinematic center. If the spaxel with the lowest value is on the edge of the spaxel grid, we expand the grid by a factor of 2, and rerun the determination of the kinematic center.

If the kinematic center is again at the edge of the grid, we do not expand the grid, but take the photometric center as the kinematic center. In this case, the kinematic center is not well-determined, often due to a disorganized velocity map, so the photometric center is a fair estimate of the kinematic center. In Stark et al. (2018), galaxies where the kinematic center is not well-determined are eliminated from the analysis, but in this case, a large fraction of galaxies have disordered kinematics, so we include them in the analysis, using the photometric center as the kinematic center.

We ultimately find that A is highly correlated with A_2 , and is therefore superfluous to the analysis. As a result, it is rejected during the RFR step.

We measure the v_{asym} and σ_{asym} predictors using `kinemetry`, previously introduced in Section 2.3. Our goal with these predictors is to quantify the chaotic velocity patterns expected for interacting systems in the velocity and velocity dispersion maps, using the degree of kinematic asymmetries from the `kinemetry` output. The `kinemetry` output includes the A_n and B_n coefficients from the harmonic expansion of the best-fitting ellipses for each ellipse. From these, we calculate the amplitude and phase coefficients (k_n and ϕ_n)²⁰:

$$k_n = \sqrt{A_n^2 + B_n^2} \quad (\text{D3})$$

²⁰ We do not explicitly apply a flux weighting for the calculation of the phase coefficients, as in Krajnović et al. (2011). However, we do use an uncertainty weighting in the `kinemetry` fitting, which indirectly incorporates a flux weighting. In addition, `kinemetry` uses adaptively sized annuli (which increase in size with radius), which already emphasizes the high flux inner regions of the galaxy.

$$\phi_n = \arctan\left(\frac{A_n}{B_n}\right) \quad (\text{D4})$$

The higher order A_n and B_n terms represent deviations from an ideal rotating disk. We quantify these perturbations with v_{asym} and σ_{asym} , which are measured from the k_n amplitude coefficients of the model velocity and velocity dispersion maps, respectively (Krajnovic et al. 2006; Shapiro et al. 2008; Bellocchi et al. 2016; Hung et al. 2016):

$$v_{\text{asym}} = \left\langle \frac{\sum_{n=2}^5 k_{n,v}/4}{B_{1,v}} \right\rangle_r, \quad \sigma_{\text{asym}} = \left\langle \frac{\sum_{n=1}^5 k_{n,v}/5}{A_{0,v}} \right\rangle_r \quad (\text{D5})$$

where the expression is averaged over all radii, r . The amplitude coefficients, $k_{n,v}$, are summed and averaged for the n higher order moments. We exclude the $k_{1,v}$ term from the calculation of v_{asym} , since it represents radial outflow, which is not associated with stars (Shapiro et al. 2008). We normalize v_{asym} by the circular velocity term, and σ_{asym} by the A_0 term, which is the amplitude of the velocity dispersion maps, as in Krajnovic et al. (2006). The v_{asym} and σ_{asym} predictors increase for disordered velocities and velocity dispersions.

We define Δx_σ as the difference in spatial position between the centroid of the galaxy in r -band imaging, and the centroid of the galaxy’s velocity dispersion map. We determine the centroid (from both the r -band image and the velocity dispersion) by applying a 10×10 pixel low-pass filter, thresholding, and then identifying the center of the brightest contour. We refer to the physical distance (in kpc) between the centroid of the r -band image and the centroid of the velocity dispersion map as Δx_σ . We use a physical distance, as opposed to a spaxel distance, in order to avoid a sensitivity to galaxy redshift. We also measure Δx_v , which is the difference (in kpc) between the centroid of the galaxy in r -band imaging and the kinematic center.

Appendix E

LDA Stability: Is the Training Set Biased?

Here, we interrogate the biases of our simulation suite. We are specifically interested in two questions:

1. Are we justified in directly comparing the classifications from the different simulations of mergers?
2. Is the classification cheating? In other words, are the merging and nonmerging galaxies biased in a way that allows the classification to identify mergers using nuisance parameters?

First, we tackle the validity of comparing different simulations. We find that the LDA differs significantly for the different merger simulations, both in terms of performance, and in the selected predictors. We want to ensure that this is a reflection of the physical properties of the stellar kinematics in these different mergers, and not merely a reflection of irrelevant differences in the data set. For instance, the classification from the q0.2_fg0.3_BT0.2 simulation has a significantly lower accuracy than any other simulation, including the q0.1_fg0.3_BT0.2 simulation, which we would naively expect to have the lowest accuracy. Could this be a result of the q0.2_fg0.3_BT0.2 simulation having fewer data points than any other simulation? We randomly discard data from all simulations so that they are limited to the same amount of data

as the q0.2_fg0.3_BT0.2 simulation. We find that both the performance and the selected predictors for all runs are stable when we decrease the number of simulated snapshots, which indicates that the number of snapshots is not biasing the classification.

Secondly, we carefully examine the properties of the isolated galaxies, relative to the merging galaxies, for each simulation. This is known as the between-class bias of the classification. Our concern is that differences between the merging and nonmerging populations (i.e., trends with inclination or size) may affect the kinematic predictors. For instance, we find that the isolated snapshots tend to be larger in size and slightly more face-on, relative to the nonmerging population, for all simulations. Since it is known that properties such as inclination can significantly affect the observed stellar kinematics, we want to ensure that the classification is not relying upon the underlying properties of the galaxies in order to cheat in its classification of merging galaxies. This could happen if, for instance, the nonmerging galaxies are more face-on, and therefore the kinematic predictors are picking up on this property, as opposed to features relating to the mergers themselves. We have chosen a relatively transparent classification method, so we can use our interpretation of the individual predictors to diagnose whether or not between-class bias is a concern.

The first line of logic we use to prove that the between-class bias is not significant is the interpretation of the important predictors throughout Section 4. The most important predictors track features that can be associated with merging galaxies. In addition, the sensitivity of the method changes with time, meaning that the kinematic features are not long-lived, as one might expect from a classification relying mainly on inclination or size effects, which do not change dramatically over time in a merger.

We also investigate between-class bias by directly introducing suspicious galaxy properties into the analysis. For instance, we expect that if the classification is instead tied to inclination or size, the kinematic predictors would be correlated with these quantities, and therefore the LDA would depend most strongly on inclination or size. To test whether this is the case, we introduce several parameters that quantify size and inclination into the LDA as “nuisance” predictors, so as to determine whether they are important. We use the galaxy ellipticity (ϵ), the number of spaxels in the velocity map, and the effective radius, R_e , of each galaxy as proxies for inclination, apparent size, and physical size, respectively. We expect these nuisance predictors to be important predictors if the merging and nonmerging galaxies are significantly unbalanced in those properties measured by the nuisance predictors. We find that the nuisance predictors are only selected for a few simulations, and when they are selected, they are far less important than the leading terms. Therefore, we conclude that the other predictors are not masquerading as proxies for these properties, and that the isolated galaxies are not significantly biased relative to the merging galaxies.

ORCID iDs

R. Nevin  <https://orcid.org/0000-0003-1056-8401>
 L. Blecha  <https://orcid.org/0000-0002-2183-1087>
 J. Comerford  <https://orcid.org/0000-0001-8627-4907>
 J. E. Greene  <https://orcid.org/0000-0002-5612-3427>
 D. R. Law  <https://orcid.org/0000-0002-9402-186X>

D. V. Stark  <https://orcid.org/0000-0002-3746-2853>
 K. B. Westfall  <https://orcid.org/0000-0003-1809-6920>
 J. A. Vazquez-Mata  <https://orcid.org/0000-0001-8694-1204>
 R. Smethurst  <https://orcid.org/0000-0001-6417-7196>
 J. R. Brownstein  <https://orcid.org/0000-0002-8725-1069>
 N. Drory  <https://orcid.org/0000-0002-7339-3170>

References

- Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, *ApJS*, **240**, 23
- Alonso-Herrero, A., Rieke, G. H., Rieke, M. J., et al. 2006, *ApJ*, **650**, 835
- Barrera-Ballesteros, J. K., Falcón-Barroso, J., García-Lorenzo, B., et al. 2014, *A&A*, **568**, A70
- Barrera-Ballesteros, J. K., García-Lorenzo, B., Falcón-Barroso, J., et al. 2015, *A&A*, **582**, A21
- Barrera-Ballesteros, J. K., Heckman, T., Sánchez, S. F., et al. 2018, *ApJ*, **852**, 74
- Barrera-Ballesteros, J. K., Utomo, D., & Bolatto, A. D. 2020, *MNRAS*, **492**, 2651
- Belfiore, F., Westfall, K. B., Schaefer, A., et al. 2019, *AJ*, **158**, 160
- Bellocchi, E., Arribas, S., & Colina, L. 2012, *A&A*, **542**, A54
- Bellocchi, E., Arribas, S., & Colina, L. 2016, *A&A*, **591**, A85
- Bellocchi, E., Arribas, S., Colina, L., & Miralles-Caballero, D. 2013, *A&A*, **557**, A59
- Bendo, G. J., & Barnes, J. E. 2000, *MNRAS*, **316**, 315
- Blanton, M. R., Bershad, M. A., Abolfathi, B., et al. 2017, *AJ*, **154**, 28
- Blanton, M. R., & Roweis, S. 2007, *AJ*, **133**, 734
- Blecha, L., Civano, F., Elvis, M., & Loeb, A. 2013a, *MNRAS*, **428**, 1341
- Blecha, L., Cox, T. J., Loeb, A., & Hernquist, L. 2011, *MNRAS*, **412**, 2154
- Blecha, L., Loeb, A., & Narayan, R. 2013b, *MNRAS*, **429**, 2594
- Blecha, L., Snyder, G. F., Satyapal, S., & Ellison, S. L. 2018, *MNRAS*, **478**, 3056
- Bloom, J. V., Fogarty, L. M. R., Croom, S. M., et al. 2017, *MNRAS*, **465**, 123
- Bois, M., Emsellem, E., Bournaud, F., et al. 2011, *MNRAS*, **416**, 1654
- Bolatto, A. D., Wong, T., Utomo, D., et al. 2017, *ApJ*, **846**, 159
- Bottrell, C., Hani, M. H., Teimoorinia, H., et al. 2019, *MNRAS*, **490**, 5390
- Bryant, J. J., Bland-Hawthorn, J., Lawrence, J., et al. 2016, *Proc. SPIE*, **9908**, 99081F
- Bundy, K., Bershad, M. A., Law, D. R., et al. 2015, *ApJ*, **798**, 7
- Cappellari, M. 2016, *ARA&A*, **54**, 597
- Cappellari, M. 2017, *MNRAS*, **466**, 798
- Cappellari, M., & Copin, Y. 2003, *MNRAS*, **342**, 345
- Cappellari, M., & Emsellem, E. 2004, *PASP*, **116**, 138
- Cappellari, M., Emsellem, E., Bacon, R., et al. 2007, *MNRAS*, **379**, 418
- Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, *MNRAS*, **413**, 813
- Cherinka, B., Andrews, B. H., Sánchez-Gallego, J., et al. 2019, *AJ*, **158**, 74
- Cole, S., Helly, J., Frenk, C. S., & Parkinson, H. 2008, *MNRAS*, **383**, 546
- Conselice, C. J., Yang, C., & Bluck, A. F. L. 2009, *MNRAS*, **394**, 1956
- Cox, T. J., Dutta, S. N., Di Matteo, T., et al. 2006, *ApJ*, **650**, 791
- Crocker, A. F., Jeong, H., Komugi, S., et al. 2009, *MNRAS*, **393**, 1255
- Croom, S. M., Lawrence, J. S., Bland-Hawthorn, J., et al. 2012, *MNRAS*, **421**, 872
- Croton, D. J., Springel, V., White, S. D. M., et al. 2006, *MNRAS*, **365**, 11
- Dasyra, K. M., Tacconi, L. J., Davies, R. I., et al. 2006, *ApJ*, **651**, 835
- Dekel, A., Birnboim, Y., Engel, G., et al. 2009, *Natur*, **457**, 451
- Di Matteo, T., Springel, V., & Hernquist, L. 2005, *Natur*, **433**, 604
- Drory, N., MacDonald, N., Bershad, M. A., et al. 2015, *AJ*, **149**, 77
- Duda, R. O., Hart, P. E., & Stork, D. G. 2001, *Pattern Classification* (2nd ed.; New York: Wiley)
- Emsellem, E., Cappellari, M., Krajnović, D., et al. 2007, *MNRAS*, **379**, 401
- Emsellem, E., Cappellari, M., Krajnović, D., et al. 2011, *MNRAS*, **414**, 888
- Fabian, A. C. 2012, *ARA&A*, **50**, 455
- Fawcett, T. 2006, *PaReL*, **27**, 861
- Feng, S., Shen, S.-Y., Yuan, F.-T., Riffel, R. A., & Pan, K. 2020, *ApJL*, **892**, L20
- Gabriel, E., Fagg, G. E., Bosilca, G., et al. 2004, in *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, ed. D. Kranzlmüller, P. Kacsuk, & J. Dongarra (Berlin: Springer), 97
- Giallalisco, M., Steidel, C. C., & Macchetto, F. D. 1996, *ApJ*, **470**, 189
- Glazebrook, K. 2013, *PASA*, **30**, e056
- Glazebrook, K., Ellis, R., Santiago, B., & Griffiths, R. 1995, *MNRAS*, **275**, L19
- Goulding, A. D., Greene, J. E., Bezanson, R., et al. 2018, *PASJ*, **70**, S37
- Graham, M. T., Cappellari, M., Li, H., et al. 2018, *MNRAS*, **477**, 4711
- Gunn, J. E., Gott, J., & Richard, I. 1972, *ApJ*, **176**, 1
- Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, *AJ*, **131**, 2332
- Guo, Y., Ferguson, H. C., Bell, E. F., et al. 2015, *ApJ*, **800**, 39
- Haan, S., Surace, J. A., Armus, L., et al. 2011, *AJ*, **141**, 100
- Harborne, K. E., Power, C., Robotham, A. S. G., Cortese, L., & Taranu, D. S. 2019, *MNRAS*, **483**, 249
- Hayward, C. C., Kereš, D., Jonsson, P., et al. 2011, *ApJ*, **743**, 159
- Hayward, C. C., Torrey, P., Springel, V., Hernquist, L., & Vogelsberger, M. 2014, *MNRAS*, **442**, 1992
- Heckman, T. M., & Best, P. N. 2014, *ARA&A*, **52**, 589
- Ho, T. K. 1995, in *ICDAR '95: Proc. Third International Conference on Document Analysis and Recognition* (Vol. 1) (Washington, DC: IEEE Computer Society), 278
- Hopkins, P. F., Cox, T. J., Hernquist, L., et al. 2013a, *MNRAS*, **430**, 1901
- Hopkins, P. F., Cox, T. J., Kereš, D., & Hernquist, L. 2008, *ApJS*, **175**, 390
- Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009, *ApJ*, **691**, 1168
- Hopkins, P. F., Croton, D., Bundy, K., et al. 2010, *ApJ*, **724**, 915
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, *ApJS*, **163**, 1
- Hopkins, P. F., Kereš, D., Murray, N., et al. 2013b, *MNRAS*, **433**, 78
- Hung, C.-L., Hayward, C. C., Smith, H. A., et al. 2016, *ApJ*, **816**, 99
- Hung, C.-L., Rich, J. A., Yuan, T., et al. 2015, *ApJ*, **803**, 62
- Hung, C.-L., Sanders, D. B., Casey, C. M., et al. 2014, *ApJ*, **791**, 63
- Hunter, J. D. 2007, *CSE*, **9**, 90
- James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013, *An Introduction to Statistical Learning with Applications in R* (Berlin: Springer)
- Jesseit, R., Naab, T., Peletier, R. F., & Burkert, A. 2007, *MNRAS*, **376**, 997
- Jonsson, P. 2006, *MNRAS*, **372**, 2
- Jonsson, P., Groves, B. A., & Cox, T. J. 2010, *MNRAS*, **403**, 17
- Kaviraj, S. 2014, *MNRAS*, **437**, L41
- Khim, D. J., Yi, S. K., Dubois, Y., et al. 2020, *ApJ*, **894**, 106
- Krajnović, D., Cappellari, M., Zeeuw, P. T. D., Copin, Y., & Lyon, D. 2006, *MNRAS*, **367**, 787
- Krajnović, D., Emsellem, E., Cappellari, M., et al. 2011, *MNRAS*, **414**, 2923
- Law, D. R., Cherinka, B., Yan, R., et al. 2016, *AJ*, **152**, 83
- Law, D. R., Steidel, C. C., Shapley, A. E., et al. 2012a, *ApJ*, **745**, 85
- Law, D. R., Steidel, C. C., Shapley, A. E., et al. 2012b, *ApJ*, **759**, 29
- Law, D. R., Yan, R., Bershad, M. A., et al. 2015, *AJ*, **150**, 19
- Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, *ApJS*, **123**, 3
- Li, T., Zhu, S., & Ogihara, M. 2006, *Knowl. Inf. Syst.*, **10**, 453
- Lilly, S. J., Tresse, L., Hammer, F., Crampton, D., & Le Fevre, O. 1995, *ApJ*, **455**, 108
- López-Sanjuan, C., Balcells, M., Pérez-González, P. G., et al. 2009, *A&A*, **501**, 505
- López-Sanjuan, C., García-Dabó, C. E., & Balcells, M. 2008, *PASP*, **120**, 571
- Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, *ApJ*, **742**, 103
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, *MNRAS*, **404**, 575
- Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010b, *MNRAS*, **404**, 590
- Lotz, J. M., Patrik, J., Cox, T. J., et al. 2008, *MNRAS*, **391**, 1137
- Mantha, K. B., McIntosh, D. H., Brennan, R., et al. 2018, *MNRAS*, **475**, 1549
- Masters, K. L., Mosleh, M., Romer, A. K., et al. 2010, *MNRAS*, **405**, 783
- McKinney, W. 2010, in *Proc. 9th Python in Science Conf.*, 56, <https://conference.scipy.org/proceedings/scipy2010/mckinney.html>
- Miralles-Caballero, D., Colina, L., Arribas, S., & Duc, P.-A. 2011, *AJ*, **142**, 79
- Naab, T., Oser, L., Emsellem, E., et al. 2014, *MNRAS*, **444**, 3357
- Narayanan, D., Dey, A., Hayward, C. C., et al. 2010, *MNRAS*, **407**, 1701
- Nevin, R., Blecha, L., Comerford, J., & Greene, J. 2019, *ApJ*, **872**, 76
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *J. Mach. Learn. Res.*, **12**, 2825
- Petty, S. M., Armus, L., Charmandaris, V., et al. 2014, *AJ*, **148**, 111
- Piqueras López, J., Colina, L., Arribas, S., Alonso-Herrero, A., & Bedregal, A. G. 2012, *A&A*, **546**, A64
- Robertson, B., Bullock, J. S., Cox, T. J., et al. 2006, *ApJ*, **645**, 986
- Robotham, A. S. G., Driver, S. P., Davies, L. J. M., et al. 2014, *MNRAS*, **444**, 3986
- Rodríguez-Gómez, V., Genel, S., Vogelsberger, M., et al. 2015, *MNRAS*, **449**, 49
- Rodríguez-Gómez, V., Snyder, G. F., Lotz, J. M., et al. 2019, *MNRAS*, **483**, 4140
- Rupke, D. 2018, *Galax*, **6**, 138
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, *A&A*, **538**, A8
- Schawinski, K., Lintott, C., Thomas, D., et al. 2009, *MNRAS*, **396**, 818
- Shapiro, K. L., Genzel, R., Förster Schreiber, N. M., et al. 2008, *ApJ*, **682**, 231
- Sheth, K., Vogel, S. N., Regan, M. W., Thornley, M. D., & Teuben, P. J. 2005, *ApJ*, **632**, 217
- Shi, Y., Rieke, G., Lotz, J., & Pérez-González, P. G. 2009, *ApJ*, **697**, 1764

- Silk, J., & Rees, M. J. 1998, *A&A*, **331**, L1
- Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, *AJ*, **146**, 32
- Snyder, G. F., Hayward, C. C., Sajina, A., et al. 2013, *ApJ*, **768**, 168
- Snyder, G. F., Rodriguez-Gomez, V., Lotz, J. M., et al. 2019, *MNRAS*, **486**, 3702
- Springel, V. 2005, *MNRAS*, **364**, 1105
- Springel, V., & Hernquist, L. 2003, *MNRAS*, **339**, 289
- Stark, D. V., Bundy, K. A., Westfall, K., et al. 2018, *MNRAS*, **480**, 2217
- Stickley, N. R., & Canalizo, G. 2016, arXiv:1610.02601
- Sweet, S. M., Glazebrook, K., Obreschkow, D., et al. 2020, *MNRAS*, **494**, 5421
- Tacconi, L. J., Genzel, R., Smail, I., et al. 2008, *ApJ*, **680**, 246
- The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- The Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22
- Veilleux, S., Kim, D.-C., & Sanders, D. B. 2002, *ApJS*, **143**, 315
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wake, D. A., Bundy, K., Diamond-Stanic, A. M., et al. 2017, *AJ*, **154**, 86
- Wang, B., Cappellari, M., Peng, Y., & Graham, M. 2020, *MNRAS*, **495**, 1958
- Waskom, M., Botvinnik, O., Hobson, P., et al. 2014, Seaborn: v0.5.0 (November 2014), Zenodo, doi:10.5281/zenodo.12710
- Weigel, A. K., Schawinski, K., Caplar, N., et al. 2017, *ApJ*, **845**, 134
- Wen, Z. Z., & Zheng, X. Z. 2016, *ApJ*, **832**, 90
- Westfall, K. B., Cappellari, M., Bershad, M. A., et al. 2019, *AJ*, **158**, 231
- White, H. A., Fisher, D. B., Murray, N., et al. 2017, *ApJ*, **846**, 35
- White, S. D. M., & Frenk, C. S. 1991, *ApJ*, **379**, 52
- White, S. D. M., & Rees, M. J. 1978, *MNRAS*, **183**, 341
- Yan, R., Bundy, K., Law, D. R., et al. 2016a, *AJ*, **152**, 197
- Yan, R., Tremonti, C., Bershad, M. A., et al. 2016b, *AJ*, **151**, 8