## THE BLACK HOLE MASS OF NGC 4151 FROM STELLAR DYNAMICAL MODELING

CAROLINE A. ROBERTS<sup>1,2</sup>, MISTY C. BENTZ<sup>1</sup>, EUGENE VASILIEV<sup>3,4,5</sup>, MONICA VALLURI<sup>6</sup>, AND CHRISTOPHER A. ONKEN<sup>7</sup> (Received; Accepted) Draft version June 8, 2021

#### **ABSTRACT**

The mass of a supermassive black hole  $(M_{\rm BH})$  is a fundamental property that can be obtained through observational methods. Constraining  $M_{\rm BH}$  through multiple methods for an individual galaxy is important for verifying the accuracy of different techniques, and for investigating the assumptions inherent in each method. However, there exist only a few galaxies where multiple  $M_{\rm BH}$  measurement techniques can be applied. NGC 4151 is one of these rare galaxies for which multiple methods can be used: stellar and gas dynamical modeling because of its proximity ( $D = 15.8 \pm 0.4$  Mpc from Cepheids), and reverberation mapping because of its active accretion. In this work, we re-analyzed H-band integral field spectroscopy of the nucleus of NGC 4151 from Gemini NIFS, improving the analysis at several key steps. We then constructed a wide range of axisymmetric dynamical models with the new orbit-superposition code FORSTAND. One of our primary goals is to quantify the systematic uncertainties in M<sub>BH</sub> arising from different combinations of the deprojected density profile, inclination, intrinsic flattening, and mass-to-light ratio. As a consequence of uncertainties on the stellar luminosity profile arising from the presence of the AGN, our constraints on  $M_{\rm BH}$  are rather weak. Models with a steep central cusp are consistent with no black hole; however, in models with more moderate cusps, the black hole mass lies within the range of  $0.25 \times 10^7 M_{\odot} \lesssim M_{\rm BH} \lesssim 3 \times 10^7 M_{\odot}$ . This measurement is somewhat smaller than the earlier analysis presented by Onken et al. (2014), but agrees with previous  $M_{\rm BH}$  values from gas dynamical modeling and reverberation mapping. Future dynamical modeling of reverberation data, as well as IFU observations with JWST, will aid in further constraining the mass of the central supermassive black hole in NGC 4151.

Subject headings: galaxies: active — galaxies: Seyfert — galaxies: supermassive black holes

#### 1. INTRODUCTION

Two of the most convincing pieces of evidence for the existence of supermassive black holes are also based on two unique methods for measuring black hole mass  $(M_{\rm BH})$ . The most robust black hole mass measurement is that based on proper motion studies of stars in the Galactic Center around Sagittarius A\*, which require a supermassive black hole with  $M_{\rm BH} = 4 \times 10^6 \, \rm M_{\odot}$  (Ghez et al. 2000; Genzel et al. 2000; Schödel et al. 2002; Gravity Collaboration et al. 2019). Unfortunately, observing the proper motions of individual stars in the centers of other galaxies is not possible due to the distances involved. The second robust black hole mass measurement is in the case of M87 which is based on image reconstruction using very-long baseline interferometry that provided the first image of emission from just outside the event horizon. These observations allowed the black hole mass to be constrained through comparison with general relativistic magnetohydrodynamic models (Event Horizon Telescope Collaboration et al. 2019). Unfortunately, imaging of the event horizon is only currently possible for M87 and for Sagittarius A\*, which has a comparable angular extent on the sky. All other

black holes in the nearby universe are too small and/or too distant to be resolved with current technology, and so other techniques must be used to study additional supermassive black holes.

Another direct method for constraining  $M_{\rm BH}$  is the use of water maser emission in edge-on nuclear gas disks (Miyoshi et al. 1995; Herrnstein et al. 2005; Greene et al. 2016). The maser emission in the disk traces out the Keplerian rotation curve of the gas, constraining the enclosed mass. Unfortunately, water maser emission is quite rare and also requires a specific set of circumstances to be fulfilled before it can be used for  $M_{\rm BH}$  measurements, so there are only a few galaxies where this technique may be used.

For larger samples of black hole masses, there are three other established direct  $M_{\rm BH}$  measurement techniques. Gas dynamical (GD) modeling relies on spatially-resolved gas kinematics in galactic nuclei to infer the geometry and inclination of the gas as well as the enclosed mass (e.g., Macchetto et al. 1997; den Brok et al. 2015), though the gas may be affected by non-circular motions and turbulence. Stellar dynamical (SD) modeling is similar, but considers a more general kinematic structure of the stellar motion, not limited to circular orbits (e.g., van der Marel et al. 1998; Cretton & van den Bosch 1999; Gebhardt et al. 2003a; Valluri et al. 2004). Reverberation mapping (RM, Blandford & McKee 1982; Peterson 1993) uses light echoes in the emission from active galactic nuclei (AGN) to constrain the kinematics of gas on spatially-unresolvable scales deep in the nuclear region. Both GD and SD modeling require high spatial resolution and are dependent on the distance to the galaxy. RM instead requires high temporal sampling and is distance independent.

For active galaxies, RM is the most prevalent  $M_{\rm BH}$  measurement technique. Continuum emission, likely arising from the

<sup>&</sup>lt;sup>1</sup> Department of Physics and Astronomy, Georgia State University, Atlanta, GA 30303, USA; roberts@astro.gsu.edu

<sup>&</sup>lt;sup>2</sup> Department of Physics and Astronomy, University of Iowa, Iowa City, IA, 52242, USA

<sup>3</sup> Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, UK

<sup>&</sup>lt;sup>4</sup> Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford,

<sup>&</sup>lt;sup>5</sup> Lebedev Physical Institute, Leninsky Prospekt 53, Moscow, 119991,

<sup>&</sup>lt;sup>6</sup> Department of Astronomy, University of Michigan, Ann Arbor, MI,

Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia

accretion disk, travels outwards at the speed of light and photoionizes gas in the broad line region (BLR), where it is processed and re-emitted as spectral lines. Variability in the continuum flux (most likely from instabilities in the disk and/or variable accretion rates) thus causes variability in the broad line flux as well. Through spectrophotometric monitoring, the time delay  $\tau$  between variations in the continuum radiation and the response in the broad emission lines can be measured. The recombination timescale of the BLR is very short compared to typical time delays, so  $\tau$  is simply the light-crossing time and  $c\tau$  is the responsivity-weighted mean radius of the BLR. When combined with a constraint on the Doppler velocities of the BLR gas (V),  $M_{\rm BH}$  can be determined via the virial theorem:

$$M_{\rm BH} = f \frac{c\tau V^2}{G},\tag{1}$$

where f is an order-unity scaling factor that depends on the detailed geometry and kinematics of the BLR gas and G is the gravitational constant. RM is a direct method of determining  $M_{\rm BH}$  based on the gravitational influence of the black hole on a luminous tracer, but its most common application relies on an average f factor that is derived from comparison with dynamical modeling (e.g., Onken et al. 2004; Grier et al. 2013; Batiste et al. 2017).

RM masses determined in this way rely on certain assumptions, such as a symmetric geometry of the BLR and the assumption that gravity dominates the kinematics of the BLR gas. The BLR exhibits ionization stratification, in that high ionization lines are observed to have shorter time delays than low ionization lines. These short time delays are accompanied by high velocities in the line widths. For those cases when it has been possible to explore the relationship between  $\tau$  and Vfor multiple broad emission lines in the same AGN, the measurements have been consistent with  $V \propto \tau^{-0.5}$ , as expected for a virial relationship (e.g., Peterson et al. 2004; Kollatschny 2003; Bentz et al. 2010) and thus supporting the assumption that gravity dominates the kinematics of the region. Velocityresolved RM (Pancoast et al. 2014; Grier et al. 2017a), on the other hand, allows the full geometry and kinematics of the BLR to be constrained and avoids many of the assumptions involved in using a mean time delay and adopting a typical f factor, thus providing a direct, primary constraint on  $M_{\rm BH}$ .

SD modeling is generally applied to quiescent galaxies – elliptical or spheroidal galaxies and the bulges of disk galaxies. It is a direct, primary method that constrains  $M_{\rm BH}$  by fitting the bulk kinematics of stars derived from spatiallyresolved spectroscopy with simulated kinematics constructed for a galaxy model with a similar surface brightness density profile to the observed galaxy. Several different approaches have been employed for stellar dynamical modeling, e.g., the solution of the spherical or axisymmetric Jeans equation (van der Marel et al. 1994; Cappellari 2008, 2014, 2020), distribution function fitting (van der Marel et al. 1994; Magorrian 2019), guided ("made-to-measure") Nbody simulations (Syer & Tremaine 1996; De Lorenzi et al. 2013), and the Schwarzschild orbit-superposition method (Schwarzschild 1979, 1993). In the past two decades, in conjunction with spatially-resolved absorption line spectroscopy, increasingly sophisticated versions of the orbit-superposition method (van der Marel et al. 1998; Cretton & van den Bosch 1999; Gebhardt et al. 2003b; Valluri et al. 2004; Thomas et al. 2004; van den Bosch et al. 2008; van den Bosch & de Zeeuw 2010; Vasiliev & Valluri 2020; Neureiter et al. 2021) have been used to construct self-consistent dynamical models of galactic nuclei and to derive their black hole masses, stellar mass-to-light ratios  $\Upsilon$ , and internal orbit distributions. Consequently, orbit superposition is now the most widely used method for black hole mass determination and is responsible for obtaining the majority of BH mass measurements from dynamical modeling to date (e.g. McConnell & Ma 2013; Saglia et al. 2016). The accuracy of any SD method depends on the ability to spatially resolve the SMBH sphere of influence (the region where the gravity of the SMBH dominates over the gravity of the stars), although some authors argue that resolving the sphere of influence is not strictly necessary (e.g., see Davies et al. 2006; Gültekin et al. 2011).

Like SD modeling, RM is also a prolific measurement technique, and has been used to determine masses of  $\sim$ 100 supermassive black holes (e.g., Bentz & Katz 2015; Grier et al. 2017b). However, for the vast majority of targets it is not possible to constrain  $M_{\rm BH}$  through both RM and SD modeling: bright broad-lined AGNs are rare in the local universe, so they are generally too far away to achieve the spatial resolution needed for dynamical modeling. This is an important point, because tens of thousands of indirect measurements have been made by large surveys and are used to explore the growth and evolution of black holes and galaxies throughout cosmic history based on the information gleaned from these smaller samples of direct  $M_{\rm BH}$  measurements, yet we do not currently know if RM and SD modeling provide consistent mass measurements when applied to the same galaxies.

Direct comparisons of RM and SD modeling through  $M_{\rm BH}$  measurements of the same targets, with their different assumptions, biases, and data and technical requirements, have only been accomplished for two galaxies to date: NGC 4151 (Bentz et al. 2006; Onken et al. 2014; De Rosa et al. 2018) and NGC 3227 (Davies et al. 2006; Denney et al. 2010). We have thus undertaken an effort to improve and increase the sample of high-quality  $M_{\rm BH}$  determinations for the small sample of nearby, bright Seyfert 1 galaxies where RM and SD modeling may be directly compared.

An important step in this program is a complete reanalysis of the SD modeling results for NGC 4151, which we describe in this manuscript. NGC 4151 is the brightest Seyfert 1 in the northern hemisphere and one of the nearest AGNs at z = 0.0033. NGC 4151 has been studied with both SD modeling (Onken et al. 2007, 2014) and RM (Bentz et al. 2006; De Rosa et al. 2018), as well as GD modeling (using H<sub>2</sub>; Hicks & Malkan 2008). The sphere of influence of NGC 4151 is estimated to be  $\sim$ 15 parsecs, or  $\sim$  0."2, which is well-matched to the spatial resolution that may be achieved from the ground with adaptive optics.

The original SD modeling analysis by Onken et al. (2014) did not take full advantage of the spatial resolution available with the instrumentation, and we have identified several additional improvements in the data reduction and analysis that we have implemented and describe below. On the modeling side, we use a powerful new orbit-superposition code FORSTAND (Vasiliev & Valluri 2020) that overcomes many limitations of other stellar-dynamical modeling codes (e.g. restriction to axisymmetry, not accounting for dark matter and inability to account for figure rotation, limited size of orbit libraries, etc.), although in the present study we do not use all its novel features.

In Section 2 of this paper, we introduce relevant integral field spectroscopy and imaging observations. Section 3 details the analysis of the photometry while Section 4 details

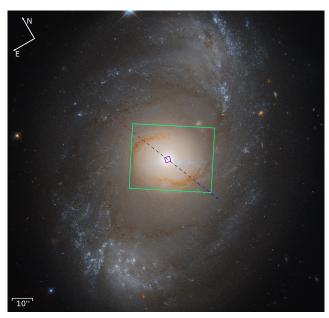


FIG. 1.— HST Wide-Field Camera 3/UVIS optical image of NGC 4151 through the F350LP, F814W, and F555W filters. North is 30° counter-clockwise from up. Small magenta and large green rectangles show the footprints of the NIFS and SAURON kinematic maps, respectively; dashed red-blue line shows the orientation of the kinematic major axis. Image credit: Judy Schmidt.

the analysis of the two spectroscopic data sets. In Section 5 we describe the application of the FORSTAND modeling code to NGC 4151. Section 6 compares the modeling results and conclusions for  $M_{\rm BH}$  with previous studies of NGC 4151, discusses the obstacles of stellar dynamical modeling in this application, and shares the upcoming direction of this project and related research. We summarize in Section 7.

## 2. OBSERVATIONS

NGC 4151 has a de Vaucouleurs et al. (1964) classification<sup>8</sup> of (R?)SAB(rs)ab, having a possible outer ring, a weak bar, an inner ring and inner spiral structure, and tightly wound spiral arms (Figure 1). The large round component that has often been identified as the bulge of the galaxy is actually a barlens, and the true bulge is significantly more compact (cf. Bentz & Manne-Nicholas 2018). It is located at  $\alpha = 182.6357$ ,  $\delta = +39.4058$  in the direction of the constellation Canes Venatici.

# 2.1. Integral Field Spectroscopy 2.1.1. Gemini NIFS

NGC 4151 was observed with the Near-infrared Integral Field Spectrograph (NIFS; McGregor et al. 2003) on the Gemini North telescope on 2008 February 16-17 and 19-24 (see Table 1). We retrieved the raw data from the Gemini Archive (GN-2008A-Q-41, PI: C. Onken).

NIFS is an image-slicer style integral field unit (IFU) with 29 slices across the  $3.''0 \times 3.''0$  field of view (FOV). The observations of NGC 4151 were taken in the H-band with the H\_G5604 grating coupled with the JH\_G0602 filter, covering the spectral range  $\sim 14900-18000$  Å with R $\approx$ 5290 and a dispersion of 1.6 Å pix<sup>-1</sup>. The instrumental resolution was measured to have FWHM = 3.2 Å pix<sup>-1</sup>. The data were acquired with the instrument rotated to a position angle (PA) of  $-15^{\circ}$ 

TABLE 1
GEMINI NIFS OBSERVATIONS

Date	Target	Type	Exp. Time	# Obs.
(Feb. 2008)	C	<i>C</i> 71		
	2700 1151		120	
16	NGC 4151	Science	120	54
	HD 98152	Telluric (A0)	40	12
	HD 116405	Telluric (A0)	30	4
17	NGC 4151	Science	120	36
	HD 98152	Telluric (A0)	40	8
	HD 116405	Telluric (A0)	30	8
19	NGC 4151	Science	120	47
	HD 98152	Telluric (A0)	40	4
	HD 116405	Telluric (A0)	30	12
20	NGC 4151	Science	120	19
	HD 98152	Telluric (A0)	40	8
	HD 116405	Telluric (A0)	30	4
21	NGC 4151	Science	120	24
	HD 98152	Telluric (A0)	40	8
	HD 116405	Telluric (A0)	30	4
22	NGC 4151	Science	120	18
	HD 98152	Telluric (A0)	40	4
	HD 116405	Telluric (A0)	30	4
23	NGC 4151	Science	120	36
	HD 98152	Telluric (A0)	40	4
	HD 116405	Telluric (A0)	30	8
	HIP 60145	Template (M0)	5.3	4
24	NGC 4151	Science	120	18
	HD 98152	Telluric (A0)	40	4
	HD 116405	Telluric (A0)	30	4
	HD 35833	Template (G0)	5.3	4
	HD 40280	Template (K0III)	5.3	4
		1		

east of north. The Altair AO system (Herriot et al. 1998) was used with the bright AGN serving as a natural guide "star".

Each exposure of NGC 4151 had a length of 120 s, and 252 individual exposures were obtained during the 8 nights of observations. The data quality and weather conditions were good; only 3% of the data did not meet the expectations in quality checks such as cloud cover, water vapor/transparency, and background counts indicated by the PI prior to the observations. At all times the data were marked as usable by the Gemini staff's quality assessment. The airmass was rarely above 1.5 (only 5% of the time), and the average airmass was

Observations of Galactic stars were also collected to produce telluric spectra that quantify the absorption of light from molecules in Earth's atmosphere (most notably water vapor). Two A0V stars were used for these telluric observations, HD 98152 and HD 116405, with exposure times of 40 s and 30 s, respectively. They were observed throughout each of the eight nights, interspersed between the science observations in order to monitor the varying telluric feature strengths due to changing airmass and weather conditions.

G, K, and M stars dominate the near-IR stellar emission of galaxies, and the high luminosities of giant stars, in particular, are responsible for the bulk of the stellar emission at these wavelengths. Three giant stars — HD 35833 (G0), HD 40280 (K0III), and HIP 60145 (M0) — were also observed on Feb. 23 and 24 to serve as velocity templates for assistance in the interpretation of the NGC 4151 spectra. For each of the three stars, four 5.3 s exposures were obtained in a single night.

Observations of NGC 4151 and the telluric and velocity template stars were typically obtained in an object-sky-object dithering pattern to allow for sky subtraction. The telescope was slewed  $\sim\!200''$  to the side of the FOV for the NGC 4151 data ( $\sim\!5''$  for telluric and velocity template stars) before recentering on the target for the next exposure. The AO was turned on for the science data and the telluric stars, but not for

<sup>8</sup> NASA/IPAC Extragalactic Database; https://ned.ipac.caltech.edu/

the velocity templates.

Reduction of the data generally followed the NIFS reduction pipeline created by Tracy Beck and Richard McDermid for IRAF<sup>9</sup>. Many of the basic reduction procedures are the same as those employed in the original study of these data by Onken et al. (2014), but we highlight the improvements we have made to the reductions below.

In particular, telluric spectra were created from the telluric star observations using the software xtellcor (Vacca et al. 2003). The telluric star spectra consist of stellar continuum and absorption with telluric features superimposed on top. xtellcor uses a high-resolution, synthetic spectrum of Vega to model the spectra of A0V stars and isolate the stellar features from the telluric features. Each telluric star spectrum was fed to xtellcor along with the B and V magnitudes of the star (8.98 and 8.93 for HD 98152, 8.27 and 8.34 for HD 116405). Our telluric standard stars, while being the same spectral type as Vega, had different absorption line widths in their spectra than those of Vega, and so xtellcor scaled and blurred the model of Vega to better match the intrinsic stellar absorption features. The best-fit model was then subtracted from the observed spectrum of the telluric standard, leaving only the telluric absorption spectrum. This process differs slightly from that of Onken et al. (2014), who instead employed the nffixa0 IRAF script written by Peter McGregor, in which the stellar absorption lines of the telluric standards were individually fit with Voigt profiles and removed in order to isolate the features arising from Earth's atmosphere. In both cases, the execution of multiple telluric template observing blocks each night allowed individual frames of NGC 4151 to be corrected using the telluric template spectrum obtained closest in time.

For observations of NGC 4151, we ensured the wavelength axis was consistent among all the cubes by adopting a fixed dispersion of  $1.6\,\text{Å}\,\text{pix}^{-1}$  and identical starting wavelengths for each cube as they were rectified. We then assessed the FWHM of the point spread function (PSF) for each cube and implemented a seeing cut at FWHM = 0."25, thus rejecting 21 of the 252 individual observations with the poorest spatial resolution (most of which had been observed in the last two nights of observations). This left 231 individual cubes that were aligned and median combined into a final data cube.

Variance information for each data frame was recorded at the time of observation, but the current version of the NIFS reduction pipeline does not carry this information through to the end of the process. Thus, we developed a method for quantifying the adjustments to the variance at each remaining step in the pipeline and producing a final noise cube to match the final data cube. The inclusion of propagated errors is an improvement on the reduction methods of Onken et al. (2014), where a constant fractional error of 2% was assumed.

#### 2.1.2. WHT SAURON

An additional data cube with wider-field IFU observations of NGC 4151 was provided by Eric Emsellem. The observations were collected as part of the study by Dumas et al. (2007), where integral field spectroscopy from the SAURON IFU was used to examine the kinematics of the stellar and gaseous components of a sample of galaxies including NGC 4151. Three 30 min exposures were collected on the 4.2-m William Herschel Telescope (WHT) in La Palma, Spain in 2004 March. SAURON is a lenslet system rather than an image-slicer, and the IFU has a field-of-view of  $33'' \times 41''$  with a spatial resolution of 0.''94, resulting in  $\sim 1500$  spectra per pointing. The spectral range is in the optical, covering 4825-5380 Å, and the instrument has a spectral resolution of 4.2 Å pix $^{-1}$ .

The SAURON data were reduced with the XSAURON software and SAURON pipeline (Bacon et al. 2001; de Zeeuw et al. 2002). In this process, a bias and dark subtraction was performed, followed by wavelength calibration, flat-fielding, cosmic ray correction, sky subtraction, and flux calibration. The three individual cubes were then aligned and drizzled onto a square grid with a spatial sampling of  $0.18 \times 0.18$ . The PSF of the final data cube has FWHM =  $2.10 \pm 0.18$ . For further information on the data, we refer the reader to Dumas et al. (2007).

#### 2.2. Photometry

High-spatial resolution *Hubble Space Telescope* (*HST*) imaging of NGC 4151 had been previously acquired with the Wide Field Camera 3 (WFC3) through the F547M (medium-*V*; GO-11661, PI: M. Bentz) and F160W (*H*; GO-13765, PI: B. Peterson) filters.

Details of the reduction and processing of the F547M images are described in Bentz & Manne-Nicholas (2018), but in brief, four images through the F547M filter were acquired in a single orbit on 2010 July 3 with a two-point dither pattern to cover the central gap between the detectors. At each point in the dither pattern, a short exposure and a long exposure were obtained, allowing for correction of saturation in the nucleus due to the AGN while also providing good sensitivity to the extended host galaxy. The total exposure time was 2310.0 s, and the drizzled image covers a field of view of  $2.7 \times 2.7$  at a pixel scale of 0.000.

Six individual F160W images were acquired in pairs with equal exposure times and a two-point dither pattern over the course of three orbits on 2015 December 7, 2015 December 28, and 2016 January 5. The total exposure time was 3317.6 s, and the drizzled image covers a field of view of  $2.3 \times 2.1$  at a pixel scale of 0.1283.

The drizzled F547M and F160W images were each fit separately with GALFIT (Peng et al. 2002, 2010) to construct two-dimensional surface brightness profiles (see Bentz et al. 2013; Bentz & Manne-Nicholas 2018 for more details). The surface brightness models allowed the PSF of the central AGN and the background sky to be removed from each image, leaving just the host galaxy for further analysis.

To compare the surface brightness profiles of the galaxy in the two filters, we rotated the F160W image to match the orientation of the F547M image, blurred and rebinned the

<sup>&</sup>lt;sup>9</sup> IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

<sup>&</sup>lt;sup>10</sup> Magnitudes of the telluric stars were retrieved from SIMBAD.

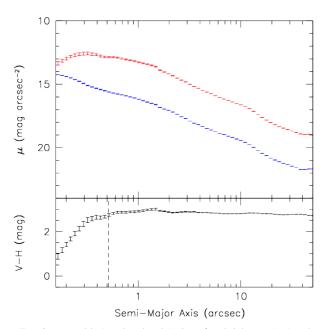


FIG. 2.— V- (blue) and H-band (red) surface brightness (top) and V-H color (bottom) as a function of radius from the center of NGC 4151. The color is relatively constant, except for the innermost 0."5 ( $\sim$  4 pixels) which are affected by the bright central AGN.

F547M image to match the PSF profile and spatial scale of the F160W image, and then cropped the two images so they were both centered on the AGN and included a common field of view. Comparing the positions of the few resolved sources in common between the two images, we found that this process aligned the two images to an accuracy of  $\sim 0.5 \,\mathrm{pix}$  at the final spatial scale. We then used the IRAF task EL-LIPSE to fit elliptical isophotes to both images and measure their one-dimensional surface brightness profiles in a consistent manner. Finally, we applied the most recent calibration of the Vegamag zeropoints, corrected for Galactic extinction based on the Schlafly & Finkbeiner (2011) recalibration of the Schlegel et al. (1998) dust map of the Milky Way, and applied small color corrections to account for the differences between a V-band filter and F547M, and an H-band filter and F160W. Figure 2 (top) displays the one-dimensional surface brightness profiles in the inner regions of NGC 4151, with V shown in blue and H shown in red. The V-H color is displayed in the bottom panel. The pixels inside a radius of  $\sim 0''\bar{5}$  (equivalent to a radius of about 4 pixels, and denoted by the vertical line in the bottom panel) are affected by residuals from the subtraction of the AGN PSF, but at radii outside this region the V-H color of the galaxy is quite flat across the inner  $\sim 1.0$ .

## 3. PHOTOMETRIC ANALYSIS

## 3.1. Image Decompositions

The surface brightness of NGC 4151 in both the WFC3 F547M and F160W images were examined for their appropriateness in constraining the stellar luminosity density within the dynamical models. The F547M image shows clear regions of dust extinction and excess stellar flux, while the F160W image shows a smoother galaxy profile but at the expense of a lower intrinsic spatial resolution than the F547M image.

We first tried a Multi-Gaussian Expansion (MGE) analysis (Emsellem et al. 1994; Cappellari 2002) using the AGN- and sky-subtracted images of the galaxy discussed above. MGE surface-brightness decomposition begins by finding the cen-

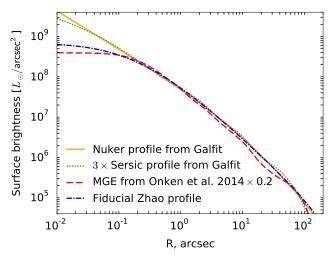


FIG. 3.— Comparison of several parametrizations of the surface brightness profiles. Orange solid line and green dotted lines show the GALFIT models to the F547M image with a single Nuker component or 3 Sérsic components, respectively. Red dashed line is the *H*-band MGE model from Onken et al. (2014) scaled by 0.2 to match the other profiles, which are based on the *V*-band HST image. Blue dot-dashed line is our fiducial Zhao (1996) profile, which closely follows the Nuker profile outside 0."1 but is less cuspy.

ter then binning the field in angle and radius. Photometry is measured for each point on the grid, and then nested two-dimensional Gaussian profiles with variable axis ratios and central intensities are fit to this two-dimensional intensity map. During this process, we required that the PAs of any elongation in the Gaussians aligned with the kinematic PA determined previously (this restriction is mandated by our choice of axisymmetric modeling strategy). Although the observed velocity field shows twisting and misalignment with the photometric PA, indicative of the presence of a bar, we defer the study of bar models to the future.

The parameters of the best-fit nested Gaussian profiles are described by the intensity in counts, the width in pixels  $(\sigma)$ , and axis ratio (q). These parameters can then be converted into surface density in  $L_{\odot}$  pc<sup>-2</sup> and  $\sigma$  in arcsec. For both the F547M and the F160W images, the MGE solutions provided a surface brightness profile that included a lot of bumps and other structures, rather than a smooth increase in surface brightness from larger radii into the galaxy center.

We then examined the potential of using the surface brightness components that had been fit to the images with GALFIT in the process of subtracting the AGN and sky. These surface brightness components included an exponential disk, Sérsic bulge, and an elongated Sérsic component to represent the weak bar, with the central positions of all the galaxy components tied together, and the position angles of the central galaxy components (not including the disk) set to the kinematic position angle of the inner galaxy. While smoother than the MGE profile, a significant amount of structure was still visible in the surface brightness profile for both the F547M and the F160W image.

Finally, we also examined the results of fitting a Nuker profile to the images using GALFIT, to promote smoothness in the surface brightness profile while still retaining some flexibility to account for the complicated morphology. Figure 3 demonstrates that the Nuker fit to the F547M image is very close to the combination of several Sérsic profiles, but is smoother. However, both these profiles overpredict the amount of stellar light in the central 0."1 – the region so dominated by the AGN emission that any residual stellar light is poorly con-

strained. A comparison with the MGE model from Onken et al. (2014), suitably scaled in normalization, shows that the latter has a central core with the radius of the innermost Gaussian component  $\sim 0.11$ , while the Nuker model has a cusp. As will be shown later in Section 5, a cuspy profile results in a best-fit black hole mass of zero, which is clearly unphysical for this bright AGN. We therefore constructed another model described by the Zhao (1996) profile, which closely follows the Nuker model except the innermost 0."1 – the region dominated by the AGN emission. We note that both Nuker and Zhao profiles are equally flexible and can represent cuspy or cored systems for different choices of parameters; here we used the former parametrization for the cuspy profile and the latter for the cored one. These two extreme cases likely encompass the true range of possible density profiles for this galaxy. The 3d luminosity density for our adopted Zhao pro-

$$j(r) = j_0 (r/r_0)^{-\gamma} \left[ 1 + (r/r_0)^{\alpha} \right]^{(\gamma - \beta)/\alpha},$$
 (2)

with  $\gamma = 0$ ,  $\beta = 2.47$ ,  $\alpha = 0.65$ ,  $r_0 = 0.113$ , and  $j_0 = 4.08 \times 10^6 L_{\odot} \, \mathrm{kpc}^{-3}$ . We note that although  $\gamma = 0$  formally corresponds to a cored profile, the zero logarithmic slope is achieved only asymptotically, and in practice the density remains weakly cuspy at the resolution limit, unlike the MGE parametrization, which is much more obviously cored within 0.111.

# 3.2. Mass-to-Light Ratio

Determination of how mass traces light is integral to understanding the gravitational potential of the galaxy for modeling. Based on OSUBSGS H-band imaging combined with SDSS g- and i-band imaging, Onken et al. (2014) produced color-color maps of NGC 4151 and found very flat colors across the galaxy bulge with average values of  $g-i=1.1\pm0.1$  mag and  $i-H=2.4\pm0.2$  mag. This is in line with our findings based on HST F547M and F160W imaging as described in Section 2.2, where we also find a very flat color across the galaxy bulge, with  $V-H=2.9\pm0.1$  mag.

Onken et al. (2014) adopted the models of Zibetti et al. (2009) to predict  $\Upsilon_H \simeq 0.4 \pm 0.2~M_{\odot}/L_{\odot}$ . Comparison of many different prescriptions for estimating  $\Upsilon$  by Roediger & Courteau (2015), however, finds that Zibetti et al. (2009) consistently predicts the lowest values at these galaxy colors. Furthermore, when compared to the known  $\Upsilon$  for resolved stars in M31 (Telford et al. 2020), the prescriptions of Zibetti et al. (2009) seem to be biased low at these colors.

Based on the analysis by Roediger & Courteau (2015), we therefore adopt  $\Upsilon_H \simeq 0.7 \pm 0.1 \, \mathrm{M}_\odot/\mathrm{L}_\odot$  and  $\Upsilon_V \simeq 2.9 \pm 0.3 \, \mathrm{M}_\odot/\mathrm{L}_\odot$  as our best estimates for the H- and V-band stellar mass-to-light ratios in the nucleus of NGC 4151.

## 3.3. Galaxy Inclination Angle

Observations of HI 21 cm emission from the large-scale disk of NGC 4151 led Davies (1973) to report an inclination of  $26^{\circ} \pm 8^{\circ}$ . Simkin (1975) examined the ratio of the minor to major axis using density tracings of wide-field photographic plate images and found the outermost isophotes of the galaxy to suggest an inclination of  $21^{\circ} \pm 5^{\circ}$ . The nearly circular shape of the galaxy disk in the HI 21 cm maps presented by Bosma et al. (1977) also support a face-on orientation.

This is at odds with what has generally been determined from optical imaging alone. For example, surface brightness decomposition of optical ground-based imaging with a modest FOV generally provides a minor-to-major axis ratio of  $\sim$  0.7 for the galaxy disk (e.g., Bentz et al. 2009), suggesting an inclination of 46°. However, as argued by Davies (1973), this elongated structure is likely to be more representative of the weak galaxy bar than the disk, although they also note that there are difficulties with this interpretation as well.

Thus the inclination of NGC 4151 is somewhat uncertain (Simkin 1975 go so far as to describe NGC 4151 as "pathological"!), but is likely to be in the range of  $20-45^{\circ}$ .

For a nearly face-on orientation, it is very difficult to estimate the intrinsic flattening q from the photometry (even a very flat disk would still appear nearly round in projection). Therefore, our strategy is to explore the parameter space of the inclination angle i and intrinsic flattening q independently. Specifically, we first take the 1d surface brightness profile I(R) and deproject it under the assumption of spherical symmetry, obtaining the luminosity density profile

$$j(r) = -\frac{1}{\pi} \int_{r}^{\infty} \frac{dR}{\sqrt{R^2 - r^2}} \frac{dI}{dR}.$$

We then assume that the intrinsic luminosity density is actually axisymmetric and ellipsoidally stratified, with the axis ratio  $q \equiv z/R$  being a free parameter in the model, where q = 1is spherical. If such a density profile is projected face-on, the result would be identical to I(R), while for a nonzero inclination angle *i*, the projected minor axis is  $q' = \sqrt{1 - (1 - q^2) \sin^2 i}$  times smaller than the projected major axis. To preserve the angularly-averaged surface brightness profile, we multiply the scale radius of the 3d profile by  $s \equiv 1/\sqrt{q'}$ , so that the geometric mean of the projected major and minor axes is the same as in the spherical model. For instance, with an inclination  $i = 30^{\circ}$  and flattening q = 0.25, the projected major and minor axes are 1.07 and 0.935 times the original scale radius of the one-dimensional surface brightness profile. We believe that the extra flexibility allowed by independent variation of flattening and inclination (with additional constraints coming from kinematics) is more important than an accurate reproduction of the density profile per se.

#### 3.4. Galaxy Distance

When performing SD modeling, the distance to the galaxy is a large source of uncertainty in determining  $M_{\rm BH}$ . The scale or radii of the galactic features being probed and the value of  $M_{\rm BH}$  scale linearly with the assumed distance. Based on the information available at the time, Onken et al. (2014) adopted a recessional velocity distance measurement of 13.9 Mpc from Pedlar et al. (1992), cautioning that this distance was highly uncertain but that no better estimates were currently available.

Indeed, the only other information available was a Tully-Fisher distance to NGC 4151 that was reported by Tully et al. (2009) as 3.9 Mpc and is highly-discrepant from the group-averaged distance to NGC 4151 of  $11.2\pm1.1$  Mpc. Bentz et al. (2013) recalculated the group-averaged distance to NGC 4151, excluding NGC 4151 itself, and reported a distance of  $16.6\pm3.3$  Mpc, albeit based on only three galaxies.

Since then, additional studies have worked to pin down the distance more accurately. Hönig et al. (2014) measured the physical size of the inner radius of the dusty torus using broad-band RM and the angular size of the torus with near-infrared interferometry, reporting a dust-parallax distance of  $19.0^{+2.4}_{-2.6}$  Mpc to NGC 4151. Tsvetkov et al. (2019) reported the discovery of a Type IIP supernova in NGC 4151.

TABLE 2 NIFS PSF CHARACTERIZATION

Component	Weight	σ (")
1	0.339	0.045
2	0.306	0.082
3	0.356	0.409

Adopting a standardizable candle approach, they find a distance of  $16.6\pm1.1\,\mathrm{Mpc}$ , but they also report a distance of  $20.0\pm1.6\,\mathrm{Mpc}$  if they instead adopt an expanding photosphere analysis.

Most recently, a Cepheid distance to NGC 4151 has been obtained with HST WFC3 optical and near-infrared imaging (GO-13765, PI: B. Peterson), resulting in a distance of 15.8 $\pm$ 0.4 Mpc (Yuan et al. 2020), which we have adopted for our modeling. At this distance, 1" corresponds to 77 pc.

## 3.5. Point Spread Function

To characterize the PSF of our NIFS datacube, we took a slice at a wavelength corresponding to an AGN broad emission line and subtracted off a slice at a wavelength corresponding to the stellar and AGN continuum, in effect creating narrowband images with an on-band image and an off-band image. The residual image thus contains an image of the unresolved AGN point source at a wavelength corresponding to a broad emission line, allowing for characterization of the PSF.

Using GALFIT, we modeled the PSF image with multiple two-dimensional Gaussian components, each having common centers but different widths  $(\sigma)$  and flux contributions. The final three-component model with circular Gaussian components is shown in Table 2. This three-component model accurately matches the wings and the core of the PSF. If only two components were used, the fit was not as good but the values agreed more closely with what was found by Onken et al. (2014).

For the SAURON datacube, the PSF was characterized by Dumas et al. (2007) as a single Gaussian component with width  $\sigma = 2.000$ .

## 4. KINEMATIC ANALYSIS

In the optical and near-infrared, the spectra of galaxies are dominated by the summation of spectra of luminous stars. The line-of-sight velocity distribution (LOSVD) gives the full range of projected stellar kinematics along a particular lineof-sight. To describe the bulk motions of unresolved populations of stars at each spatial position, the convolution kernel necessary to transform stellar template spectra into the absorption profiles of observed galaxy spectra must be determined. In addition to a central wavelength shift indicating typical velocity V, the detailed shapes of the convolution kernels can be approximated with higher-order Gauss-Hermite terms, including the width of the profile  $\sigma$  (related to the velocity dispersion),  $h_3$  (related to skewness),  $h_4$  (related to kurtosis),  $h_5$ , and  $h_6$ . Data with S/N>30 is generally necessary for making sure that the higher order moments of the Gauss-Hermite polynomials can be constrained (Bender et al. 1994). We employed the Penalized Pixel-Fitting method (pPXF; Cappellari & Emsellem 2004; Cappellari 2017), which was developed to constrain the Gauss-Hermite approximation to absorption line profiles through determination of the shifting and blurring required to match a stellar absorption template to an observed galaxy spectrum.

#### 4.1. Kinematic Position Angle

The kinematic position angle was derived using the method of Krajnović et al. (2006). The kinematic measurements derived from an initial pPXF run were smoothed and bisymmetrized before determining the best-fit systemic velocity along with the kinematic position angle of the system as measured from the y-axis of the data cube. This angle represents the line of maximum galaxy rotation, perpendicular to the axis of rotation, and for our kinematics is 37.2° counterclockwise from the y-axis.

The photometric major axis of the galaxy is somewhat uncertain. At intermediate radii ( $\gtrsim 30^{\prime\prime}$ ), the galaxy is visibly elongated in a direction almost perpendicular to the kinematic major axis (see Figure 1), while at smaller radii (within the footprint of the SAURON dataset, shown by a green rectangle), its elongation is still offset by some 30° from the kinematic major axis. Furthermore, 21 cm imaging of the galaxy shows a nearly circular gas disk extending several arcmin beyond the high surface brightness stellar features, with a position angle of 26° (Davies 1973). The intrinsic galaxy shape is thus non-axisymmetric and radius-dependent. However, in the present study we do not attempt to model this complex morphology and instead assume an intrinsically axisymmetric shape. The use of axisymmetric models requires that the kinematic and photometric major axes are aligned, which motivates the constraint on the position angle during the photometric fit.

#### 4.2. Voronoi Binning

Next, we employed the adaptive binning method of Cappellari & Copin (2003). This binning procedure is based on Voronoi (1908) tessellations and assigns adjacent spaxels into bins in a scheme that approximates constant signal-to-noise (S/N) across the FOV. Bins that are similar in shape to the spaxels are prioritized, i.e. not overly irregularly shaped, even at the expense of some irregularity in the S/N. This Voronoi binning method preserves the high spatial resolution at the center of the galaxy where the S/N is highest while adaptively binning the spaxels at the edges of the field to increase their collective S/N.

The data and noise cubes were first cropped, ensuring that the center of the FOV was defined by the centroid of the AGN position, and that spaxels at the edges of the cube with very low counts due to dithering between exposures were rejected. This left us with  $49 \times 49$  spaxels, with the odd number of spaxels in each dimension ensuring that the central spaxel contained the AGN. We then produced both a slice of the data cube (the signal) and a slice of the noise cube that well-represented the cubes as a whole, focusing on the region between the  $16200\,\text{Å}$  CO (6-3) bandhead and the strong [FeII] emission line at  $16440\,\text{Å}$ . These signal and noise slices were then used to assign the individual spaxels to bins.

We identified the Voronoi bin pattern for only one-quarter of the FOV, bounded by the kinematic major and minor axes, and then we replicated the bin pattern for the other three quadrants of the FOV. This scheme ensures that each bin has symmetric analogs in the other quadrants of the FOV, while the central spaxel containing the AGN was assigned to its own single bin. With this approach, we divided the FOV into 165 total bins.

Once the Voronoi bin assignments were determined, the spaxels that were assigned to each particular bin were combined for both the data cube and the noise cube. For all the

spaxels assigned to a particular bin, the spectra from the data cube were co-added, channel by channel, and the spectra from the noise cube were added in quadrature. Voronoi binning the NGC 4151 integral field spectroscopy is an improvement of the analysis over that of Onken et al. (2014), where the cubes were rectified with a constant spatial sampling of  $0.2 \times 0.2$  across the FOV, thus degrading the spatial resolution in the crucial region near the black hole sphere of influence.

#### 4.3. NIFS Kinematic Fits

The NIFS spectra cover the wavelengths  $\sim 15100-17700\,\text{Å}$ . Notable features within the wavelength range include forbidden Fe emission on the blue side, including the strongest line at 16440 Å, and hydrogen Brackett 11 and 10 lines on the red side at 16811 Å and 17367 Å, respectively. We focused on the wavelength range  $\sim 15200-16400\,\text{Å}$ , which included the  $\sim\!16200\,\text{Å}$  CO (6-3) bandhead and blueward absorption. The spectra were fit within the wavelength ranges  $\sim 15150-15330, 15450-15500, 15530-16000, 16060-16130,$  and  $16180-16350\,\text{Å}$ , the same as Onken et al. (2014). These ranges satisfactorily masked the strong emission lines while focusing the analysis on the strongest expected stellar absorption features.

As LOSVDs can be described by Gauss-Hermite polynomials, when the noise of the spectra is high or the data are undersampled with a low velocity resolution, the proper use of pPXF is to bias or penalize higher order Gauss-Hermite terms toward zero, defaulting them to more Gaussian shapes. In the original study, Onken et al. (2014) adopted no penalty. Based on the details of the NIFS integral field spectroscopy, however, we adopted a BIAS of 0.3. We also examined the largest changes in kinematic measurements between BIAS = 0.3 and BIAS = 0 to determine which bins produce unreliable kinematic measurements that might need to be masked during the modeling process.

While convolving the velocity template spectra to match the observations, pPXF provides the option to include Legendre polynomials (Whittaker & Watson 1920) to improve the fit to the continuum. The additive polynomials (DEGREE) can mediate some of the effects of template mismatch (where poor fits of kinematics are obtained due to templates that incompletely represent the observations) and imperfect sky subtraction in the data reduction, while the multiplicative polynomials (MDEGREE) can aid with slight imperfections in the flux calibration and reduce or remove the need for reddening correction. We found a good balance between the weights of the various polynomial components and the goodness of fits to the galaxy spectra when adopting 2nd order additive and 2nd order multiplicative Legendre polynomials, identical to what was found by Onken et al. (2014). These low-order polynomials improved the overall shape of the best fits to the spectra without introducing localized fluctuations that might have inhibited our ability to constrain the stellar kinematics.

We adopted the point-symmetric two-sided pPXF fitting of our spectra, in which two identically-shaped bins symmetric across the center (x, y and -x, -y) are fed to pPXF and fitted simultaneously. Point-symmetric fitting improves the S/N of the data and provides symmetric measurements that utilize the full information contained in each of the two spectra. This is another improvement in the analysis over that of Onken et al. (2014), where the bins were fit independently and then the output measurements  $(V, \sigma, h_3, h_4)$  were four-fold symmetrized (also known as bisymmetrization, described below)

instead and their errors from individual bins were added in quadrature and divided by two. We opted not to perform such bisymmetrization, in which spectra from four bins at positions  $(\pm x, \pm y)$  are fitted simultaneously, for two reasons. First, bisymmetrized kinematic maps hide any non-axisymmetric features (although we currently explored only axisymmetric models, in the future we plan to extend the analysis to a more general geometry). Second, in the point-symmetrized case we have essentially two independent variants of kinematic maps (one in the opposing pair of quadrants x > 0, y > 0and x < 0, y < 0, the other – in the second pair x > 0, y < 0and x < 0, y > 0; here x and y are coordinates aligned with the kinematic major axis). Even though for most models we fitted both halves of the dataset simultaneously (and hence the model maps, which are bisymmetric by construction, lie halfway between the two variants of the observed maps), we can fit the two variants separately to quantify the systematic differences in the model parameters arising from the asymmetries in the data - this is explored later in Section 5.2 and in Figure 8.

In our tests, we found that the G star was only used in 20% of the best-fit spectra, and even when it was used its contribution had a low weight, and so we omitted it as a template. The M star contributed 98% of the time and the K star contributed 97% of the time, and so we included both the K and M templates in our final fit. This is a slight difference from the method of Onken et al. (2014), who used only the M star template in the final kinematic fits.

With all of these adopted parameters, we determined the final stellar kinematics for the central region of NGC 4151 as shown in Figure 4. We were not able to reliably recover the stellar kinematic signature at the center because of the very bright AGN, so for the NIFS data we omit these bins by masking the central  $3\times3$  spaxels. This left 156 bins in our final kinematic maps, which are shown in Figure 4

In summary, the stellar kinematic maps presented here are based on the following improvements over the analysis procedures of Onken et al. (2014): (1) an improved telluric absorption correction; (2) the inclusion of noise information carried through the full reduction pipeline (rather than assuming a constant noise of 2%); (3) the adoption of Voronoi binning to preserve the native spatial resolution near the galaxy nucleus and maintain a constant S/N across the field; (4) the adoption of a penalty term in the implementation of pPXF; and (5) fitting the co-added symmetrized spectra in Voronoi bins with pPXF instead of symmetrizing the pPXF measurements in 0.2"bins.

## 4.4. SAURON Kinematic Fits

We also reanalyzed the SAURON data of NGC 4151. The data cube was provided to us having already been Voronoi binned, with 482 total bins that were not symmetric across the FOV. We followed many of the same procedures as the original analysis carried out by Dumas et al. (2007), but we updated a few details, including the use of the MILES stellar template library of Sánchez-Blázquez et al. (2006) with the updates and corrections of Falcón-Barroso et al. (2011). The library contains 985 empirical stellar spectra covering 3525 – 7500 Å with a well-constrained spectral resolution of 2.51 Å (Beifiori et al. 2011). We focused on a subset of 148 stars to consider as templates, selecting those stars that were in common with both the Indo-US and Elodie libraries (Valdes et al. 2004; Prugniel & Soubiran 2001; Soubiran et al.

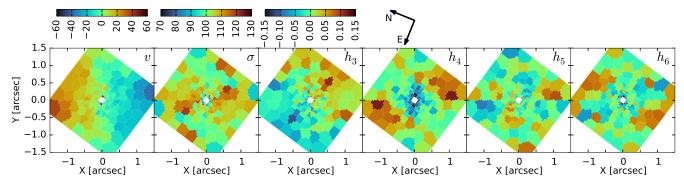


FIG. 4.— Kinematics measured from the NIFS data. X axis is aligned with the kinematic major axis (position angle  $202^{\circ}$  from North through East). The color bar for Gauss-Hermite coefficients  $h_4 - h_6$  is the same as for  $h_3$ 

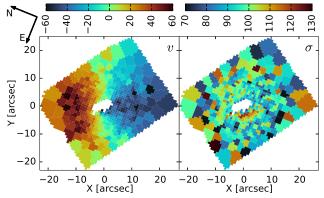


FIG. 5.— Kinematics measured from the SAURON data. The orientation and the color scales are the same as in Figure 4. The central region is excluded due to AGN contamination.

1998; Katz et al. 1998) and spanned spectral types F0 to K7. During the fitting with pPXF, we included the entire spectral range of 4825 – 5380 Å except for small regions around the strong AGN emission lines, including H $\beta$  and [OIII]  $\lambda\lambda 4959,5007$  Å. Following Cappellari et al. (2011), we determined a single best-fit template spectrum that was then adopted for all the bins, to avoid issues that might arise in the derived kinematics from imperfect velocity calibration of the various template stars in the library. We adopted a BIAS parameter of 0.4 and additive and multiplicative Legendre polynomials of degree 4 and 0, respectively. Because the data cube was already binned with a non-symmetric binning pattern, we were unable to carry out the kinematic fits using the two-sided point-symmetric option, so each bin was fit individually. Only the first two Gauss-Hermite moments, V and  $\sigma$ , were included in the fit, with the other terms set to zero. We then masked several dozen AGN-contaminated spaxels within a few arcseconds from the center, retaining 447 bins in the final dataset, which is displayed in Figure 5.

## 5. DYNAMICAL MODELING

## 5.1. Method

The Schwarzschild (1979) orbit-superposition method is a standard approach for dynamical modelling of galaxies based on integrated-light stellar kinematics. In the last two decades, several independent implementations of this method have been widely used to determine black hole masses: the NUKER code (Gebhardt et al. 2000, 2003a; Thomas et al. 2004; Siopis et al. 2009), the LEIDEN code (van der Marel et al. 1998; Cretton et al. 1999; Cappellari et al. 2002), the MASMOD code (Valluri et al. 2004, 2005), and the HEIDELBERG code (van den Bosch et al. 2008; van den Bosch & de Zeeuw 2010).

In this paper, we use yet another recently developed code FORSTAND<sup>11</sup> (Vasiliev & Valluri 2020). Although it can be applied to galaxies of any geometry, in the present case we limit ourselves to the axisymmetric case. We refer to that paper for a detailed description of the method, and here only briefly summarize the general workflow of orbit-superposition modelling and the particular aspects used in this study of NGC 4151.

- 1. First, we determine the 3d luminosity density profile from the surface brightness profile. As explained in Section 3.3, we vary the assumed flattening q and inclination i independently, and each combination of these two parameters produces a different 3d density profile (although the spherically averaged profiles are all very similar). The *mass* density profile  $\rho(r)$  is  $\Upsilon$  times the *luminosity* density j(r), where the mass-to-light ratio  $\Upsilon$  is another free parameter that is varied at a later stage in the modelling pipeline. We denote the mass density profile with a fixed fiducial value of  $\Upsilon_0$  as the "base-line" profile  $\rho_0$ .
- 2. The total gravitational potential  $\Phi(r)$  contains the contributions from the stars  $\Phi_{\star}$  (which is related to  $\rho_0$  by the Poisson equation), the black hole  $\Phi_{\rm BH} \equiv -GM_{\rm BH}/r$ , and optionally the dark matter halo  $\Phi_{\rm h}$ , for which we adopt a spherical NFW profile with a scale radius  $r_{\rm h}$  and peak circular velocity  $v_{\rm h}$ . For simplicity, we fix  $r_{\rm h}$  at 50" and vary only  $v_{\rm h}$ ; since the inner part of the model is self-similar, it is sufficient to vary only the overall density normalization ( $\rho \propto v_{\rm h}^2$ ).
- 3. For each choice of model parameters i, q,  $M_{\bullet}$ , and  $v_{\rm h}$ , we construct an orbit library by integrating  $N_{\rm orb}$  orbits for 100 dynamical times in the given potential and recording the LOSVDs of each orbit in each Voronoi bin (convolved with the PSF). The initial conditions for the orbits are sampled randomly: the positions are sampled from the stellar density profile, and the velocities from a 3d Gaussian distribution with mean and dispersions given by the solution of an axisymmetric Jeans equation for the stellar density in the total potential. The random sampling ensures that all possible orbits supported by the given potential can enter the library, but the results (kinematic fit quality expressed in terms of  $\chi^2$ ) do depend on the specific random realization, because some realizations might contain orbits that are

<sup>&</sup>lt;sup>11</sup> The code is publicly available, for details see Vasiliev & Valluri (2019).

more suitable for reproducing certain kinematic fluctuations caused by observational errors than other realizations. We stress that the scatter in  $\chi^2$  values between different orbit libraries with the same model parameters is a general consequence of the highly flexible nature of the method, and the ability to consider different sets of initial conditions does not cause this effect but merely uncovers it. Nevertheless, this stochasticity is clearly undesirable when comparing  $\chi^2$  values between models with different parameters, and we use two approaches to reduce its impact: (1) for each series of models with given i, q, and  $v_h$  but different  $M_{\bullet}$ , we use the same set of orbital initial conditions, and (2) we run series of models for several random realizations of the orbital initial conditions and consider the "ensemble average"  $\chi^2$  contours to determine the range of best-fit model pa-

4. We then seek a solution for orbit weights that (1) reproduces the 3d density discretized on a  $20 \times 15$  cylindrical grid in the R, z plane, and (2) minimizes the objective function  $\mathcal{F} \equiv \mathcal{F}_{kin} + \mathcal{F}_{reg}.$  The first term is the measure of fit quality for the kinematic constraints (v,  $\sigma$ ,  $h_3 \dots h_6$ ). The second term is the regularization score, which penalizes large variations between orbit weights w; in our case,  $\mathcal{F}_{\text{reg}} = \lambda N_{\text{orb}}^{-1} \sum_{i=1}^{N_{\text{orb}}} (w_i/\overline{w})^2$ , where the mean orbit weight is  $\overline{w} \equiv M_{\star}/N_{\text{orb}}$ , and the regularization coefficient  $\lambda$  controls the tradeoff between smoothness (for large  $\lambda$ ) and closeness of reproduction of kinematic constraints (for negligible  $\lambda$ ). With a too small value of  $\lambda$ , there is a risk of overfitting (the model trying to reproduce noise in the data), which is undesirable because  $\mathcal{F}_{kin}$  of the best-fit solution exhibits large fluctuations between adjacent points in the parameter space. While a detailed analysis of model sensitivity and fidelity as a function of regularization coefficient is beyond the scope of this study, after some experiments we settle on a value  $\lambda = 100$  that produces reasonably smooth likelihood surfaces.

The solution for orbit weights in step 4, and the corresponding  $\chi^2$  score, are obtained separately for each value of  $\Upsilon$ , but reusing the same orbit library constructed in step 3, only rescaling the velocity in the model by  $\sqrt{\Upsilon/\Upsilon_0}$  (which corresponds to a rescaling of all masses by  $\Upsilon/\Upsilon_0$ ). We start the search from a plausible initial guess for  $\Upsilon$  and increase or decrease it by a factor  $2^{1/16} \approx 1.04$ , until reaching a difference between  $\chi^2(\Upsilon)$  and  $\chi^2_{\min}$  greater than 100. The whole process is then repeated from steps 1–3 for a different choice of other model parameters  $(i, q, M_{\bullet}, \text{and } v_h)$ . In total, we considered a few thousand models, which took  $\sim 10^3$  CPU hours (the code is parallelized for multi-core architectures, so the wall-clock time is far shorter).

A technical detail worth mentioning is that when using Gauss–Hermite moments as observational constraints, one needs to decompose the LOSVDs of orbits in the model in the same basis, using the observed values v and  $\sigma$  as the parameters of the Gaussian–Hermite series for each bin, but expressing the measurement uncertainties  $\delta v$ ,  $\delta \sigma$  as uncertainties on the first two coefficients  $\delta h_1$ ,  $\delta h_2$ , whose measured values are zero by construction. This makes the kinematic objective function  $\mathcal{F}_{\rm kin} = \sum_{b=1}^{N_{\rm bins}} \sum_{c=1}^{6} \left[ (h_{b,c}^{\rm model} - h_{b,c}^{\rm data}) / \delta h_{b,c} \right]^2$  quadratic in orbit weights and amenable to efficient quadratic

optimization solvers. After obtaining the best-fit solution for orbit weights, we then compute the final  $\chi^2$  with respect to the original measured values  $v, \sigma, h_3 \dots h_6$  and their associated uncertainties, which is somewhat different from  $\mathcal{F}_{kin}$ . Nevertheless, the shapes and locations of the minima are similar for both  $\chi^2$  and  $\mathcal{F}_{kin}$  as functions of model parameters.

#### 5.2. Analysis of the Model Grid

Given the relatively large number of free parameters  $(i, q, \nu_h, M_{\rm BH}, \text{ and } \Upsilon)$ , we do not attempt to cover this parameter space exhaustively, but use a multi-stage strategy.

It is clear that  $M_{\rm BH}$  is only constrained by the NIFS kinematics, since the SAURON data has lower spatial resolution, and moreover, we excise the central few arcsec of SAURON dataset because of AGN contamination. At the same time, the small spatial coverage of NIFS makes it insensitive to the dark matter halo normalization  $v_h$ , which is merely a nuisance parameter in the present context. Since our primary goal is to determine  $M_{\rm BH}$ , we first consider a series of models fitted to NIFS kinematics alone and ignore the dark halo.

Figure 6 shows a four-dimensional grid of models in the parameter space of inclination i (increasing from top to bottom rows), intrinsic axis ratio q (increasing from left to right columns), and in each panel, black hole mass  $M_{\rm BH}$  (abscissa) and mass-to-light ratio  $\Upsilon$  (ordinate). The two leftmost panels in the lowest row show examples of models that use the cuspy Nuker density profile, and the best-fit  $M_{\rm BH}$  remains near zero for all such models regardless of i and q. The radius of influence  $r_{\rm infl} \equiv G M_{\rm BH}/\sigma^2$  for a  $2 \times 10^7 M_{\odot}$  black hole and  $\sigma \simeq 100 \; \mathrm{km} \, \mathrm{s}^{-1}$  is  $\sim 0.11$ , at the limit of resolution of both kinematic and photometric data. Naturally, the stellar mass within this radius is comparable to  $M_{\rm BH}$ , but also varies by a factor of a few between the models with cuspy ( $\sim 4 \times 10^7 M_{\odot}$ ) or cored ( $\sim 1.5 \times 10^7 M_{\odot}$ ) profiles. We therefore conclude that a cuspy Nuker profile has too high stellar mass in the innermost region, which obviates the need for a black hole. As this is clearly in contradiction with observational evidence for a black hole as demonstrated by AGN activity, in the rest of the paper we consider only the models with the cored Zhao profile; we obtained very similar results with the cored MGE profile from Onken et al. (2014).

The lowest values of  $\chi^2$  differ between panels, but the contours of  $\Delta\chi^2 \equiv \chi^2(M_{\rm BH},\Upsilon) - \chi^2_{\rm min}$  look similar in all panels, showing a "tilted valley" of acceptable models: higher  $M_{\rm BH}$  values correspond to lower  $\Upsilon$ , such that the total gravitating mass within the region  $\lesssim 0.5-0.0^{\prime\prime}6$  is approximately constant. In most panels, the marginalized 1d profiles of  $\Delta\chi^2$  as a function of  $M_{\rm BH}$ , shown by green curves, have large, shallow minima in the range  $0.5\times10^7 M_\odot \lesssim M_{\rm BH} \lesssim (3-4)\times10^7 M_\odot$ .

Figure 7 shows a series of models constrained by both NIFS and SAURON kinematics, with each panel plotting the contours of  $\Delta\chi^2$  as functions of  $M_{\rm BH}$  and  $\Upsilon$  for a given choice of other parameters. We fix the inclination to  $i=30^\circ$  (the overall trends are similar for other values of i), and consider different choices of flattening q (thickness increases from left to right) and dark matter halo normalization  $v_{\rm h}$  (increases from top to bottom). We plot the contributions of NIFS and SAURON datasets to the total  $\chi^2$  of each model separately by red and blue contours. It is clear that the NIFS contours are very similar to the ones in the third row of Figure 6, independently of  $v_{\rm h}$ , and they constrain a certain linear combination of  $\Upsilon$  and  $M_{\rm BH}$ , as discussed earlier. By contrast, the SAURON contours are insensitive to  $M_{\rm BH}$ , but constrain a combination of

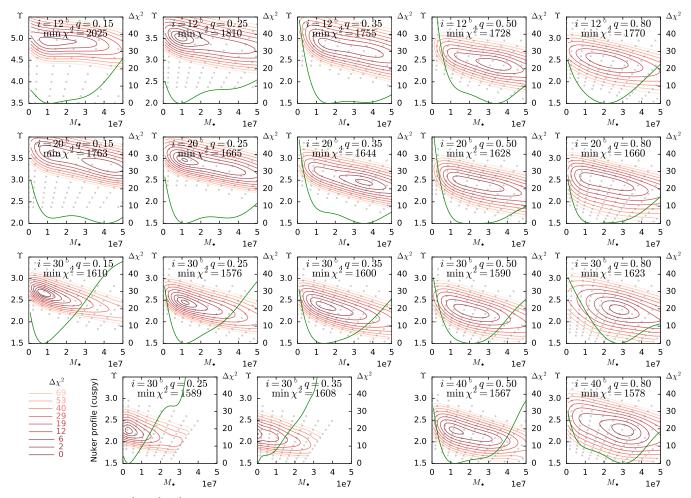


FIG. 6.— Contours of  $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\rm min}$  as a function of two model parameters (black hole mass  $M_{\rm BH}$  and mass-to-light ratio  $\Upsilon$ ) in each panel, for a series of models constrained only by the NIFS kinematics. The contours are placed at  $\Delta\chi^2 = 2.3, 6.2, 11.8, \ldots$ , equivalent to  $1\sigma, 2\sigma, 3\sigma$  confidence intervals for two degrees of freedom. The inclination varies from top to bottom row as  $i = 12^{\circ}, 20^{\circ}, 30^{\circ}, 40^{\circ}$ , and the flattening is q = 0.15, 0.25, 0.35, 0.5, 0.8 from left to right column, except the two leftmost panels in the bottom row, which show models with the cuspy Nuker density profile with  $i = 30^{\circ}, q = 0.25, 0.35$  (the remaining models use our fiducial Zhao profile). The marginalized 1d intervals of  $\Delta\chi^2$  as a function of  $M_{\rm BH}$  alone are shown by green lines in each panel. Gray dots show the actual values of  $M_{\rm BH}$  and  $\Upsilon$  for the grid of models.

 $\Upsilon$  and  $\nu_h.$  One could reasonably guess that a "universally acceptable" model must have compatible best-fit  $\Upsilon$  from both datasets, which indeed minimizes the overall  $\chi^2.$  By adjusting the halo normalization  $\nu_h,$  we can shift the location of best-fit SAURON  $\Upsilon$  up or down, and it is natural to select such a value that maximizes the overlap with best-fit NIFS  $\Upsilon.$  Therefore, the dark matter halo normalization is not really an independent free parameter in our case, but rather should be determined from the consistency between the two datasets.

Comparing the series of models with different flattening q, we see that more disky models (smaller q) have higher best-fit  $\Upsilon$  (especially for low inclination) and correspondingly lower  $M_{\rm BH}$ . This can be understood as follows: since we observe the galaxy at a nearly face-on orientation, the line-of-sight velocity dispersion  $\sigma$  is determined by how far the stars travel in the vertical direction (i.e., thickness) and how large is the restoring force (i.e., the mass density). Indeed, for a thin isothermal disk with a surface mass density  $\Sigma$  and scale height h, the vertical velocity dispersion is  $\sigma = \sqrt{2\pi G \Sigma h}$  (Binney & Tremaine 2008, problem 4.21). Therefore, when decreasing q towards more disky models and hence decreasing h, we must simultaneously increase  $\Upsilon$  and hence  $\Sigma$  to keep the velocity dispersion at the observed value. On the other hand, when

adding a dark matter halo, we increase the gravitating mass and hence reduce the stellar  $\Upsilon$ , but only in the outer parts (i.e., in the SAURON dataset), where stars are not overwhelmingly dominating the total potential. This suggests that the flattening q is largely degenerate with  $\Upsilon$  and  $\nu_{\rm h}$ , but nevertheless we find that the overall  $\chi^2$  is lower for values of q in the range 0.2-0.4

Finally, the assumed inclination does have a significant impact on the model properties. The mean rotational velocity of stars in the equatorial plane  $\overline{v_{\phi}}(R)$  cannot exceed the circular velocity  $v_{\rm circ}(R) \equiv \sqrt{R \partial \Phi / \partial R}$ . In the case of a relatively cold disk, the difference (called asymmetric drift) is  $v_{\rm circ} - \overline{v_{\phi}} \propto \sigma_R^2$ (Binney & Tremaine 2008, equation 4.228). The projected rotational velocity is  $\sim \overline{v_{\phi}} \sin i$ , but for small inclination angles, the line-of-sight velocity dispersion is nearly independent of i, being determined by the mass density and thickness of the galaxy. Thus the observed line-of-sight velocity gradient sets the lower limit on the inclination angle  $i_{min}$ , for which stars need to move on nearly circular ("cold") orbits to produce the given rotational signal. For larger inclinations, stellar orbits have to be "warmer" (have higher eccentricity or even rotate in the opposite direction) in order not to exceed the projected rotational velocity.

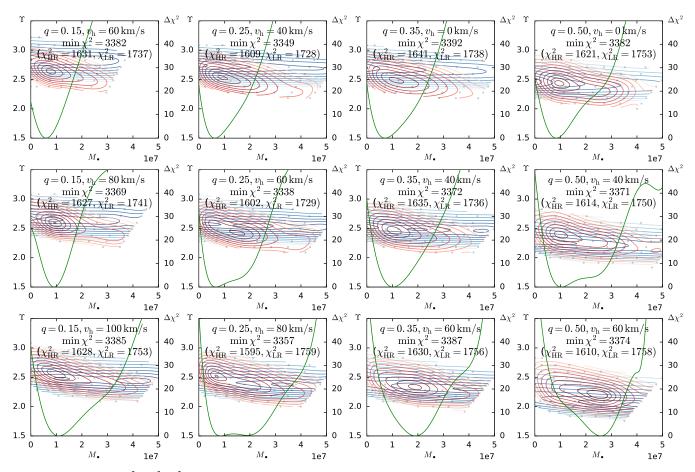


FIG. 7.— Contours of  $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min}$  as a function of two model parameters (black hole mass  $M_{\rm BH}$  and mass-to-light ratio  $\Upsilon$ ) in each panel, for a series of models constrained simultaneously by NIFS and SAURON kinematics. In this case, we plot separately the contours corresponging to both kinematic datasets: NIFS in red (as in Figure 6), SAURON in blue; the spacing between contours is the same as in the previous plot. All models in this series use the Zhao density profile and have the inclination angle  $i=30^\circ$ , while the flattening varies from left to right as q=0.15, 0.25, 0.35, 0.5, and the normalization of the dark matter halo increases from top to bottom, parametrized by  $\nu_h$  (the ranges differ between columns). The marginalized 1d intervals of the total  $\Delta\chi^2$  as a function of  $M_{\rm BH}$  alone are shown by green lines in each panel.

Comparing different rows of Figure 6, we see that models with low inclination generally have higher  $\chi^2$  even if the location of the minimum in the  $M_{\rm BH}-\Upsilon$  plane is not strongly varying (except when both the inclination i and the thickness q are so low that the model is forced to have higher  $\Upsilon$  to match the line-of-sight velocity dispersion). The tendency of orbitsuperposition models to produce better fits at high inclination angles (edge-on orientations), even if the actual orientation is closer to face-on, has been noted in many studies (e.g., Section 5.3 in Thomas et al. 2007, or Lipka & Thomas 2021), and can be attributed to a greater flexibility in rearranging the orbits in the case of sub-maximal rotation. However, as discussed in Section 3.3, NGC 4151 appears fairly round at large radii, and certainly resembles a disk galaxy seen close to faceon rather than an intrinsically round galaxy seen edge-on. For this reason, we only consider models with  $i \leq 30^{\circ}$ , with an exception of two models with  $i = 40^{\circ}$  and  $q \ge 0.5$ . In any case, the range of acceptable values of  $M_{\rm BH}$  does not strongly depend on the inclination, and we take  $i = 30^{\circ}$ , q = 0.25 for our fiducial series of models.

As mentioned in Section 4.3, point-symmetrized NIFS kinematic maps provide two independent variants of observational constraints and allow us to test the sensitivity of global model parameters to some asymmetries in the LOSVD. Indeed, the kinematic maps shown in Figure 4 are not symmetric

with respect to reflection about the kinematic major axis (exchanging the upper and lower halves of the maps), although they are symmetric (for even moments) or antisymmetric (for odd moments) between the opposite quadrants (upper left and lower right quadrants are identical and represent one variant of data reduction, while upper right and lower left quadrants represent another, independent variant). In most cases we fitted both pairs of quadrants simultaneously, but in Figure 8 we show the results obtained from a series of models constrained by only one of the two variants of kinematic maps. The similar morphology of the  $\chi^2$  contours supports the robustness of our results with respect to small variations in the kinematic maps.

Figures 9 and 10 show the NIFS and SAURON kinematic maps, respectively, for a fiducial model with  $i = 30^{\circ}$ , q = 0.25,  $M_{\rm BH} = 1.3 \times 10^7 \, M_{\odot}$ , and  $\Upsilon = 2.5$ , which has one of the lowest  $\chi^2$  values. The maps actually look very similar for other nearby choices of parameters, even though the difference in  $\chi^2$  may be rather significant (of order few tens). Variation of the black hole mass most noticeably affects  $\sigma$  and  $h_4$  maps in the innermost  $\sim 0.12$ .

## 5.3. Black Hole Mass

Having considered the overall trends in the grid of models, we finally return to the main science question of this study –

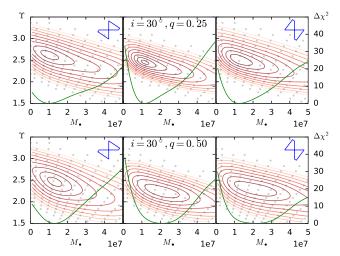


FIG. 8.— Contours of  $\Delta\chi^2 \equiv \chi^2 - \chi^2_{\min}$  for variants of models constrained either by the full NIFS datacube (central panels) or by two independent pairs of quadrants separately. The point-symmetric kinematic maps consist of two essentially independent pairs of opposing quadrants (illustrated by blue boundary polygons in the left and right columns). The central panels are identical to the 2nd and 4th panels in the 3rd row of Figure 6 (inclination  $i=30^\circ$ , flattening q=0.25 and q=0.5); these models are fitted to both variants of kinematic maps simultaneously. The left and right columns instead are fitted to only one of these variants. Despite some differences in the precise location of minima, the shapes of the  $\chi^2$  contours are qualitatively very similar between the three variants (although the contours are necessarily tighter in the central column, which has twice as many constraints), thus we conclude that the results are robust to small variations in the observed kinematics.

the measurement of the black hole mass. Summarizing the above discussion, we find that the addition of the SAURON kinematics does not tighten constraints on  $M_{\rm BH}$ , since it also brings another free parameter (dark matter normalization) that is largely degenerate with  $\Upsilon$ . The limited spatial extent of the SAURON data does not allow us to detect the spatial gradients of dynamical mass-to-light ratio associated with the gradual increase of the halo contribution with radius and hence to disentangle these two parameters. Since the halo properties are irrelevant for the present study, we focus primarily on the NIFS-only series of models (Figure 6), most of which have similar and rather broad ranges of  $M_{\rm BH}$  masses with small  $\Delta\chi^2$ .

It is difficult to establish statistically strict constraints on  $M_{\rm BH}$  for several reasons. First, the discrete nature of orbitsuperposition models necessarily implies some noise in the value of  $\chi^2_{\min}$ . For each of the panels in that figure, we considered several random realizations of the orbit library, which exhibited variations of  $\chi^2_{\min} \sim \mathcal{O}(10)$ , and the locations of the minima were randomly scattered within regions roughly bounded by  $\Delta\chi^2\lesssim 10$  (see, e.g., bottom row of Figure 2 in Vasiliev & Valluri 2020); in the plot, we show the contours averaged across these realizations. Second, the value of  $\chi^2$ per constraint (reduced  $\chi^2$ ) is  $\sim 1.7$ , indicating a moderately poor fit (or, more likely, underestimated measurement uncertainties). Third, there is little systematic study of statistical foundations of the Schwarzschild method, in particular, regarding the rigorous choice of confidence intervals on model parameters in terms of  $\Delta \chi^2$ . This choice cannot be guided by standard considerations applicable to the normally distributed errors (e.g.,  $\Delta \chi^2 = 1, 4, \dots$  for  $1\sigma$ ,  $2\sigma$ , etc.), since this does not take into account the highly non-parametric nature of the method: for each choice of "real" model parameters ( $M_{\rm BH}$ ,  $\Upsilon$ , etc.), we consider the  $\chi^2$  value produced by only one of

 $\begin{array}{c} \text{TABLE 3} \\ \text{Measurements of } \textit{M}_{BH} \text{ for NGC 4151} \end{array}$ 

Method	$M_{\rm BH}~(10^7~{ m M}_\odot)$	Reference
SD Modeling	0.25-3.0	This work
SD Modeling	$4.27_{-1.31}^{+1.31} \\ 3.6_{-2.6}^{+0.9} \\ 3.57_{-0.37}^{+0.45} \\ 2.06_{-0.05}^{+0.05}$	Onken et al. (2014)
GD Modeling	$3.6^{+0.9}_{-2.6}$	Hicks & Malkan (2008)
$H\beta$ RM	$3.57_{-0.37}^{+0.45}$	Bentz et al. (2006)
$H\beta$ RM	$2.06_{-0.05}^{+0.05}$	De Rosa et al. (2018)

NOTE. — All masses from dynamical modeling have been adjusted to an assumed galaxy distance of  $15.8\,\mathrm{Mpc}$ .

many possible combinations of orbit weights ("hidden" parameters), instead of marginalizing over them (see Magorrian 2006 for a discussion).

Pending a rigorous statistical analysis, we adopt a simple qualitative prescription guided by our experiments with a large suite of models. Namely, we take the  $1d \Delta \chi^2$  profile, marginalized over  $\Upsilon$  in each panel, and plot these profiles for several choices of i, q, and in the case of NIFS+SAURON models,  $v_{\text{halo}}$ , to examine the overall range of acceptable  $M_{\text{BH}}$  values. Figure 11 shows  $\sim 20$  such profiles, whose minima span the range  $0.5 \times 10^7 M_{\odot} \lesssim M_{\text{BH}} \lesssim 2 \times 10^7 M_{\odot}$ ; given the inherent noise in  $\chi^2 \gtrsim \mathcal{O}(10)$ , we extend the range of acceptable values to  $0.25 \times 10^7 M_{\odot} \lesssim M_{\text{BH}} \lesssim 3 \times 10^7 M_{\odot}$ . The corresponding mass-to-light ratios lie in the range  $\Upsilon \simeq 2.5 \pm 0.3$ , consistent with the photometric estimates (Section 3.2).

#### 6. DISCUSSION

#### 6.1. Comparison with Previous Results

The results presented in the previous section generally agree with those of Onken et al. (2014), which is unsurprising given the large error bars in both studies. Onken et al. (2014) found  $M_{\rm BH} = (3.76 \pm 1.15) \times 10^7 \, M_{\odot}$  but assumed a distance of  $D = 13.9 \, \rm Mpc$ . When adjusted for the recently published Cepheid distance to NGC 4151 of  $D = 15.8 \pm 0.4 \, \rm Mpc$  (Yuan et al. 2020), their best-fit mass becomes  $M_{\rm BH} = (4.27 \pm 1.31) \times 10^7 \, M_{\odot}$ . Our range of  $M_{\rm BH}$  values is somewhat lower, which may be attributed to a somewhat lower velocity dispersion  $\sigma$  in the central parts inferred from our reanalysis of the observational data. Their mass-to-light ratio  $\Upsilon_H = 0.34 \pm 0.03$  is quoted for a different frequency band, and is generally consistent with our inferred  $\Upsilon_V = 2.5 \pm 0.3$ , as discussed in Section 3.2.

We also list in Table 3 all other  $M_{\rm BH}$  determinations for NGC 4151 from other direct methods. As we did for the Onken et al. (2014) SD mass, we have adjusted the GD modeling mass to account for our adopted distance of 15.8 Mpc. We have also adjusted the RM masses so they have the same adopted value of  $\langle f \rangle = 4.3$  (Grier et al. 2013). The SD mass range we report here generally agrees with the two RM masses from Bentz et al. (2006) and De Rosa et al. (2018) and the H<sub>2</sub> GD modeling mass of Hicks & Malkan (2008).

## 6.2. Limitations of the Dynamical Models

Despite the improvements in the data reduction and analysis process, our uncertainties on  $M_{\rm BH}$  are much larger than in previous studies of the same galaxy, primarily due to the fact that we have explored a wide range of parameters but have only reported results for marginalizing over  $\Upsilon$ . We acknowledge that the intrinsic shape and orientation of the galaxy are not well constrained, and instead of considering only one

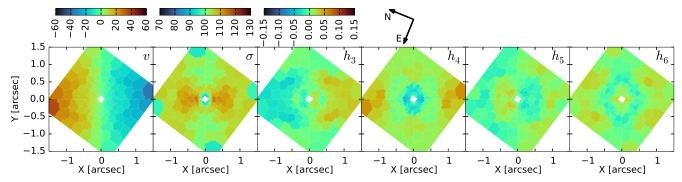


FIG. 9.— Kinematic maps of the central region (the NIFS dataset) for the fiducial model with  $i = 30^{\circ}$ , q = 0.25,  $M_{\rm BH} = 1.3 \times 10^7 \, M_{\odot}$ , and  $\Upsilon = 2.5$ ; these maps can be directly compared to the observations shown in Figure 4.

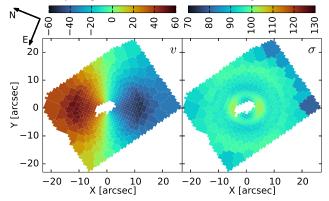


FIG. 10.— Kinematic maps (v and  $\sigma$ ) of the larger region (the SAURON dataset) for the model with  $i=30^\circ$ , q=0.25,  $Mbh=1.3\times10^7\,M_\odot$ ,  $\Upsilon=2.5$ , and  $v_h=60\,\mathrm{km\,s^{-1}}$ , which can be compared to the observations shown in Figure 5. The input data did not have higher-order Gauss–Hermite moments, but we did constrain them in the model to have zero mean and associated formal uncertainty of 0.05; the actual values in the models are indeed fairly small and contribute negligibly to the total  $\chi^2$ .

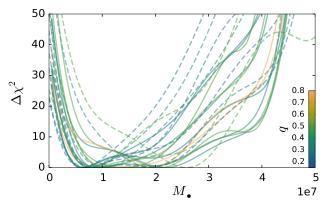


FIG. 11.— Marginalized 1d plots of  $\Delta\chi^2$  as a function of  $M_{\rm BH}$ . Each curve shows the difference  $\chi^2-\chi^2_{\rm min}$  for a series of models with a fixed inclination, flattening (shown by color) and dark matter fraction, where the minimum value varies between series. Solid lines show NIFS-only models, which generally have broader range of acceptable black hole masses, and dashed lines – NIFS+SAURON models.

formally best-fit choice of these nuisance parameters, we accepted the  $M_{\rm BH}$  values from a broad range of models. However, Figures 6, 7 demonstrate that even for a given choice of galactic geometry, there is a relatively large uncertainty on  $M_{\rm BH}$  made possible by suitable rearrangement of orbits in our extensive orbit libraries. This confirms the expectations stemming from the analysis of mock data in Section 3.2 of Vasiliev & Valluri (2020), but stands at odds with more optimistic con-

clusions about the precision of  $M_{\rm BH}$  measurements reached by some other studies, e.g., Neureiter et al. (2021). As these experiments did not use the same setup, it is clear that a more thorough investigation of this question is needed in the future.

The rather weak constraints on  $M_{\rm BH}$  that we obtain in this study may seem overly pessimistic, but we recall that this galaxy is a rather difficult case from the observational perspective, since the AGN strongly dominates the light profile and limits the precision of kinematic measurements at small radii. For our estimated mass range  $0.25 \times 10^7 M_{\odot} \lesssim M_{\rm BH} \lesssim 3 \times 10^7 M_{\odot}$ , and our measured central value of  $\sigma \sim 100 {\rm km \ s^{-1}}$ in the inner few pixels, the sphere-of-influence of the black radius of influence of the black hole is barely resolved by the innermost few pixels in the NIFS dataset. Moreover, the stellar luminosity profile in the same central region is also very difficult to constrain due to overwhelming AGN brightness, and as we have seen, for a plausible cuspy stellar profile consistent with observations, dynamical models prefer no black hole, which is clearly counterfactual. These considerations highlight the importance of spatially resolving the sphere of influence both photometrically and kinematically, especially in late type galaxies where black hole masses and stellar velocity dispersions are typically smaller than in elliptical galax-

Previous authors (e.g. Gültekin et al. 2009; Batcheldor 2010; Gültekin et al. 2011) have focused on the estimated parameters of the  $M_{\rm BH}$ - $\sigma$  relation when  $M_{\rm BH}$  measurements that do not resolve the black hole sphere-of-influence are included or excluded from the sample. Gültekin et al. (2011) found, for a sample of  $M_{\rm BH}$  in early type galaxies with central velocity dispersions with median  $\sigma \gtrsim 260 {\rm km \ s^{-1}}$ , that excluding measurements which do not resolve the radius-of-influence of the black hole biases the estimated  $M_{\rm BH}$ - $\sigma$  relation. They concluded that instead of being excluded these measurements should be included as upper limits or with large error bars. For the elliptical galaxies and massive bulges in these studies, it is argued based on empirical evidence from repeated measurements, that resolving the radius-of-influence only helps to reduce the errors on the black hole mass estimate.

Using mock long-slit kinematic data Valluri et al. (2004) showed that if the radius-of-influence of a black hole is not resolved, nuclear kinematical data can be fitted without a black hole. As far as we are aware the work presented in this paper is the first modeling study in a late type galaxy that clearly demonstrates the drawbacks of not resolving the influence radius on the stellar dynamical  $M_{\rm BH}$  measurement. Despite the fact that the AGN makes it difficult to resolve the influence radius of the black hole, its very presence necessitates a black

hole. Yet we saw that because the central luminosity profile is not resolved below 0."1, the assumption of a cuspy Nuker density profile (which is generally considered reasonable for a late type galaxy) would require no black hole to fit the kinematics. Therefore this study clearly illustrates the need for both photometric and kinematic data that resolve the influence radii of supermassive black holes, data that will become more readily available with JWST and the ELTs.

## 6.3. Future Prospects

Efforts to model the RM observations presented by De Rosa et al. (2018) are currently underway and will remove the dependence on  $\langle f \rangle$  and provide a fully-independent RM mass for comparison with the SD and GD-based masses. Such analyses rely on strong AGN variability during the monitoring campaign, as well as careful management of all noise sources and the observing cadence, and have only been possible for a handful of objects thus far. The observations presented by De Rosa et al. (2018) are of similar quality to previously successful RM modeling attempts (Pancoast et al. 2014; Grier et al. 2017a; Williams et al. 2018), and should therefore be sufficient for an accurate mass constraint.

Furthermore, NGC 4151 is the target of an Early Release Science program with JWST (ERS 1364, PI Bentz). Observations with the NIRSpec IFU will probe the nuclear stellar dynamics and will be directly compared with the observations and analysis presented in this work. NIRSpec is expected to provide some crucial advantages over AO-assisted ground-based observations with its stable and diffraction-limited PSF and significantly lower backgrounds. However, NIFS provides a higher spectral resolution, which may allow for more accurate measurements of the higher-order moments of the stellar absorption profiles. We will thus revisit the topic of the black hole mass in NGC 4151 from stellar dynamical modeling in the near future, once JWST has successfully launched.

## 7. SUMMARY

We have presented a full reanalysis of the black hole mass derived from nuclear stellar dynamics in NGC 4151, beginning with the raw data cubes observed with Gemini NIFS and Altair adaptive optics. We implemented several improvements to the data reduction, including modifications to the NIFS pipeline to allow the variance information to be carried through the full reduction process to the final combined cubes. We also improved the spectral measurements derived from the final combined data cubes, preserving the spatial resolution near the central AGN through the use of adaptive binning, and jointly fitting the spectra for point-symmetrized bins rather than bisymmetrizing the Gauss-Hermite terms af-

terwards. With the new orbit-superposition code FORSTAND, we conducted a thorough exploration of the parameter space within the dynamical models, investigating the use of a dark matter halo and including models with a range of  $M_{\rm BH}$ ,  $\Upsilon$ , galaxy inclination i, and bulge flattening q. We have also adopted the first measurement of an accurate and precise distance to NGC 4151 based on observations of Cepheid stars. We find the black hole mass lies within the range of  $0.25\times 10^7\,M_\odot \lesssim M_{\rm BH} \lesssim 3\times 10^7\,M_\odot$ , which is in general agreement with results from reverberation mapping and gas dynamical modeling. Future dynamical modeling of reverberation data as well as IFU observations with JWST will further constrain the mass of the central supermassive black hole in NGC 4151.

We thank the anonymous referee for comments that improved the presentation of this work. CAR and MCB gratefully acknowledge support from the National Science Foundation through CAREER grant AST-1253702 and AAG grant AST-2009230. CAO acknowledges support from the Australian Research Council through Discovery Project DP190100252. MV gratefully acknowledges support from the National Science Foundation through AAG grants AST-1515001 and AST-2009122, and the Space Telescope Science institute through HST-AR-13890.001 and JWST-ERS-01364.002-A. EV acknowledges support from STFC via the Consolidated grant to the Institute of Astronomy. We thank Eric Emsellem for providing SAURON IFU observations of NGC 4151.

This work is based on observations acquired through the Gemini Science Archive and processed using the Gemini IRAF package, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation on behalf of the Gemini Observatory partnership: the National Science Foundation (United States), National Research Council (Canada), Agencia Nacional de Investigación y Desarrollo (Chile), Ministerio de Ciencia, Tecnología e Innovación (Argentina), Ministério da Ciência, Tecnologia, Inovações e Comunicações (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea). This work was enabled by observations made from the Gemini North telescope, located within the Maunakea Science Reserve and adjacent to the summit of Maunakea. We are grateful for the privilege of studying the Universe with observations that were acquired from a place that is unique in both its astronomical quality and its cultural significance. The SAURON observations were obtained at the William Herschel Telescope, operated by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofisica de Canarias.

#### REFERENCES

Bacon, R., Copin, Y., Monnet, G., et al. 2001, MNRAS, 326, 23

Batcheldor, D. 2010, ApJ, 711, L108

Batiste, M., Bentz, M. Ĉ., Raimundo, S. I., Vestergaard, M., & Onken, C. A. 2017, ApJ, 838, L10

Beifiori, A., Maraston, C., Thomas, D., & Johansson, J. 2011, A&A, 531, A109

Bender, R., Saglia, R. P., & Gerhard, O. E. 1994, MNRAS, 269, 785

Bentz, M. C., Denney, K. D., Cackett, E. M., et al. 2006, ApJ, 651, 775Bentz, M. C., Denney, K. D., Grier, C. J., et al. 2013, ApJ, 767, 149

Bentz, M. C., & Katz, S. 2015, PASP, 127, 67

Bentz, M. C., & Manne-Nicholas, E. 2018, ApJ, 864, 146

Bentz, M. C., Peterson, B. M., Netzer, H., Pogge, R. W., & Vestergaard, M. 2009, ApJ, 697, 160

Bentz, M. C., Walsh, J. L., Barth, A. J., et al. 2010, ApJ, 716, 993

Binney, J., & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)

Blandford, R. D., & McKee, C. F. 1982, ApJ, 255, 419

Bosma, A., Ekers, R. D., & Lequeux, J. 1977, A&A, 57, 97

Cappellari, M. 2002, MNRAS, 333, 400

—. 2008, MNRAS, 390, 71

—. 2014, JAM: Jeans Anisotropic MGE modeling method

—. 2017, MNRAS, 466, 798

—. 2020, MNRAS, 494, 4819

Cappellari, M., & Copin, Y. 2003, MNRAS, 342, 345

Cappellari, M., & Emsellem, E. 2004, PASP, 116, 138

Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 413, 813

Cappellari, M., Verolme, E. K., van der Marel, R. P., Verdoes Kleijn, G. A., Illingworth, G. D., Franx, M., Carollo, C. M., & de Zeeuw, P. T. 2002, ApJ, 578, 787

Cretion, N., de Zeeuw, P. T., van der Marel, R. P., & Rix, H.-W. 1999, ApJS, 124, 383

Cretton, N., & van den Bosch, F. C. 1999, ApJ, 514, 704

Davies, R. D. 1973, MNRAS, 161, 25P

Davies, R. I., Thomas, J., Genzel, R., et al. 2006, ApJ, 646, 754

De Lorenzi, F., Hartmann, M., Debattista, V. P., Seth, A. C., & Gerhard, O. 2013, MNRAS, 429, 2974

De Rosa, G., Fausnaugh, M. M., Grier, C. J., et al. 2018, ApJ, 866, 133 de Vaucouleurs, G. H., de Vaucouleurs, A., & Shapley, H. 1964, Reference

Catalogue of Bright Galaxies (Austin: University of Texas Press) de Zeeuw, P. T., Bureau, M., Emsellem, E., et al. 2002, MNRAS, 329, 513

den Brok, M., Seth, A. C., Barth, A. J., et al. 2015, ApJ, 809, 101
Denney, K. D., Peterson, B. M., Pogge, R. W., et al. 2010, ApJ, 721, 715

Dumas, G., Mundell, C. G., Emsellem, E., & Nagar, N. M. 2007, MNRAS, 379, 1249

Emsellem, E., Monnet, G., & Bacon, R. 1994, A&A, 285, 723

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019, ApJ, 875, L1

Falcón-Barroso, J., Sánchez-Blázquez, P., Vazdekis, A., et al. 2011, A&A, 532, A95

Gebhardt, K., Richstone, D., Kormendy, J., et al. 2000, AJ, 119, 1157 Gebhardt, K., Richstone, D., Tremaine, S., et al. 2003a, ApJ, 583, 92 —. 2003b, ApJ, 583, 92

Genzel, R., Pichon, C., Eckart, A., Gerhard, O. E., & Ott, T. 2000, MNRAS, 317, 348

Ghez, A. M., Morris, M., Becklin, E. E., Tanner, A., & Kremenek, T. 2000, Nature, 407, 349

Gravity Collaboration, Abuter, R., Amorim, A., et al. 2019, A&A, 625, L10 Greene, J. E., Seth, A., Kim, M., et al. 2016, ApJ, 826, L32

Grier, C. J., Martini, P., Watson, L. C., et al. 2013, ApJ, 773, 90

Grier, C. J., Pancoast, A., Barth, A. J., et al. 2017a, ApJ, 849, 146

Grier, C. J., Trump, J. R., Shen, Y., et al. 2017b, ApJ, 851, 21

Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, ApJ, 698, 198

Gültekin, K., Tremaine, S., Loeb, A., & Richstone, D. O. 2011, ApJ, 738, 17 Herriot, G., Morris, S., Roberts, S., et al. 1998, Proc. SPIE, 3353, 488

Herrnstein, J. R., Moran, J. M., Greenhill, L. J., & Trotter, A. S. 2005, ApJ, 629, 719

Hicks, E. K. S., & Malkan, M. A. 2008, ApJS, 174, 31

Hönig, S. F., Watson, D., Kishimoto, M., & Hjorth, J. 2014, Nature, 515, 528 Katz, D., Soubiran, C., Cayrel, R., Adda, M., & Cautain, R. 1998, A&A, 338, 151

Kollatschny, W. 2003, A&A, 407, 461

Krajnović, D., Cappellari, M., de Zeeuw, P. T., & Copin, Y. 2006, MNRAS, 366, 787

Lipka, M., & Thomas, J. 2021, MNRAS

Macchetto, F., Marconi, A., Axon, D. J., et al. 1997, ApJ, 489, 579

Magorrian, J. 2006, MNRAS, 373, 425

—. 2019, MNRAS, 484, 1166

McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

McGregor, P. J., Hart, J., Conroy, P. G., et al. 2003, Proc. SPIE, 4841, 1581

Miyoshi, M., Moran, J., Herrnstein, J., et al. 1995, Nature, 373, 127

Neureiter, B., Thomas, J., Saglia, R., Bender, R., Finozzi, F., Krukau, A., Naab, T., Rantala, A., & Frigo, M. 2021, MNRAS, 500, 1437

Onken, C. A., Ferrarese, L., Merritt, D., et al. 2004, ApJ, 615, 645

Onken, C. A., Valluri, M., Brown, J. S., et al. 2014, ApJ, 791, 37

Onken, C. A., Valluri, M., Peterson, B. M., et al. 2007, ApJ, 670, 105 Pancoast, A., Brewer, B. J., Treu, T., et al. 2014, MNRAS, 445, 3073

Pedlar, A., Howley, P., Axon, D. J., & Unger, S. W. 1992, MNRAS, 259, 369 Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

—. 2010, AJ, 139, 2097Peterson, B. M. 1993, PASP, 105, 247

Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, ApJ, 613, 682

Prugniel, P., & Soubiran, C. 2001, A&A, 369, 1048

Roediger, J. C., & Courteau, S. 2015, MNRAS, 452, 3209

Saglia, R. P., Opitsch, M., Erwin, P., et al. 2016, ApJ, 818, 47

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schödel, R., Ott, T., Genzel, R., et al. 2002, Nature, 419, 694

Schwarzschild, M. 1979, ApJ, 232, 236

—. 1993, ApJ, 409, 563

Simkin, S. M. 1975, ApJ, 200, 567

Siopis, C., Gebhardt, K., Lauer, T. R., et al. 2009, ApJ, 693, 946

Soubiran, C., Katz, D., & Cayrel, R. 1998, A&AS, 133, 221

Sver, D., & Tremaine, S. 1996, MNRAS, 282, 223

Telford, O. G., Dalcanton, J. J., Williams, B. F., Bell, E. F., Dolphin, A. E., Durbin, M. J., & Choi, Y. 2020, ApJ, 891, 32

Thomas, J., Jesseit, R., Naab, T., Saglia, R. P., Burkert, A., & Bender, R. 2007, MNRAS, 381, 1672

Thomas, J., Saglia, R. P., Bender, R., et al. 2004, MNRAS, 353, 391 Tsyetkov, D. Y., Baklanov, P. V., Potashov, M. S., et al. 2019, MNRAS, 487,

Tully, R. B., Rizzi, L., Shaya, E. J., et al. 2009, AJ, 138, 323

Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, PASP, 115, 389

Valdes, F., Gupta, R., Rose, J. A., Singh, H. P., & Bell, D. J. 2004, ApJS, 152, 251

Valluri, M., Ferrarese, L., Merritt, D., & Joseph, C. L. 2005, ApJ, 628, 137 Valluri, M., Merritt, D., & Emsellem, E. 2004, ApJ, 602, 66

van den Bosch, R. C. E., & de Zeeuw, P. T. 2010, MNRAS, 401, 1770 van den Bosch, R. C. E., van de Ven, G., Verolme, E. K., Cappellari, M., & de Zeeuw, P. T. 2008, MNRAS, 385, 647

de Zeeuw, P. 1. 2008, MINRAS, 385, 647 van der Marel, R. P., Cretton, N., de Zeeuw, P. T., & Rix, H.-W. 1998, ApJ,

493, 613 van der Marel, R. P., Evans, N. W., Rix, H. W., White, S. D. M., & de

Zeeuw, T. 1994, MNRAS, 271, 99 Vasiliev, E., & Valluri, M. 2019, FORSTAND: Flexible ORbit Superposition

Toolbox for ANalyzing Dynamical models

—. 2020, ApJ, 889, 39

Voronoi, G. F. 1908, Journal für die reine und angewandte Mathematik, 134, 198

Whittaker, E. T., & Watson, G. N. 1920, A Course of Modern Analysis (Cambridge University Press)

Williams, P. R., Pancoast, A., Treu, T., et al. 2018, ApJ, 866, 75

Yuan, W., Fausnaugh, M. M., Hoffmann, S. L., et al. 2020, ApJ, 902, 26

Zhao, H. 1996, MNRAS, 278, 488 Zibetti, S., Charlot, S., & Rix, H.-W. 2009, MNRAS, 400, 1181