



**HAL**  
open science

## Three-Dimensional Optimal Spectral Extraction (TDOSE) from integral field spectroscopy

K. B Schmidt, L. Wisotzki, T. Urrutia, J. Kerutt, D. Krajnović, E. C Herenz,  
R. Saust, T. Contini, B. Epinat, H. Inami, et al.

► **To cite this version:**

K. B Schmidt, L. Wisotzki, T. Urrutia, J. Kerutt, D. Krajnović, et al.. Three-Dimensional Optimal Spectral Extraction (TDOSE) from integral field spectroscopy. *Astronomy and Astrophysics - A&A*, 2019, 628, pp.A91. 10.1051/0004-6361/201935857 . hal-02325144

**HAL Id: hal-02325144**

**<https://hal.science/hal-02325144>**

Submitted on 22 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Three-Dimensional Optimal Spectral Extraction (TDOSE) from integral field spectroscopy<sup>★</sup>

K. B. Schmidt<sup>1</sup>, L. Wisotzki<sup>1</sup>, T. Urrutia<sup>1</sup>, J. Kerutt<sup>1</sup>, D. Krajnović<sup>1</sup>, E. C. Herenz<sup>2</sup>, R. Saust<sup>1</sup>, T. Contini<sup>3</sup>, B. Epinat<sup>3,4</sup>, H. Inami<sup>5,6</sup>, and M. V. Maseda<sup>7</sup>

<sup>1</sup> Leibniz-Institut für Astrophysik Potsdam (AIP), An der Sternwarte 16, 14482 Potsdam, Germany  
e-mail: kbschmidt@aip.de

<sup>2</sup> Department of Astronomy, Stockholm University, AlbaNova University Centre, 106 91 Stockholm, Sweden

<sup>3</sup> IRAP, Université de Toulouse, CNRS, CNES, UPS, Toulouse, France

<sup>4</sup> Aix Marseille Univ., CNRS, CNES, LAM, Marseille, France

<sup>5</sup> Univ. Lyon 1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon (CRAL) UMR5574, 69230 Saint-Genis-Laval, France

<sup>6</sup> Hiroshima Astrophysical Science Center, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan

<sup>7</sup> Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands

Received 9 May 2019 / Accepted 13 June 2019

## ABSTRACT

The amount of integral field spectrograph (IFS) data has grown considerably over the last few decades. The demand for tools to analyze such data is therefore bigger now than ever. We present a flexible Python tool for Three-Dimensional Optimal Spectral Extraction (TDOSE) from IFS data cubes. TDOSE works on any three-dimensional data cube and bases the spectral extractions on morphological reference image models. By default, these models are generated and composed of multiple multivariate Gaussian components, but can also be constructed with independent modeling tools and be provided as input to TDOSE. In each wavelength layer of the IFS data cube, TDOSE simultaneously optimizes all sources in the morphological model to minimize the difference between the scaled model components and the IFS data. The flux optimization produces individual data cubes containing the scaled three-dimensional source models. This allows the efficient de-blending of flux in both the spatial and spectral dimensions of the IFS data cubes, and extraction of the corresponding one-dimensional spectra. TDOSE implicitly requires an assumption about the two-dimensional light distribution. We describe how the flexibility of TDOSE can be used to mitigate and correct for deviations from the input distribution. Furthermore, we present an example of how the three-dimensional source models generated by TDOSE can be used to improve two-dimensional maps of physical parameters like velocity, metallicity, or star formation rate when flux contamination is a problem. By extracting TDOSE spectra of  $\sim 150$  [OII] emitters from the MUSE-Wide survey we show that the median increase in line flux is  $\sim 5\%$  when using multi-component models as opposed to single-component models. However, the increase in recovered line emission in individual cases can be as much as 50%. Comparing the TDOSE model-based extractions of the MUSE-Wide [OII] emitters with aperture spectra, the TDOSE spectra provides a median flux (S/N) increase of 9% (14%). Hence, TDOSE spectra optimize the S/N while still being able to recover the total emitted flux.

**Key words.** methods: data analysis – methods: observational – techniques: imaging spectroscopy

## 1. Introduction

With the advent of large-area three-dimensional (3D) integral field spectrographs (IFSs) and multi-object spectrographs over the last few decades, larger spectroscopic samples of galaxies have become available. In particular with the growth of the field of view (FoV) of the IFSs, areas of more than 100 square arcminutes now have complete spectroscopic coverage down to the limiting depth of the observations, which are often well below fluxes of  $10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ \AA}^{-1}$ .

In particular, the optical Multi Unit Spectroscopic Explorer (MUSE; Bacon et al. 2010, 2014) on ESO's Very Large Telescope (VLT) and the *Hobby-Eberly* Telescope Dark Energy Experiment (HETDEX; Hill & HETDEX Consortium 2016; Hill et al. 2012) are current IFS facilities mapping large areas on the

sky more efficiently than has previously been possible. Since its start of operations in 2014, MUSE has been instrumental in providing sensitive wide-area (Herenz et al. 2017; Urrutia et al. 2019) and deep pencil-beam surveys (Bacon et al. 2017, 2015) with complete medium-resolution spectroscopic coverage. Most of these data have been taken over already well-known legacy fields with extensive ancillary photometric data available, but has nevertheless revealed new understanding and insights about the general galaxy population thanks to its blind spectroscopic nature. Among these results, it is worth noting the spectroscopic identification of Ly $\alpha$  emitting galaxies undetected in even the deepest existing *Hubble* Space Telescope (HST) photometry (Inami et al. 2017; Maseda et al. 2018), the discovery of ubiquitous extended Ly $\alpha$  halos in Ly $\alpha$  emitters (Leclercq et al. 2017; Wisotzki et al. 2016, 2018, Saust et al., in prep.), and a likely bias in previous estimates of the faint end of the Ly $\alpha$  luminosity function (LF; Drake et al. 2017a,b; Herenz et al. 2019). These studies, focused on Ly $\alpha$  emission, were all enabled by the

<sup>★</sup> TDOSE version 3.0 presented in this paper is publicly available at <https://github.com/kasperschmidt/TDOSE> and Schmidt (2019).

wide-area blind IFS searches with MUSE. Detailed studies of samples of more nearby galaxies have also become considerably more sophisticated in recent years thanks to the advancement of the IFS capabilities. In particular, the significant progress has been driven by dedicated IFS surveys on individual objects like SAURON (de Zeeuw et al. 2002), SINS (Förster Schreiber et al. 2009), ATLAS3D (Cappellari et al. 2011), CALIFA (Sánchez et al. 2012, 2016; García-Benito et al. 2015; Husemann et al. 2013), SAMI (Scott et al. 2018; Green et al. 2018; Allen et al. 2015), MaNGA (Bundy et al. 2015), and KMOS-3D (e.g., Wisnioski et al. 2015). Similarly, the capability of efficiently surveying large areas on the sky, has become possible, for instance by taking advantage of the 3D capabilities of the HST grisms (e.g., Schmidt et al. 2013; Nelson et al. 2016a,b, 2013; Vulcani et al. 2015, 2016, 2017; Wang et al. 2017), or by exploiting the IFSs surveying the sky for HETDEX or as part of the numerous MUSE programs currently being carried out. Taking advantage of the large FoV and increased sensitivity, studies with MUSE have presented parameters including the metallicity, kinematics, emission line diagnostics, and cluster masses of samples of both nearby and distant objects (e.g., Contini et al. 2016; Guerou et al. 2017; Drake et al. 2017a; Lagattuta et al. 2017; Poggianti et al. 2017; Swinbank et al. 2017; Finley et al. 2017; Carton et al. 2018; Krajnović et al. 2018; Patricio et al. 2018; Paalvast et al. 2018; Feltre et al. 2018; Mahler et al. 2018). Such studies have only become possible with large complete spectroscopic samples from IFSs. And with new IFS capabilities being planned and becoming available on upcoming telescopes including the *James Webb* Space Telescope (JWST), ESO’s Extremely Large Telescope (ELT), and the Thirty Meter Telescope (TMT), the amount of IFS data available shows no signs of stagnating.

As shown by the studies mentioned above, IFS data are particularly useful for pixel-by-pixel spectral analysis to obtain maps of metallicity, kinematics, and ionization parameters, among other parameters. Samples of integrated one-dimensional (1D) spectra of complete flux limited galaxy samples for population statistics like LF analysis, emission line characteristics, and comparisons with parameters derived from photometric and spectral galaxy models are also key areas where the large amount of IFS data has been transformative.

Irrespective of whether the science case is focused on resolved maps or integrated 1D spectra, a crucial part of any IFS data extraction is accounting for contaminating light. Here contamination refers to any light coming from foreground or background objects not part of the object(s) of interest. To perform such de-blending (i.e., accounting for the flux contribution of all objects to all pixels in the FoV) a spectral extraction accounting for the morphology of individual objects and the wavelength dependent point spread function (PSF) optimizing the signal-to-noise ratio (S/N) of the spectra is needed. Tools for extracting point-sources, for example stars (PampelMuse; Kamann et al. 2013), and extended objects (AUTOSPEC; Griffiths & Conselice 2018) have been developed with this in mind. Such spectral extractions where the source morphology is accounted for are often referred to as optimal spectral extractions.

In this paper, we present a complementary tool for Three-Dimensional Optimal Spectral Extraction (TDOSE) to accommodate the large amount of deep spectroscopic IFS data of galaxies available today. TDOSE builds on the same principles as the spectral extraction tool for crowded stellar fields, PampelMuse, applied to extensive MUSE data by Husser et al. (2016) and Kamann et al. (2016, 2018). However, TDOSE expands these concepts for applicability to nonpoint sources (i.e., galaxies and other nonstellar objects). TDOSE performs simultaneous

extraction taking both the wavelength dependent PSF and the morphology of individual galaxies in the FoV into account to facilitate the de-blending of flux and hence the spectra of neighboring sources. We show that neighboring sources do not need to be distinguishable in the IFS data as long as they are marginally resolved in the reference imaging. This represents one of the main advantages of including prior information from ancillary data in the extraction as opposed to only relying on the IFS data alone.

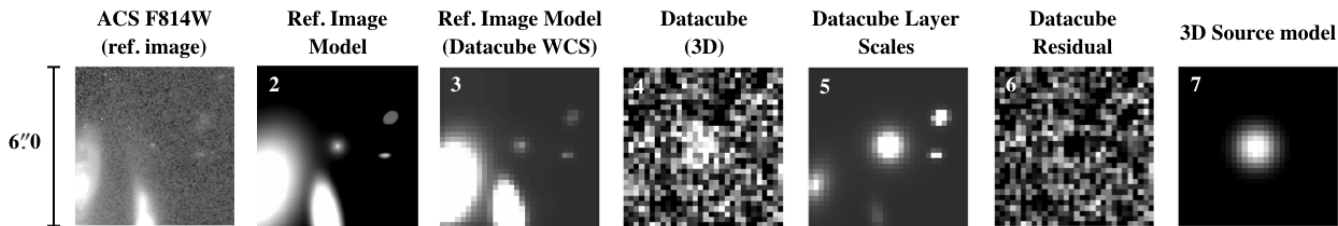
Throughout this paper a “source” refers to a light source, which does not necessarily correspond to a single object. An “object” corresponds to a collection of sources, such that a multi-component galaxy can be extracted combining the fluxes of multiple sources (e.g., different [OII] regions, spiral features, or extended halos).

The paper is structured as follows. In Sect. 2 we discuss the term “optimal extractions”. In Sect. 3 we describe the framework of TDOSE and the individual stages of spectral extractions performed with the software. We describe the capabilities and main limitation of the software, using examples of spectral extractions from MUSE data cubes in Sect. 4. This include examples of recovering spectra only partially covered in the IFS data, de-blending sources, comparing extractions based on single-source and multi-source object models (e.g., generated with GALFIT; Peng et al. 2010, 2002), and correcting spatial maps of galaxy properties by correcting IFS data cubes for contaminating flux. In Sect. 5 we give a summary, and conclude the paper. Appendix A describes the main routines and setup files of TDOSE and Appendix B provides a few examples of execution sequences to perform spectral extractions and to modify data cubes with TDOSE.

## 2. Optimal spectral extraction

Spectral extraction has been a topic of debate for more than three decades since Horne (1986) and Robertson (1986) formulated a complementary method of performing standard aperture extractions of slit spectra that optimizes the S/N in each pixel of the extracted spectrum. They referred to this as an optimal extraction. With the advent of 3D IFS observations this debate has continued, as an optimal spectral extraction of a 1D spectrum from a 3D IFS data cube should account for the wavelength dependent object morphology, the object’s spectral energy distribution variation, any kinematic effects on the spatial distribution of flux as a function of wavelength, and the instrumental effects and their variations both spatially and spectrally. If this can be done, the S/N of the resulting 1D spectrum would also be optimized, i.e., the weighting between actual signal and pixels contributing mostly noise would be accounted for, as was the case for early descriptions of methods used to perform optimal extractions from slit-based spectroscopy.

Often spectral extractions are performed with a specific science question in mind, and hence becomes dependent on the science case being investigated. For instance, if the goal is to assemble a large sample of spectra for emission line identification and classification, a PSF or white light weighted extraction is often enough to obtain the desired results. Here “white light” refers to an image obtained by collapsing the IFS data cube along the dispersion direction. PSF weighted extraction is also useful for studies involving emission line ratios, like metallicities and BPT studies, but to get a proper star formation rate estimate for example, correct flux estimates are needed as much of the line flux could be spatially extended compared to the continuum (e.g., Förster Schreiber et al. 2009; Nelson et al. 2016a; Wisotzki et al. 2016). For these



**Fig. 1.** Overview of the individual steps and data products produced in the workflow of a standard spectral extraction with TDOSE. *From left to right:* (1) HST ACS F814W reference image, (2) five-component model of the reference image, (3) reference image model converted to the pixel scales of the data cube, (4) one wavelength layer of the 3D data cube to extract the spectra from, (5) flux scaling of the individual model components at the data cube layer and convolved with the data cube PSF, (6) residual between the scaled model components and the data cube, and (7) 3D source model at the given wavelength layer for the central model component. Each panel has a size of  $6''.0 \times 6''.0$  corresponding to a  $30 \times 30$  pixel cutout of the MUSE data cube shown in panel 4.

estimates, noisy aperture spectra (with or without aperture correction factors) increases the uncertainty in the ability to derive the information from the spectra. If the goal is to estimate emission line equivalent widths of sources with well-detected continuum, it is mainly the flux ratios and the S/N on the measurements that are important. Lastly, if the science is focusing on emission expected to deviate from the overall (continuum) morphology of the object, tying the spectral extraction to the object morphology will bias the final results. In these cases an optimal extraction will be far from optimal, as the optimized S/N only applies to extractions where the assumed morphology represents the actual data well.

Examples of the latter are studies exploring the spatial extent of Ly $\alpha$  emission, which is known to deviate significantly from the continuum morphology of the host galaxy (Wisotzki et al. 2016, 2018; Leclercq et al. 2017). More fundamentally, any nebular emission line, which by definition is not coincident with the stars making up the continuum light distribution, will be biased by a spectral extraction tied to the continuum morphology if the extent and light distribution of the two are significantly different. Of course, the significance of such a discrepancy is dependent on the nature of the line, where resonant emission lines, like Ly $\alpha$ , must be considered to be the more extreme cases.

Therefore, depending on the science question the extracted spectra are intended to address, alternative methods for spectral extraction might be advisable. Generally, for studies where obtaining high S/N of the spectrum is the driving factor, an optimal extraction that accounts for the object morphology and optimizes S/N is preferable. TDOSE provides a broadly applicable software package for performing such optimal spectral extraction from 3D IFS data cubes.

### 3. TDOSE

TDOSE is a versatile Python software package for extracting one-dimensional spectra and de-blending flux from 3D IFS data cubes. In this paper we describe version 3.0 of TDOSE (Schmidt 2019) but the current front-end version of TDOSE is always available on GitHub<sup>1</sup>. The main purpose of TDOSE is to optimally extract the flux for individual sources in a given FoV accounting for both the object morphology and the flux from neighboring contaminating sources. However, as different science cases potentially require different extraction approaches, TDOSE also enables aperture extractions and PSF weighted point source extractions. Given that all three methods are performed within the same framework and conserve flux, the data products are easily comparable.

<sup>1</sup> <https://github.com/kasperschmidt/TDOSE>

Figure 1 illustrates the individual steps in the workflow of a standard 3D optimal spectral extraction with TDOSE. The workflow can be divided into three main stages:

1. Determining the sources in the reference image and generating a two-dimensional (2D) morphological model for those (Sect. 3.1);
2. Converting the reference image model to the IFS reference frame, and determining the flux contribution from each source at each wavelength layer (Sect. 3.2);
3. Combining and de-blending sources in the IFS to extract the 1D spectra of objects in the considered FoV (Sect. 3.3).

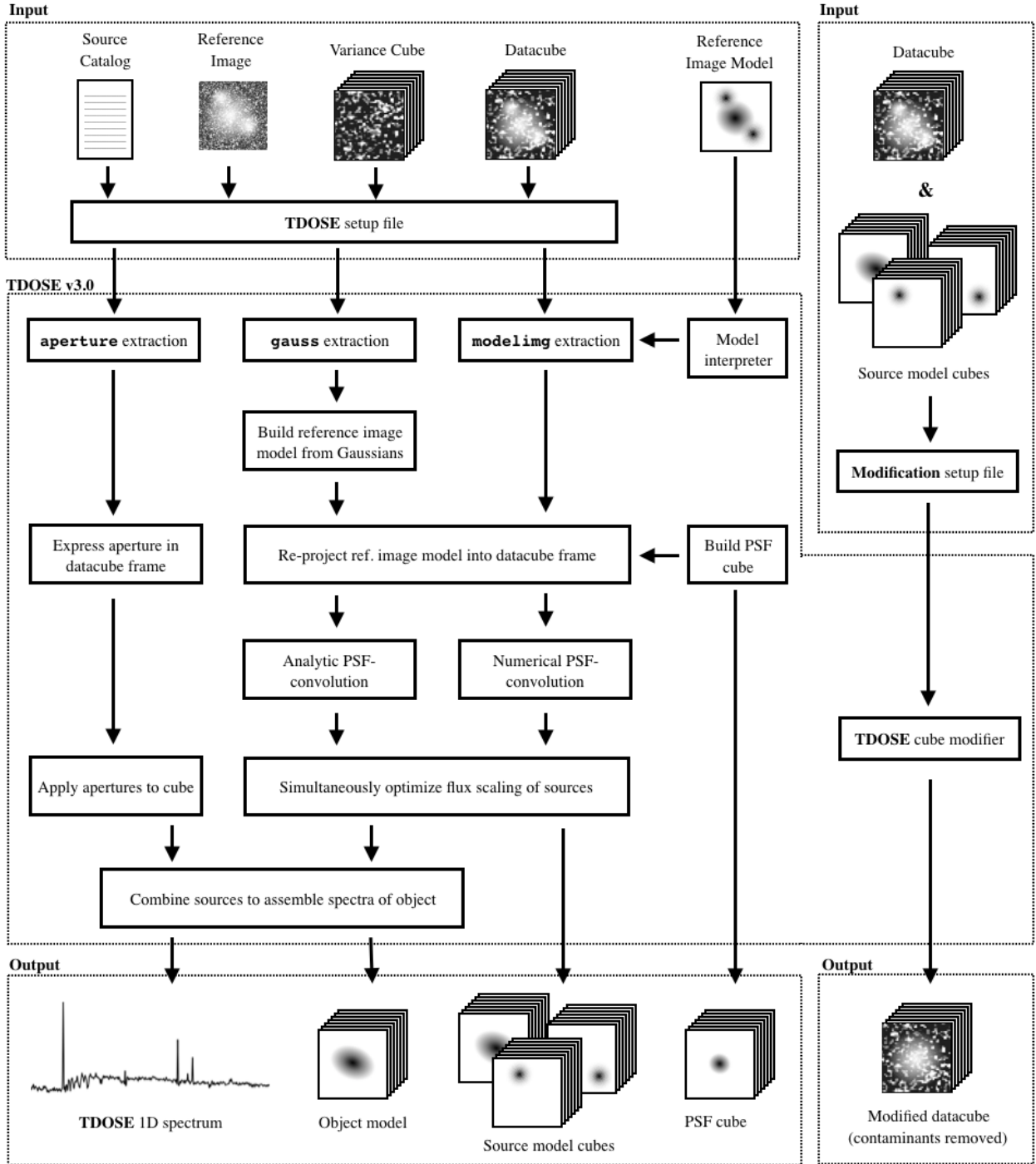
In the following we describe each of these stages in detail, and explain how aperture and point source spectral extractions are also enabled in the TDOSE software package. Figure 2 presents a flow chart of the different spectral extractions, and how IFS data cubes can be modified and corrected for undesirable flux based on the 3D source models generated by TDOSE (see Sects. 3.4 and 4.6). The TDOSE version 3.0 scripts and setup files used to extract spectra and generate the main outputs, some of which are displayed in Fig. 1 and illustrated in Fig. 2, are presented in Appendix A. Appendix B provides examples of a selection of TDOSE tasks and scripts.

The minimum required inputs for TDOSE is a data cube, a variance cube (to propagate and estimate noise on the extracted spectra), a reference image, a model for the PSF wavelength dependence, and a source catalog (see Fig. 2). TDOSE is therefore agnostic to the type of IFS data cube the spectra are actually extracted from, as long as the spatial and spectral dimensions are provided in the FITS cube header. TDOSE was developed with MUSE in mind, and the examples presented in this paper therefore all show MUSE data and spectra. Spectral extractions from both CALIFA and MaNGA data cubes have been performed successfully with TDOSE but are not presented here.

#### 3.1. Determining sources in reference images

TDOSE performs optimal spectral extraction and optimizes the S/N by accounting for the spatial morphology and extent of a given object. Ideally, estimating this morphology would be done on reference imaging with infinite resolution and a depth exceeding that of the IFS data cube to eliminate any bias from instrumental effects. Infinite resolution imaging does not exist, but for essentially all existing IFS data, HST imaging or ground-based imaging taken under good conditions can be used. In Sect. 4 we use HST ACS F814W imaging as reference imaging when defining the sources that contribute flux to the MUSE data cubes. Formally, a reference image of the same resolution as the data cube itself, for instance a white light image, can also be used as





**Fig. 2.** Flowchart of TDOSE version 3.0. The top layer defines the required inputs (source catalog, reference image, variance cube, and data cube) and optional inputs (reference image model and source model cubes) expected by TDOSE either for one of the three extraction methods (aperture, gauss, and modeling; center left) or for the modification of intrinsic data cubes shown on the right. The bottom part of the panel illustrates the outputs generated by each of the reductions and the cube modifier. Input reference image models can be converted to the format expected by TDOSE with a “model interpreter”. Tools for interpreting GALFIT multi-component models and models of individual objects are provided as part of TDOSE (for details on individual steps, formats, and outputs, see Sect. 3 and Appendix A).

the starting point for spectral extractions with TDOSE if ancillary data is unavailable. The use of ancillary (higher resolution) reference images was deliberately avoided in the spectral extraction tool AUTOSPEC (Griffiths & Conselice 2018) to make the method self-contained. However, providing a higher-resolution reference image allows the de-blending of sources that are unresolved at the IFS PSF resolution, which is one of the main

strengths of the approach taken by TDOSE. Avoiding the use of ancillary data could result in extractions of spectra containing flux contamination that could otherwise be avoided. Examples of these scenarios are provided in Sect. 4.4.

After selecting the reference image, it is necessary to determine which sources to base the modeling on. This can be done using standard imaging source detection software like

SExtractor (Bertin & Arnouts 1996). Modeling can only be performed on sources. To be able to account for sources not showing up in the reference image (like emission line sources with faint continuum or sources with abrupt changes in their spectral energy distributions), point sources or manually generated models can be used for the spectral extraction. Hence, combining standard source detection software and source detection tools applied to the IFS data cube itself, like the Line Source Detection and Cataloging Tool (LSDCAT; Herenz & Wisotzki 2017), the MUSE Line Emission Tracker (MUSELET; Bacon et al. 2016), the detectiON and extRactiON of Galaxy emIssion liNes tool (ORIGIN; Bacon et al. 2017; Inami et al. 2017, Mary et al., in prep.), or the Source Emission Line FINDER (SELFI; Meillier et al. 2016), can be useful for assembling the most complete source lists and corresponding source models for the spectral extraction. The important thing to note is that only sources included in the source catalog provided to TDOSE can be accounted for in the spectral extraction.

Having determined which sources to account for in the spectral extraction, a reference image model has to be generated (Fig. 1, panel 2). This model can either be an empirical representation of the source morphologies in the FoV based on binary or weighted flux segmentation regions similar to the segmentation maps SExtractor produces, or it can be a sample of analytic parametric 2D light profiles, like Sérsic (1963) profiles or multivariate Gaussians. The key point is that each source has a unique representation that mimics its reference image morphology or, rather, the expected underlying morphology of the IFS flux.

By default, TDOSE models the sources in the reference image by representing each source in the source catalog by a multivariate Gaussian defined as

$$f(\mathbf{c}) = \frac{1}{2\pi \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2} (\mathbf{c} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{c} - \boldsymbol{\mu})\right), \quad (1)$$

where  $\mathbf{c}$  represents the coordinate set  $(x, y)$  and  $\boldsymbol{\mu}$  contains the mean values  $(\mu_x, \mu_y)$ . The covariance matrix is given by

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}, \quad (2)$$

where  $\rho$  is the correlation between  $x$  and  $y$ . The morphological multivariate Gaussian models are generated and optimized using Scipy<sup>2</sup> (Jones et al. 2001). Should the source list contain objects that are faint or undetectable in the reference image, these would be challenging to model automatically with TDOSE. In these cases, TDOSE can be instructed to add a point source fixed at each source location to the model. When to use point sources and when to trust the model depends, among other things, on the completeness of the source catalog and the quality of the reference image.

Alternatively, a custom 2D model of the (reference image) FoV can be provided to TDOSE. Such a model can be generated with GALFIT models (Peng et al. 2010, 2002). TDOSE provides tools to enable de-blending of the individual model components of GALFIT (see Appendix B). Custom models are treated numerically in the spectral extraction, which in cases of large FoVs increases the computation time and hence the time it takes to extract spectra.

Generating models or providing custom models of the sources in the reference image informs the flux optimization in the second stage of TDOSE about the number and light distribution of the sources to be accounted for during the spectral extraction and source de-bending.

<sup>2</sup> <https://www.scipy.org>

### 3.2. Building a source model cube via flux optimization

Using the information from the reference image source model, TDOSE optimizes the flux distribution of each wavelength layer, assigning fluxes to each source according to its morphological representation in the reference image model. The reference image source model is turned into a cube by convolving the reference image model with the wavelength dependent IFS PSF, after pixelating the high-resolution reference image models to the spaxel size of the IFS (Fig. 1 panel 3). Numerical convolution over large spatial scales at thousands of wavelength layers is computationally expensive. TDOSE version 3.0 therefore uses a Gaussian PSF model and by default the multivariate Gaussian source models, such that the PSF convolution is carried out analytically. Providing the wavelength dependent IFS PSF as an analytic or empirical function has not been implemented in TDOSE yet. For source models with non-Gaussian source components a direct (non-FFT) numerical convolution with the PSF is performed. After transforming the reference image model into a 3D data cube, the flux scalings of the individual source models that best represent the IFS data cube (Fig. 1 panel 4) can be determined by solving the set of linear equations defined by

$$\chi^2 = |\mathbf{A}\mathbf{a} - \mathbf{d}|^2. \quad (3)$$

Here,  $\mathbf{A}$  represents a list of the 3D models for each of the  $n$  sources in the reference image FoV. The factor  $\mathbf{a}$  is a matrix of (flux) scalings for each of the individual source representations in  $\mathbf{A}$ . The matrix  $\mathbf{d}$  represents the IFS data cube that the source models are supposed to represent. Hence, given the source models, by minimizing the  $\chi^2$  expression the flux scalings that best represent the IFS data cube can be derived for all sources simultaneously. The  $\chi^2$  minimization can be done with matrix algebra for each of the  $m$  wavelength layers on the IFS data cube using the ordinary least squares (OLS) estimator in matrix form:

$$\mathbf{a}_m = (\mathbf{A}_m^T \mathbf{A}_m)^{-1} (\mathbf{A}_m^T \mathbf{d}_m). \quad (4)$$

Here  $\mathbf{A}_m$  is a matrix of dimension  $(n, N_{\text{pix}})$  representing the source models for each individual wavelength layer. The models are normalized by the square root of the variance in each voxel of the wavelength layer. The  $\mathbf{d}_m$  is a vector of dimension  $N_{\text{pix}}$  containing the data flux values normalized by the square root of the variance in each of the  $m$  layers. Finally,  $\mathbf{a}_m$  is the total flux in layer  $m$  assuming the models are normalized (Fig. 1 panel 5). Hence,  $\mathbf{a}_m$  gives the optimized flux; when combined with the given source models, this parameter best represents the data in the  $m$ th layer of the data cube  $\mathbf{d}$  (Fig. 1 panel 6). At the wavelength layer of the data cubes displayed in Fig. 1 the central source, a Ly $\alpha$  emitter at  $z = 3.25$ , outshines the other sources in the FoV. The central model component is scaled accordingly at this particular wavelength when performing the extraction with TDOSE.

The approach described above is identical to the approach used in PampelMuse (Kamann et al. 2013; Kamann 2018) developed to extract and de-blend stellar spectra in crowded IFS data cubes, with the exception that in TDOSE  $\mathbf{A}$  consists of extended source models, as opposed to point sources in PampelMuse.

### 3.3. Extracting spectra: de-blending and combining sources into objects

Having minimized the disagreement between the scaled 3D source models and the IFS data cube, the spectrum of any object in the modeled FoV can be extracted. An object consists of any

number  $k$  of the  $n$  source model cubes produced by TDOSE. Any sources that are not included in the object are considered to be contaminants. Due to the algebraic treatment of the  $\chi^2$  minimization described in Sect. 3.2, the flux values are optimized simultaneously for all sources in the FoV ensuring ideal conditions for flux de-blending of neighboring sources. And as the models are based on the high-resolution reference imaging, it is in principle the resolution of the reference image that drives the ability of TDOSE to de-blend objects from contaminating sources. Due to the simultaneous flux scaling determined for all sources, the relative flux contribution of each source in the FoV model to each of the IFS voxels is accounted for, and the object flux cube is obtained by simply summing up the flux-scaled source model cubes of all the sources contributing to the object (Fig. 1 panel 7). Collapsing this object model cube in the spatial dimensions results in the optimally extracted 1D TDOSE spectrum. The 1D TDOSE spectrum, the 3D object model, and the individual source models can all be returned by TDOSE (see Fig. 2). The individual source modules can be used to modify the original IFS data cube (see Sects. 3.4 and 4.6).

The noise on the extracted de-blended 1D spectrum is propagated from the variance of the IFS data cube. Following Eq. (16) of Kamann et al. (2013), TDOSE estimates the noise at each wavelength in the 1D spectrum as

$$N_m = \left( \sum_{i,j} f_{i,j,m}^2 / v_{i,j,m}^2 \right)^{-0.5}. \quad (5)$$

Here  $f_{i,j,m}$  is the fraction of flux in each data cube voxel with respect to the total flux in each wavelength layer  $m$  of the data cube, and  $v_{i,j,m}$  is the variance of each voxel in the data cube. The sum is performed over the spatial indices  $i, j$  of the data cube. Equation (5) results in a S/N at each of the  $m$  wavelengths in the extracted 1D spectrum of

$$\left( \frac{S}{N} \right)_m = \frac{\sum_k \mathbf{a}_{m,k}}{N_m}, \quad (6)$$

where  $\sum_k \mathbf{a}_{m,k}$  represents the sum of the flux scales in the  $m$ th wavelength layer of the  $k$  (out of the total  $n$ ) sources contributing to the object's spectrum.

In the example of the TDOSE extraction workflow presented in Fig. 1, only the source model shown in panel 7 contributes to the object spectrum (i.e.,  $k = 1$ ). The remaining  $n - k = 4$  sources seen in the five-component reference image model (panel 2) are considered contaminants. In Sect. 4.3 we present examples of the gain in flux and S/N that can be obtained from MUSE data using multi-source models instead of single-source models.

As mentioned, a default spectral extraction with TDOSE is based on an object model consisting of only Gaussian sources combined with a Gaussian PSF model. This makes the extraction of spectra fully analytic. So-called multi-Gaussian expansion (MGE) models, also known as mixture of Gaussian (MoG) models, have been shown to successfully represent the light distribution of most galaxy types and morphologies in both 1D and 2D (e.g., Monnet et al. 1992; Emsellem et al. 1994a,b; Bendinelli & Parmeggiani 1995; Kochanek et al. 2000; Cappellari 2002; Hogg & Lang 2013; Scott et al. 2013). This makes the 2D version particularly interesting for optimal spectral extraction from IFS data cubes similar to the one performed by TDOSE. Representing the PSF by a multi-component MGE model itself adds flexibility and precision to the PSF model without removing the benefits of a fully analytic spectral extraction. However, handling of a multi-component MGE PSF model has not yet been implemented in TDOSE.

Simultaneously accounting for and assigning the flux in the IFS data cube to multiple sources in the FoV is exactly what is required for performing reliable de-blending of objects as it keeps track of the fractional contribution of light in each individual voxel in the IFS data cube from all objects in the FoV. It is this information that TDOSE uses to de-blend the objects of interest from contaminating sources in the IFS. The framework that TDOSE and PampelMuse (Kamann et al. 2013) is based on has previously been shown to effectively handle source de-blending in some of the most crowded fields on the sky, namely globular clusters (Husser et al. 2016). In a similar manner, TDOSE reliably de-blends extended objects in crowded fields like galaxy clusters or deep extragalactic exposures. PampelMuse takes advantage of the fact that stars are point sources, and hence are well represented by a scaled PSF at all wavelengths. TDOSE, on the other hand, is designed to handle extended objects and hence requires source modeling; it is worth noticing that the de-blending capabilities of TDOSE are naturally limited by the accuracy of the source model's representation of the actual data, and by the galaxy morphology that changes as a function of wavelength, for example due to emission line regions/extent and continuum color variations. As described in Sect. 2 a key limitation of any optimal spectral extraction tool using a morphological model as prior, and therefore also TDOSE, is that extractions of spectral features that are poorly represented by the continuum morphology described by the reference image model will be biased. Especially capturing the strength of nebular emission lines, which by definition are not described by the continuum light from stars, is potentially biased. To remedy the mismatch between the continuum model and intrinsic wavelength dependent 2D morphology when extracting spectra with TDOSE, several approaches can be taken. For example, the spectral feature (e.g., emission line morphology) can be modeled and extracted independently in smaller spectral regions around the observed emission line wavelength, in an attempt to capture their sizes with dedicated models. Alternatively, secondary halo source components can be added to the intrinsic object model. This provides TDOSE with the opportunity to account for emission that extends beyond the compact central continuum emission by assigning extended emission to a secondary halo source when scaling the model components. For example, the extended Ly $\alpha$  emission of the emitter shown in Fig. 1 is not well represented by the extent of the continuum emission as seen in panel 6 where the extended emission is not fully recovered by the scaled model, and the central region is over-subtracted. We caution the use of continuum models for extracting fluxes of emission lines which are poorly represented by the continuum's 2D surface brightness profile. Instead a more extended halo source component can be added to the reference image source model to account for this. A third alternative is to estimate the potential bias that the continuum-based modeling extraction introduces, and then correct for the bias. Section 4.4 presents a few representative examples of de-blending with TDOSE.

#### 3.4. Removing sources from IFS data cubes with TDOSE

As TDOSE simultaneously models all sources in the FoV, the TDOSE source model cubes can be used to manipulate the original data cube. One of the main outputs from TDOSE is a flux scaled model of each individual source, which can be collapsed and combined into extracted 1D spectra. However, instead of collapsing the individual source model cubes, any number of sources can be subtracted directly from the intrinsic data cube. Hence, TDOSE allows data cubes to be generated where individual sources (or



objects) are removed. This is useful for science cases where the main goal is to work in 2D or 3D, instead of with extracted 1D spectra. Such studies include the description of maps where satellite galaxies or foreground and/or background galaxies are contaminating the flux and therefore affecting estimates of the probed parameters of the main galaxy. With the optimal de-blending of sources from TDOSE, instead of simply masking the contaminating source positions, for example in kinematic, metallicity, emission line, or continuum color maps, unmasked full FoV maps can be generated after the source models contributing to the contaminating objects are removed from the original IFS data cubes. This will increase both the area covered by the maps, and the precision of the estimates in regions affected by the contaminating sources. In addition, searches for extended low surface brightness emission could potentially benefit from the inspection of source-subtracted data cubes. If all sources are removed, the residual data cubes would ideally only contain noise and extended emission not captured by the continuum source models. Section 4.6 presents an example of how to improve a kinematic map of a galaxy by removing contaminating sources. Appendix B.6 provides an example of the TDOSE commands needed to perform modifications to the original IFS data cube.

### 3.5. Point source and aperture extractions

As a supplement to the optimal spectral extraction, TDOSE also performs point source and aperture extractions. A TDOSE point source extraction is simply done by adding a point source to the reference image model, which will then be convolved with the wavelength dependent IFS PSF during the extraction. This enables extractions using point sources in combination with extended sources. Such extractions correspond to standard PSF-weighted extractions, except that TDOSE conserves the flux by normalizing the source models before estimating the flux scalings. As PampelMuse, which is based on the same framework as TDOSE, is dedicated to spectral extraction in crowded stellar fields like globular clusters, we suggest using this software, as opposed to TDOSE, for such applications.

Aperture extractions with TDOSE are performed by representing each source in the input source catalog by a cylinder in the 3D source model skipping the flux optimization step. We note that the apertures are defined based on the high-resolution reference image, and are then converted to the IFS voxel scales afterwards, leading to irregular apertures when the IFS's spatial resolution is comparable to the chosen aperture size (i.e., for small apertures with respect to the spatial resolution of the IFS). Section 4.5 presents a comparison between aperture and model-based spectral extractions with TDOSE.

## 4. TDOSE extractions from MUSE data cubes

In the previous section the theoretical framework of the TDOSE software was outlined and described. In this section we illustrate these concepts and the results obtainable by describing and comparing spectra extracted with TDOSE from MUSE IFS data cubes. In particular, we focus the comparisons and extractions on objects from the MUSE-Wide survey (see Sect. 4.1 below and Herenz et al. 2019; Urrutia et al. 2019).

### 4.1. The MUSE-Wide Survey

The majority of the data presented in this paper (except for the data in Sect. 4.6) were taken from the MUSE-Wide survey

(P. I. L. Wisotzki). MUSE-Wide is part of the MUSE Guaranteed Time Observations (GTO) campaigns, and comprises 100 MUSE pointings of one square arcminute and one-hour exposure, each mosaiced over the CANDELS (Grogin et al. 2011; Koekemoer et al. 2011), GOODS-South, and COSMOS regions. The first data release from MUSE-Wide (DR1<sup>3</sup>) is described in detail by Urrutia et al. (2019), and presents 44 consecutive MUSE pointings collected in CANDELS/GOODS-South. The nine pointings of MUSE-Wide covering the *Hubble* Ultra Deep Field (HUDF) were released (with increased depth) by Bacon et al. (2017). The remaining 16 fields in GOODS-S, the  $2 \times 4$  pointings covering the HUDF parallel fields, and the 23 pointings in COSMOS will be released at a later stage. MUSE-Wide DR1 presents a sample of more than 9000 spectra of photometrically selected objects, including absorption line galaxies, as well as emission line selected galaxies, including more than 400 Ly $\alpha$  emitters (LAEs). As part of MUSE-Wide DR1, we provided 1D spectra extracted with TDOSE of all objects in GOODS-South from the Guo et al. (2013) source catalog. Each MUSE-Wide DR1 spectrum was extracted using the default extraction method of TDOSE version 3.0 (also available in TDOSE version 2.0<sup>4</sup>) described in this paper. Hence, to represent both the main objects and the contaminating objects for all Guo objects in MUSE-Wide DR1, we used multivariate Gaussian models of the HST F814W image morphology. To make the extraction process fully analytic, we used Gaussian models for the IFS PSF. For each MUSE-Wide pointing the wavelength-dependent Gaussian PSF is provided in the master PSF catalog selection of MUSE-Wide DR1 (Table 2 of Urrutia et al. 2019). For faint marginally detected objects in the Guo catalogs, where Gaussian modeling was suboptimal, the point source extractions described in Sect. 3.5 were used.

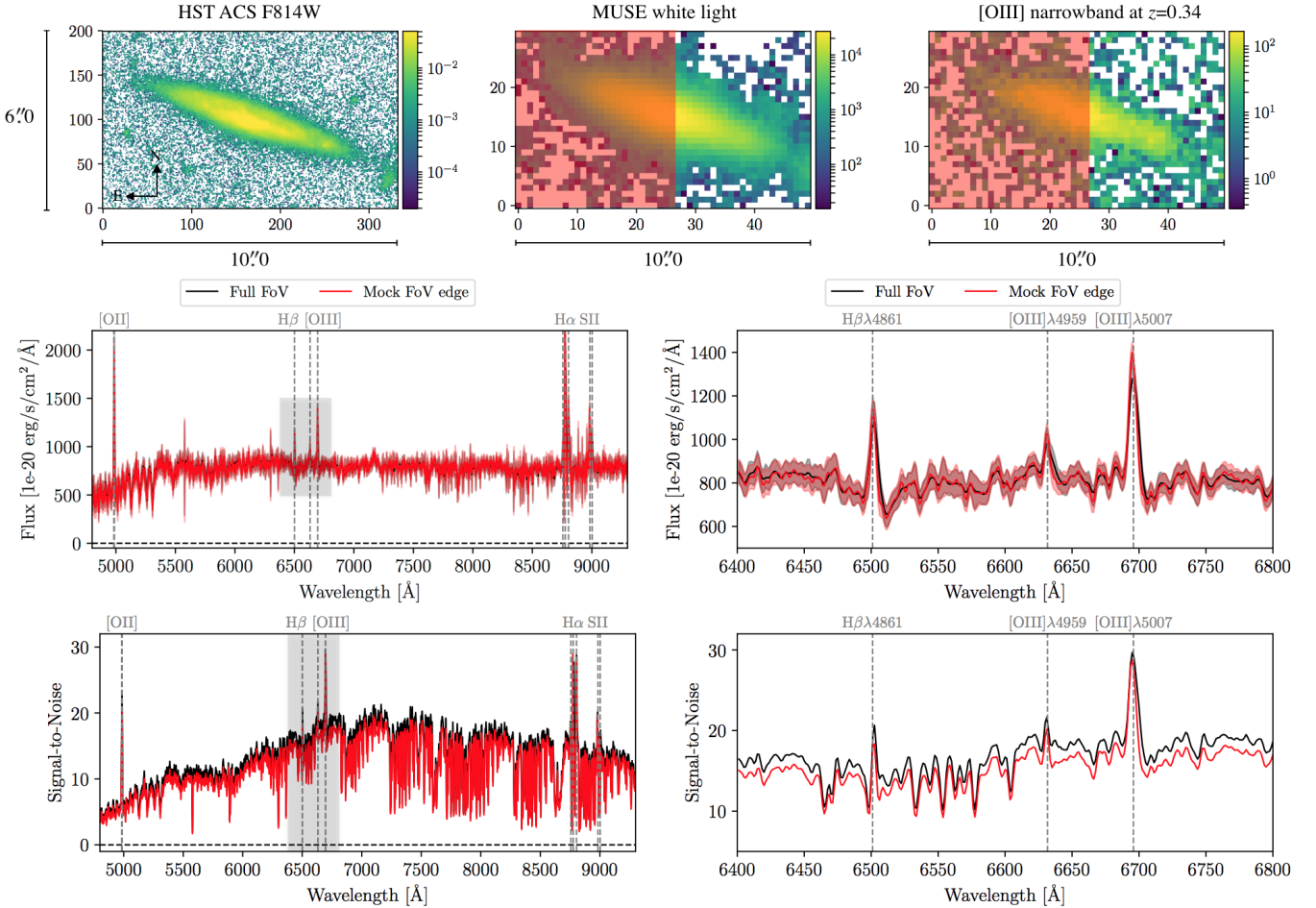
### 4.2. Recovering spectra of sources only partially covered

As TDOSE bases the spectral extraction on a source model of a high-resolution reference image scaled according to the observed flux in the IFS data cube to obtain the resulting spectrum, the intrinsic object flux is predicted at each wavelength based on the input model. This implies that the flux prediction at each wavelength layer of the IFS data cube is insensitive to edges or holes in the data as it is only based on a scaling of the input model, which is assumed to represent the whole galaxy. In Fig. 3 the flux spectrum of ID<sub>Guo</sub> = 10701 (ID<sub>MUSE-Wide</sub> = 125034103) extracted based on the full MUSE data cube (black spectrum) agrees with the flux spectrum extracted from a data cube including a mock edge (red spectrum), where only half of the data (red shaded region in the top panels) were used when determining the flux scalings in Eq. (4). The two extracted spectra agree within a median flux difference below one percent (central panels of Fig. 3), illustrating the ability to recover the intrinsic flux of an object's spectrum given an assumed morphological model even though in this case almost half of the object falls off the IFS detector. On the other hand, the extracted S/N spectra (lower panels) differ by a median value of 10% in the shown example. This loss in S/N is caused by the lower number of voxels available when extracting the spectrum from only half of the data cube, mimicking that the object falls off the IFS detector. Hence, assuming that the object model represents the overall source morphology well (in the example shown in Fig. 3 a single Gaussian model was used), the intrinsic flux can be reliably

<sup>3</sup> <http://musedata.aip.de>

<sup>4</sup> <https://github.com/kaspschmidt/TDOSE/releases>





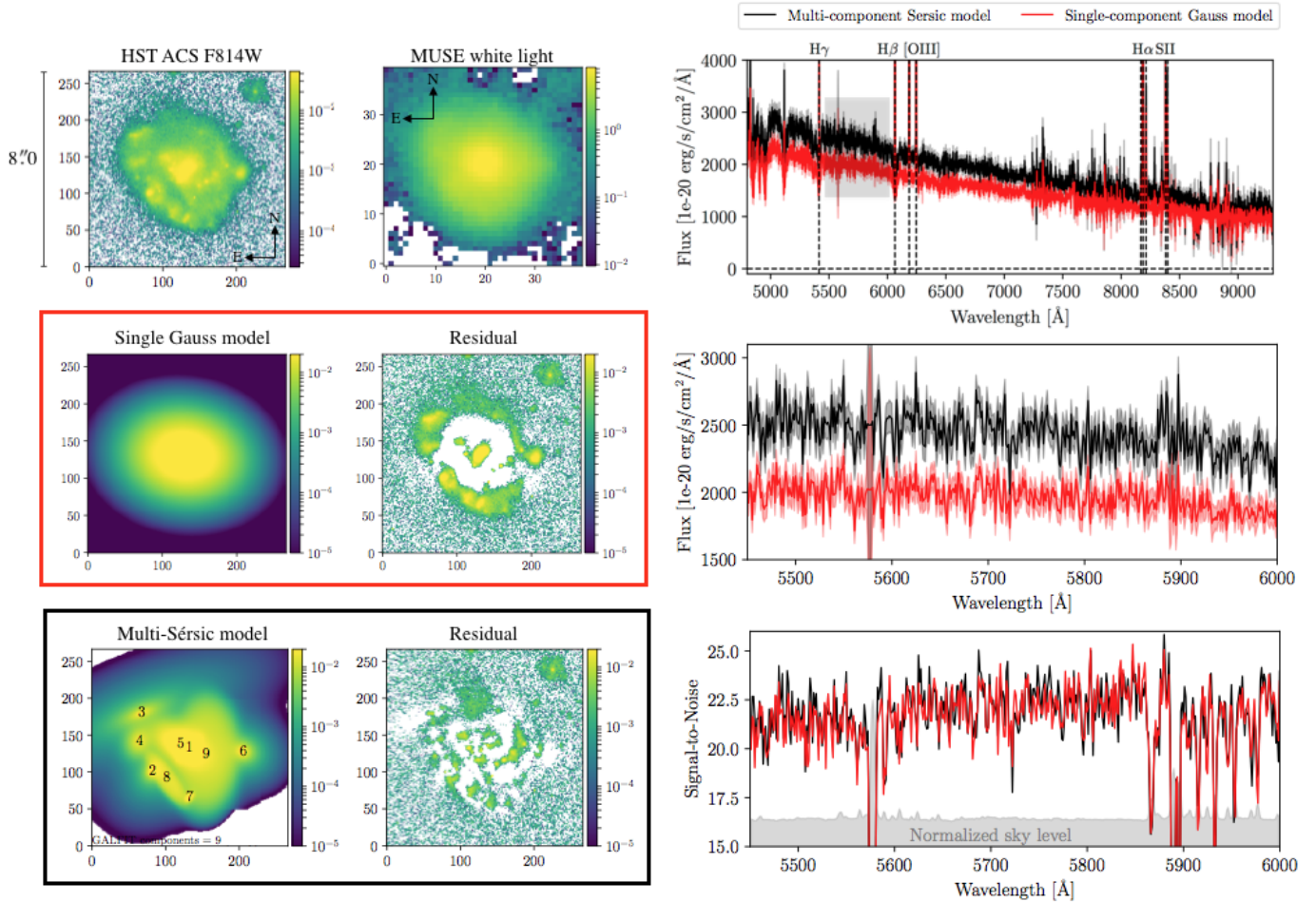
**Fig. 3.** Illustration of the recovery of a spectrum from an object ( $ID_{\text{Guo}} = 10701$ ,  $ID_{\text{MUSE-Wide}} = 125034103$ ) only partially covered (red shaded region) in the IFS data cube. *Top panels (from left to right):*  $6''0 \times 10''0$  postage stamps of the HST F814W image, the MUSE white light image, and a narrowband image of width  $1000 \text{ km s}^{-1}$  (rest-frame) around  $[\text{OIII}]\lambda 5007 \text{ \AA}$ . *Bottom panels:* flux (*top*) and S/N (*bottom*) spectra for the full MUSE wavelength range (*left*) and a zoom-in on the  $\text{H}\beta$  and  $[\text{OIII}]\lambda 4959, 5007 \text{ \AA}$  emission lines region (indicated by the gray box in the left panels). The black spectrum shows the results from a TDOSE extraction based on a single-source Gauss model of the full FoV shown in the top panels. The red spectrum on the other hand shows a TDOSE extraction using the same HST source model, but mimicking that the object falls off the edge of the MUSE detector, by only using the area of the MUSE data shaded in red for the model flux scaling in the TDOSE extraction.

estimated irrespective of missing data using TDOSE with only a minor loss in S/N. Multi-component object models, where individual sources fall fully within the excluded edge region, will naturally be biased as those model components cannot be scaled. The results will also be biased if only a small fraction of voxels are available or regions that poorly reflect the overall light distribution of the object are used for the flux scaling.

#### 4.3. Extractions based on single-source and multi-source models

As described in Sect. 3, TDOSE is based on a simultaneous scaling a pre-defined sample of sources in the reference image model. This model consists of any number of sources that can be combined into spectra of individual objects. A model is generated by the TDOSE software itself, but can also be provided as an input (see Fig. 2), and hence can be generated manually or with existing image modeling software, for instance GALFIT (Peng et al. 2010, 2002). The default modeling approach of TDOSE is to assign one multivariate Gaussian (see Eq. (1)) to each source in the source catalog. This allows for both efficient de-blending (Sect. 4.4) and for flexibility to recover the

intrinsic flux of non-uniform galaxies as accurately as possible. However, as described in Sect. 3.3, depending on the science goal, describing an object by a single source might not be sufficient. Figure 4 presents an example of the difference between a spectral extraction using a single-source object model and combining multiple source models into a single object. The figure shows  $ID_{\text{Guo}} = 10843$  ( $ID_{\text{MUSE-Wide}} = 112003032$ ), a star forming galaxy with pronounced features in its 2D light distribution. The red and black spectra were extracted with TDOSE using the multivariate single-Gaussian source model ( $n = 1$ ), and the multi-component source model ( $n = 9$ ) shown in the bottom left panels. The models and the image residuals clearly show an improvement in the representation of the 2D light distribution of the galaxy when moving from a single-source (red box and spectra) to a multi-source (black box and spectra) object model. The recovered continuum flux from the multi-component model is on average 20% higher than the flux level recovered from the single-component model. The continuum S/N of the two spectra is roughly identical (a median change of 0.5%) due to a higher noise in the spectrum extracted based on the nine-component model compared to the single-Gaussian extraction. However, the  $\text{H}\alpha$  peak S/N increases by roughly 10% when extracting the



**Fig. 4.** Spectral extractions performed with TDOSE for  $ID_{\text{G10}} = 10843$  ( $ID_{\text{MUSE-Wide}} = 112003032$ ), a star forming  $H\alpha\lambda 6563 \text{ \AA}$  emitter at  $z = 0.2475$ . *Left panels:*  $8''0 \times 8''0$  postage stamps of the HST F814W image, the collapsed MUSE white light image, a single-component Gaussian model and the corresponding HST residual (red box), and a multi-component Sérsic model generated with GALFIT and the corresponding HST residual (black box). The nine individual components of the GALFIT model are indicated by the black numbers in the bottom left panel. *Righthand panels:* TDOSE spectra extracted based on the single-component (red) and multi-component models (black) from the MUSE data cube. The black spectrum combines the flux from all nine sources marked in the bottom left panel. *Top right panel:* full spectrum, whereas *bottom right panels:* flux and S/N spectra of the continuum-only region between  $H\gamma$  and  $H\beta$  marked by the gray square in the top right panel. *Lower right panel:* normalized sky flux level responsible for the S/N drops is indicated by the gray shaded region. The median flux increase of  $\sim 20\%$  when using a multi-component model shows the power of basing spectral extractions on multi-component models when the object's light distribution is inhomogeneous.

spectrum based on the more detailed multi-component model as the individual star forming regions seen as subclumps of the galaxy (and potential differences in their kinematics; see Sect. 4.7) are better represented in the multi-component object model. This illustrates the power of combining models of individual sources into single objects when extracting TDOSE spectra for galaxies of inhomogeneous 2D light profiles.

The object shown in Fig. 4 is particularly well suited for a multi-source model representation. To quantify the effects of using multiple versus single source models when extracting spectra of objects in more general terms, we considered a sample of [OII] emitters from MUSE-Wide DR1. We selected 153 galaxies with apparent magnitudes in the HST F814W filter of  $23 < m_{814} < 24$ , a high-confidence [OII] $\lambda 3726, 3729 \text{ \AA}$  emission line doublet identified in the MUSE data cube (confidence 3; see Urrutia et al. 2019), and a clear match to the photometric Skelton et al. (2014) catalog ( $ID_{\text{Skelton}} < 0.3 \text{ arcsec}$ ). For this sample we generated GALFIT multi-source morphological models (with up to four independent sources) based on the existing F814W imaging. We then extracted spectra with TDOSE using

both the default single-Gaussian modeling approach (which are the spectra released as part of MUSE-Wide DR1), and by providing multi-source models from GALFIT. The top panel of Fig. 5 shows a comparison of the peak flux of the [OII] emission obtained from spectra using the two different source models. A red line shows the best fit linear relation between the two flux estimates, obtained from Scipy's (Jones et al. 2001) orthogonal distance regression (ODR) in which the uncertainties on both parameters are accounted for. The fit to the flux values in the top panel, which is slightly offset from the one-to-one line, indicates that the median increase in peak flux obtained by using an input model with multiple sources of the object of interest is 5%. However, the spread of the data around the best fit shows that the flux increase (or even decrease) gained by using a source model with multiple components can be as high as 50% in extreme cases. Hence, for sample statistics the gain in using models containing multiple sources to model and extract spectra for [OII] emitters in MUSE data is relatively modest, but for individual objects and in special cases the difference can be significant. Given the extent of the selected [OII] emitters, which is often comparable to the

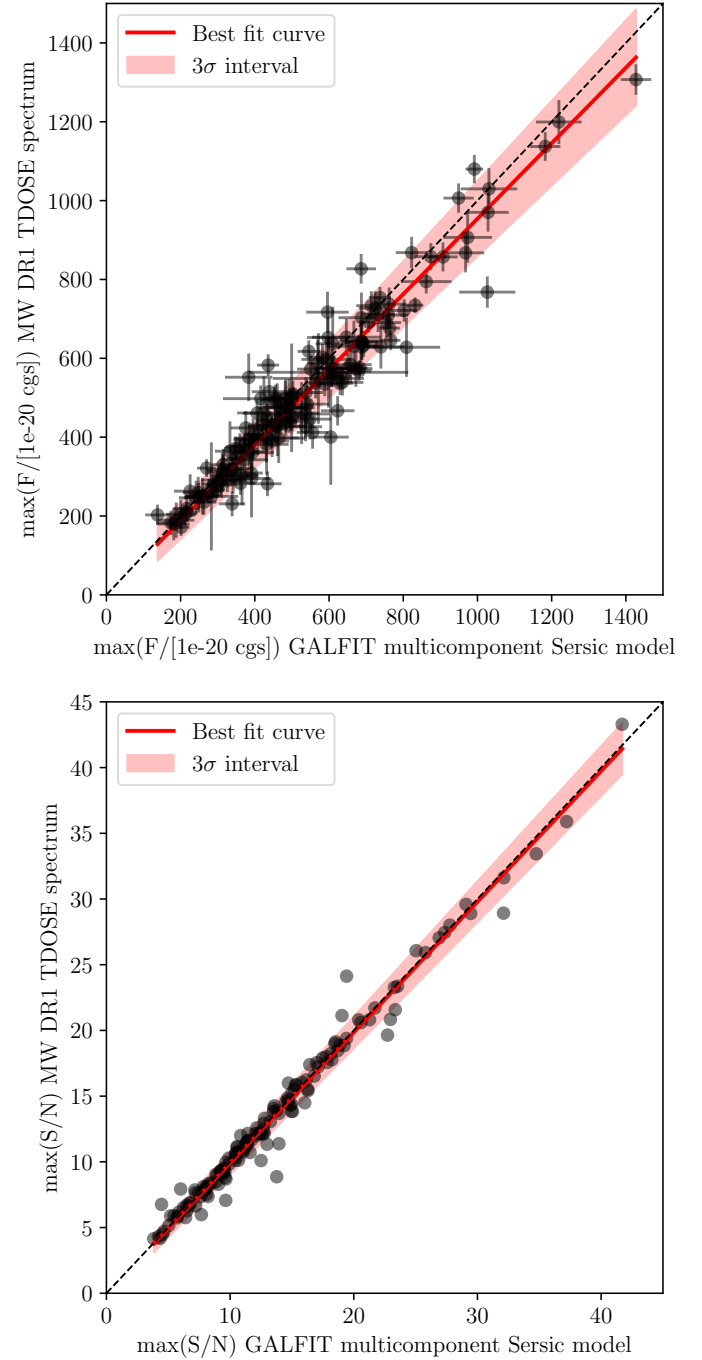
size of the MUSE-Wide PSF ( $\lesssim 1''$ ) it is not surprising that the majority of the differences in the HST-based models are eliminated by convolution with the MUSE-Wide PSF. By comparing the extent of white light, continuum and [OII] emission line narrowband images, we confirmed that the [OII] emission is well represented by the MUSE PSF convolved continuum models and extent. Objects that extend well beyond the IFS PSF scales (like  $ID_{\text{Guo}} = 10843$  shown in Fig. 4) will show more significant differences between extractions based on object models with a single and multiple sources. The bottom panel of Fig. 5 shows the corresponding S/N for the two spectral extractions. These estimates are in good agreement, with just a few outliers. Hence, the main effect of using multiple sources in the object model for the spectral extractions appears to be on the recovered flux levels, as is also confirmed by the extractions for  $ID_{\text{Guo}} = 10843$ .

The above examples and comparisons show that the gain in flux between a simple single-source object model and a more complex model including multiple sources is generally only at a level of a few percentage points, but can in special cases be quite high. The S/N, on the other hand, appears more stable against variations in the object model.

#### 4.4. De-blending of objects

Apart from improving the flux recovery of individual sources, extracting spectra based on models containing multiple sources can be used to efficiently de-blend spectra from independent objects that appear close to each other when projected on the sky. For the case of MUSE data taken without adaptive optics (AO), HST reference images generally have a resolution that is at least five times higher as the PSF FWHM of HST ACS images is generally below  $0''.1$  (Guo et al. 2013). The MUSE-Wide non-AO data were in seeing between  $0''.7$  and  $1''.2$  (Urrutia et al. 2019). This implies that it is straightforward to distinguish and reliably model objects that are blended in the MUSE data.

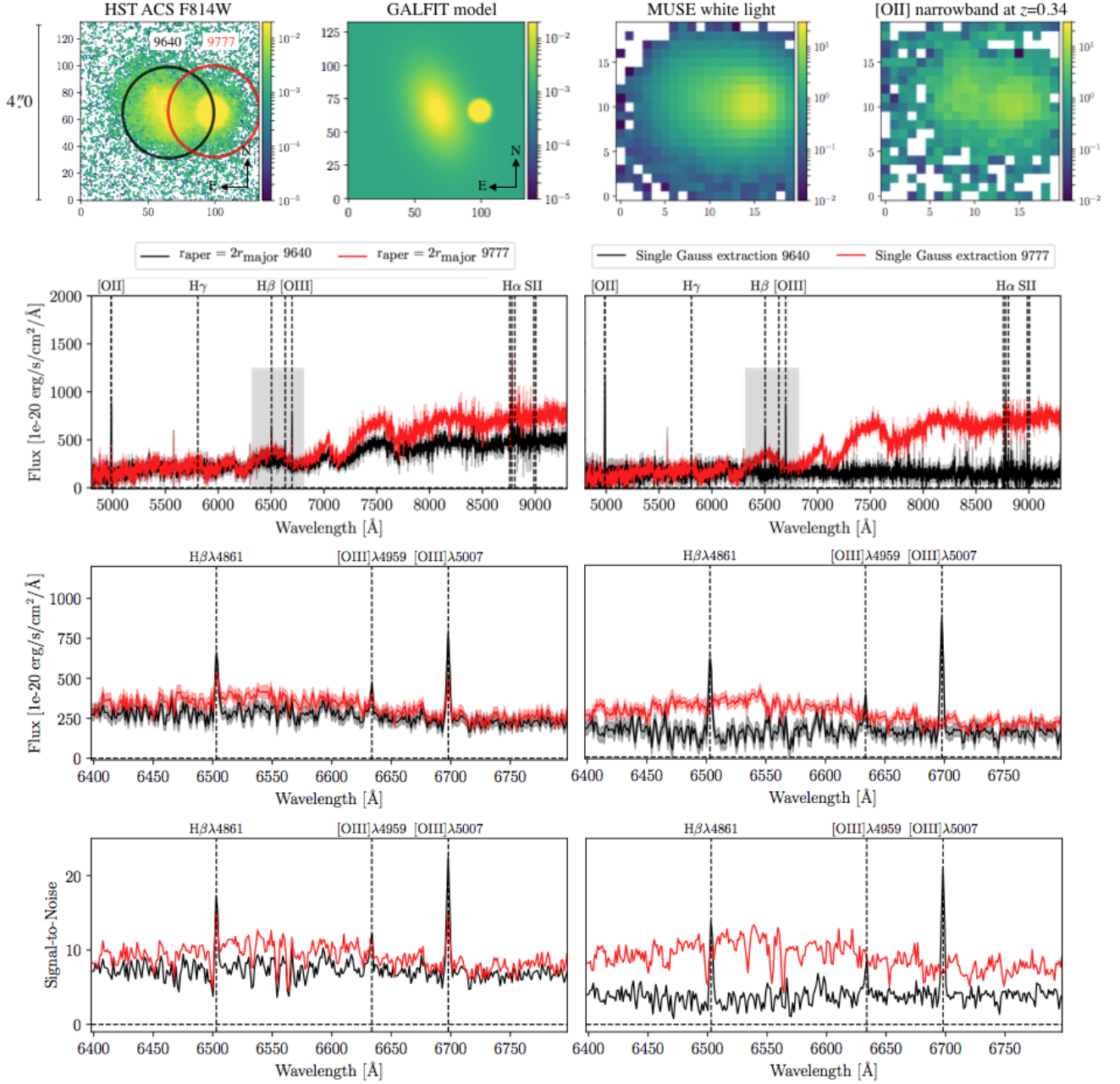
Figure 6 shows images and spectra of the object  $ID_{\text{Guo}} = 9640$  ( $ID_{\text{MUSE-Wide}} = 102009072$ ) at redshift  $z = 0.3377$ , which has strong [OII] $\lambda 3726, 3729 \text{ \AA}$ , H $\beta$ , [OIII] $\lambda 4959, 5007 \text{ \AA}$ , and H $\alpha$  emission in the MUSE-Wide DR1 data cube. Projected on the sky, the object appears close to the foreground M star  $ID_{\text{Guo}} = 9777$ . The two objects are separated in the HST data (top left panel), but are marginally resolved in the MUSE data cube, as illustrated in the white light and [OII] narrowband postage stamps in the top right panels given the MUSE PSF size of just below  $1''.0$ . The [OII] narrowband has a rest-frame width of  $1000 \text{ km s}^{-1}$ . The amount of contamination (i.e., blending between the spectra of the two objects) strongly depends on the spectral extraction method. The bottom left panels show aperture spectra for the two objects extracted using an aperture size of  $2 \times r_{\text{major}} = 1''.02$ , where  $r_{\text{major}}$  is the isophotal major axis of  $ID_{\text{Guo}} = 9640$  provided in the photometric Guo et al. (2013) source catalog. As we show in Sect. 4.5,  $2 \times r_{\text{major}}$  provides a good compromise between recovered flux and optimal S/N for aperture extractions when you cannot obtain both. As the FWHM of the MUSE PSF is  $\sim 1''.0$  this aperture size also recovers the vast majority of the light from the star. The characteristic ‘‘wavy’’ continuum of the M star (red spectrum) is clearly seen imprinted on the extracted galaxy spectrum (black spectrum). In a similar way, the stellar spectrum has been contaminated by the galaxy emission lines. Likewise, a MUSE PSF-weighted spectral extraction results in spectra with considerable contamination. The spectra extracted with TDOSE shown in the bottom right panels are based on a single-source model for each of the two objects (generated with GALFIT and shown in the second panel,



**Fig. 5.** Comparison of the peak [OII] flux (*top panel*) and S/N (*bottom panel*) from the spectra of 153 MUSE-Wide DR1 [OII] emitters extracted with TDOSE based on multi-component Sérsic models of the HST F814W imaging generated with GALFIT (*x*-axes), and the MUSE-Wide DR1 TDOSE spectra based on single Gaussian models (*y*-axes). In both panels the dashed line indicates the one-to-one relation, and the red line is the best linear fit to the flux and S/N measurements with the  $3\sigma$  uncertainty on the fit indicated by the shaded region.

top), and are significantly cleaner than the aperture (and PSF weighted) spectra. Due to the efficient de-blending by TDOSE the galaxy now has a flat low-level continuum without obvious imprints from the neighboring star, and the stellar spectrum has been cleaned for the galaxy emission lines. The de-blending and spectral extraction by TDOSE was done without any loss in flux or S/N compared to the  $2 \times r_{\text{major}}$  aperture extractions.





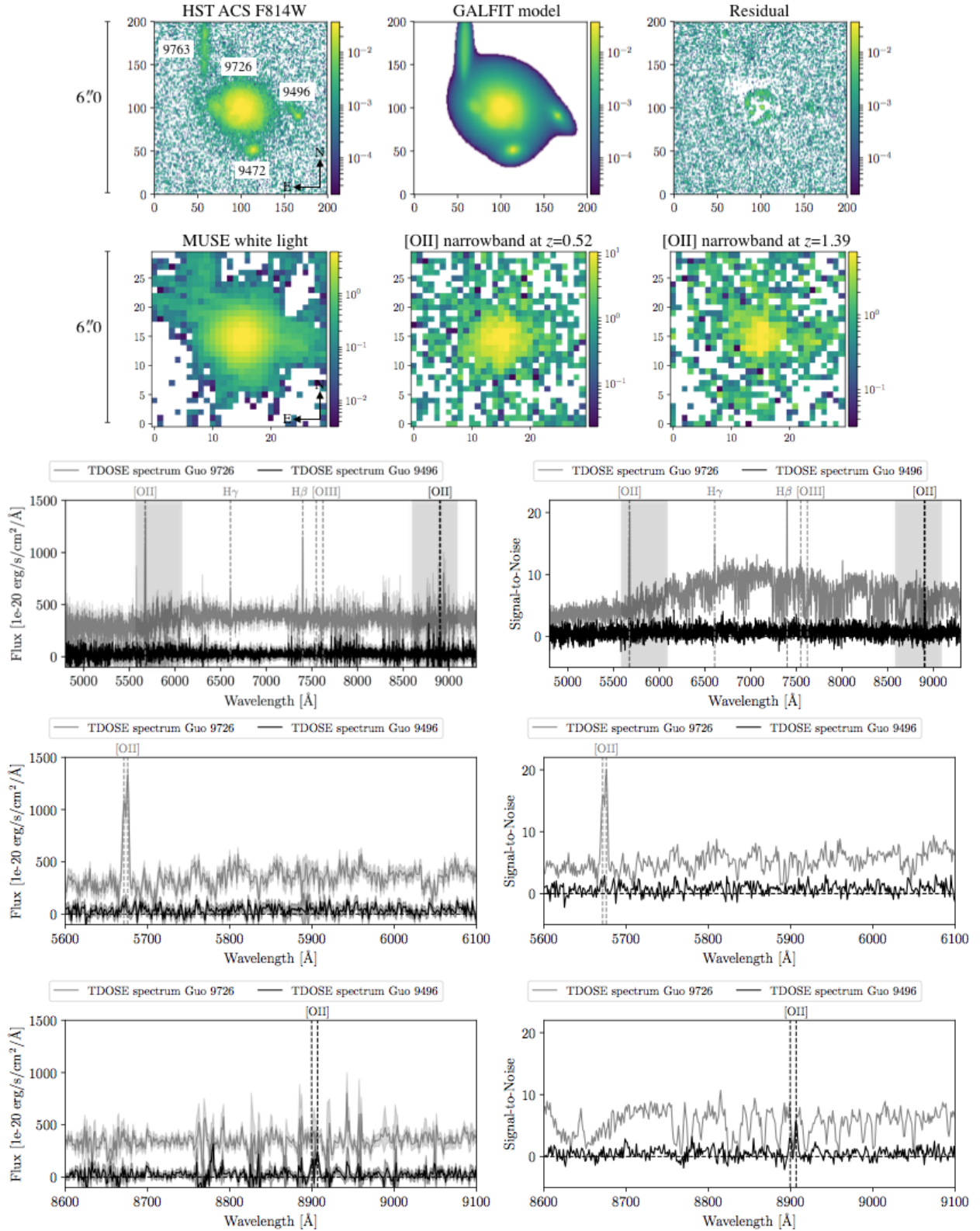
**Fig. 6.** Illustration of extraction of galaxy (object ID<sub>Guo</sub> = 9640, ID<sub>MUSE-Wide</sub> = 102009072) and star (object ID<sub>Guo</sub> = 9777) spectrum while de-blending with TDOSE. *Top panels:* 4'0 × 4'0 postage stamps of the HST F814W image, the two-component GALFIT model, the MUSE white light image, and a narrowband image around the [OII] emission of the galaxy at  $z = 0.34$  from the MUSE data cube. *Bottom left panels:* aperture spectra extracted for the two objects, indicated by the circles in the top left panel. The aperture sizes were set to  $2 \times r_{\text{major}} = 1''.02$  of the galaxy ID<sub>Guo</sub> = 9640, providing a good compromise between flux and S/N (see Fig. 9). The three panels show the full spectrum and zoom-ins on the H $\beta$  and [OIII] $\lambda$ 4959,5007 Å emission lines region (gray box in the top panel). *Bottom right panels:* identical to the bottom left panels, except that now the spectra shown were extracted with TDOSE based on the two-component GALFIT model (*top second panel*).

Figure 7 shows a second example of de-blending with TDOSE from solving the equations presented in Sect. 3.2. The complex of four individual Guo objects (top left panel) were modeled using multiple sources with GALFIT (top central and right panel). This reference image model was used to extract spectra from the MUSE-Wide data cube where the flux is blended, as illustrated by the MUSE white light image that is also shown. The bottom panels show the spectra for ID<sub>Guo</sub> = 9726 (ID<sub>MUSE-Wide</sub> = 125017033, gray spectra) and ID<sub>Guo</sub> = 9496 (ID<sub>MUSE-Wide</sub> = 125068147, black spectra). The bright continuum of ID<sub>Guo</sub> = 9726 has been clearly de-blended from the

fainter line emitter ID<sub>Guo</sub> = 9496. The cross-contamination from the [OII] emission in the two objects which are at  $z = 0.52$  and  $z = 1.39$ , respectively, has also been removed with the de-blending by TDOSE.

Another case of de-blending happens when objects in the reference (HST) imaging are only barely resolved, but the IFS data show clear spectral features of an unresolved superposition of multiple sources. In such cases the centroid of the spectral features in the IFS data cube can be used to de-blend and assign flux to independent sources of one (or more) objects in the FoV. Such de-blending can determine, for instance, the





**Fig. 7.** Illustration of extraction of multiple spectra while de-blending with TDOSE. *Top panels* (from top left to bottom right):  $6''.0 \times 6''.0$  postage stamps of the HST F814W image, the multi-component GALFIT model and residual, the MUSE white light image, and two narrowband images around [OII] of widths  $1000 \text{ km s}^{-1}$  in rest-frame positioned at redshifts 0.52 and 1.39, corresponding to the redshifts of object  $\text{ID}_{\text{Guo}} = 9726$  ( $\text{ID}_{\text{MUSE-Wide}} = 125017033$ ) and  $\text{ID}_{\text{Guo}} = 9496$  ( $\text{ID}_{\text{MUSE-Wide}} = 125068147$ ), respectively. *Bottom panels:* TDOSE spectra extracted for object  $\text{ID}_{\text{Guo}} = 9726$  (gray) and  $\text{ID}_{\text{Guo}} = 9496$  (black) based on the multi-component GALFIT model (*top central panel*). *First two panels:* full spectra and the position of prominent emission lines at the redshift of the two objects. The gray shaded regions show the location of the flux and S/N zoom-ins shown in the *bottom two panels*. Due to the efficient de-blending by TDOSE, the continuum and emission line contamination in the spectra is low, and reliable flux estimates can be obtained based on these extractions. Figure 8 shows a comparison between aperture spectra and the model-based TDOSE spectrum of  $\text{ID}_{\text{Guo}} = 9726$ .

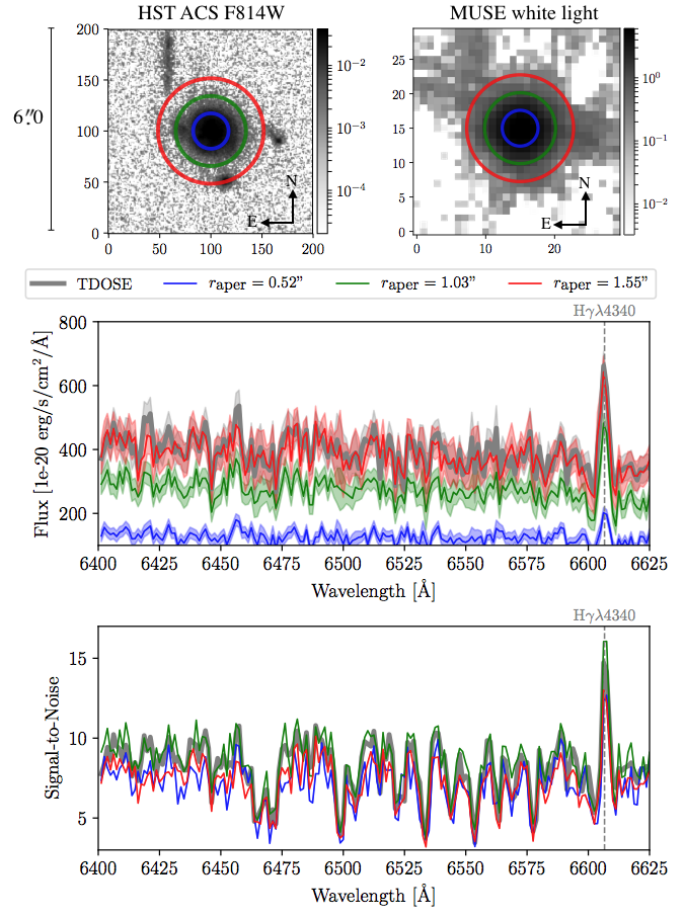
origin of emission lines and other prominent spectral features, and through them can reliably determine redshifts of objects that are only marginally resolved at the resolution of the reference imaging used. In such cases, photometric catalogs often only assign a single ID to the unresolved objects. To avoid too aggressive photometric de-blending this is likely also the correct thing to do, when the de-blending based on the IFS data is unavailable. However, if physical parameters like stellar mass, equivalent width, or SFR are estimated from fitting templates to the photometry of the combined flux the results would be biased, given that the combined flux from the unresolved objects is assumed to be from a single source with the spectral features (e.g., emission lines) from the IFS data cube.

We note that the simultaneous modeling and de-bending of objects with TDOSE introduces nonzero covariances between the different source models. These covariances can become significant for especially unresolved sources. Nevertheless, the above examples illustrate the importance of careful spectral extraction, using the best possible reference imaging, to avoid incorrect identification and misinterpretations, especially at high redshift where blending is very prominent in most IFS data. If the use of ancillary data had deliberately been avoided in the examples shown in this section, the scientific analysis of the extracted spectra and the corresponding broadband photometry would have been biased.

#### 4.5. Comparison of model-based and aperture-based extractions

As described in Sect. 2 an optimal spectral extraction attempts to recover the intrinsic flux of the considered object as accurately as possible, while still providing a high S/N by limiting the number of voxels with limited information included in the extraction. On the other hand, standard aperture extractions implicitly requires a choice between accurate flux measurement or high S/N. An aperture extraction cannot provide both. This is illustrated in Fig. 8, where we show the emission line free region of the continuum blueward of the  $H\gamma$  line for the object  $ID_{\text{Guo}} = 9496$  also shown in Fig. 7. The three apertures marked by the colored circles in the top panels, have a size of  $1 \times r_{\text{major}} = 0''.52$  (blue),  $2 \times r_{\text{major}} = 1''.03$  (green), and  $3 \times r_{\text{major}} = 1''.55$  (red). Here  $r_{\text{major}}$  corresponds to the isophotal major axis (`a_image`) of  $ID_{\text{Guo}} = 9496$  as measured by SExtractor and provided in the Guo et al. (2013) catalog. The bottom panels show that, as expected, the largest aperture captures the largest amount of flux at a cost of a lower overall S/N (bottom panel). A good compromise between flux recovery and S/N appears to be the extraction using an aperture with a radius of  $2 \times r_{\text{major}}$  (green circle and spectra). The thick gray curve shows the TDOSE spectrum of  $ID_{\text{Guo}} = 9496$  also presented in Fig. 8. The TDOSE spectrum recovers the flux at the level of the  $3 \times r_{\text{major}}$  aperture spectrum (red, central panel), but simultaneously provides a S/N per pixel comparable to the  $2 \times r_{\text{major}}$  aperture spectrum (green, bottom panel).

To estimate the efficiency of the TDOSE extractions compared to aperture extractions for a larger sample of objects, we return to the  $\sim 150$  MUSE-Wide [OII] emitters considered in Sect. 4.3. We extracted aperture spectra for all objects again using aperture radii of  $1 \times r_{\text{major}}$  (blue),  $2 \times r_{\text{major}}$  (green), and  $3 \times r_{\text{major}}$  (red). The  $r_{\text{major}}$  value for each object was taken from the Skelton et al. (2014) catalog. In Fig. 9 we compare the peak flux (top) and S/N (bottom) of the [OII] emission line from the TDOSE extractions based on GALFIT multi-component models ( $x$ -axis) and the aperture spectra ( $y$ -axes). As expected, the largest apertures on average recovers the most flux, whereas a



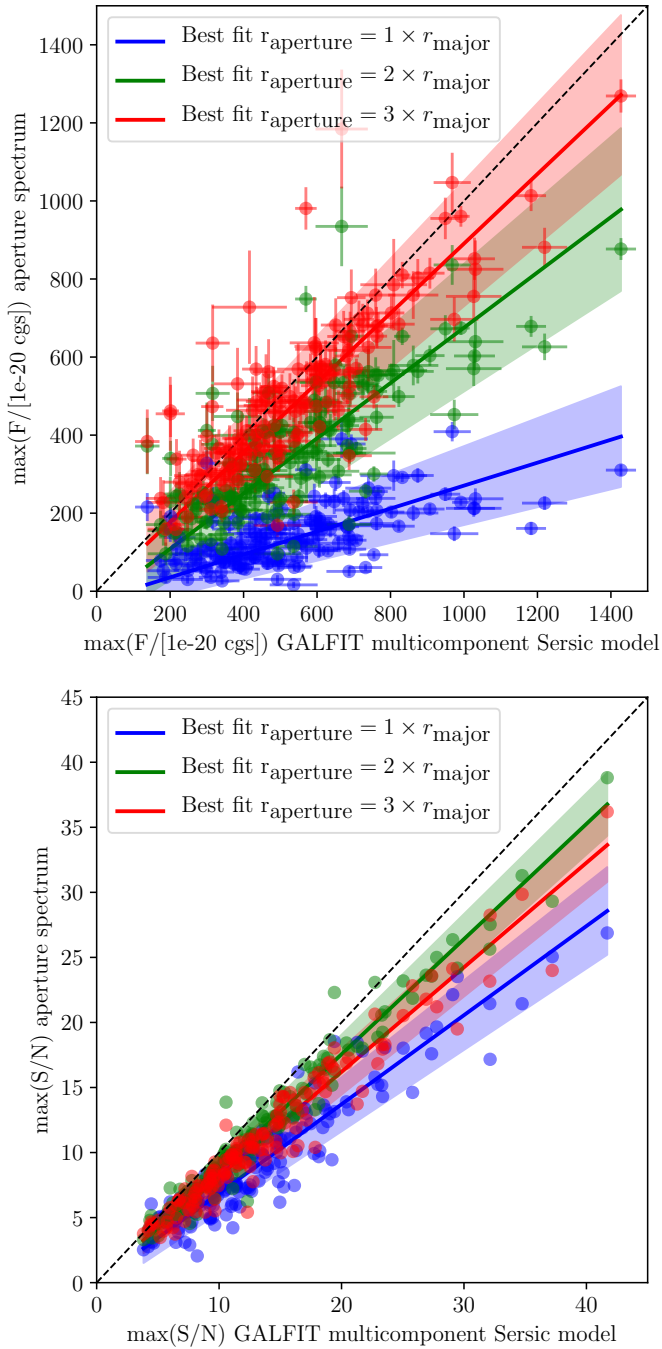
**Fig. 8.** Comparison of three different aperture spectra (blue, green, and red) with the TDOSE spectrum of  $ID_{\text{Guo}} = 9726$  shown in Fig. 7 (thick gray) extracted based on a multi-component object model. The radii of the apertures are  $1 \times r_{\text{major}} = 0''.52$  (blue),  $2 \times r_{\text{major}} = 1''.03$  (green), and  $3 \times r_{\text{major}} = 1''.55$  (red). *Top panels:*  $6''.0 \times 6''.0$  postage stamps of the HST F814W and the MUSE white light images with the aperture sizes marked by the colored circles. *Bottom panels:* zoom-in on the emission line free continuum blueward of the  $H\gamma$  line. The largest aperture (red) recovers the most flux, but suffers from a lower S/N. On the other hand, the median-aperture extraction (green) provides the highest S/N, but suffers from a lower recovered flux. The spectrum extracted with TDOSE (thick gray) recovers the same flux as the large-aperture spectrum, but also has a S/N similar to the median-aperture extraction. See Fig. 9 for similar effects for a sample of 153 fainter [OII] emitters.

more modest aperture of just  $2 \times r_{\text{major}}$  achieves the highest S/N on average in agreement with the extractions for  $ID_{\text{Guo}} = 9496$  in Fig. 8. As was the case for  $ID_{\text{Guo}} = 9496$  the model-based TDOSE extractions are capable of providing a high S/N while still delivering a reliable estimate of the peak [OII] flux for the MUSE-Wide [OII] emitter sample. The median increase in flux (S/N) of the TDOSE spectra is even 9% (14%) when compared to the  $3 \times r_{\text{major}}$  ( $2 \times r_{\text{major}}$ ) aperture extractions.

Hence, the model-based spectral extractions from TDOSE provide optimal extractions bringing the “best of two worlds” by optimizing the S/N, while still recovering a large fraction of the emitted flux.

#### 4.6. Generating contamination-free 2D maps

As described in Sect. 3.4 the 3D source models produced by TDOSE are useful for removing unwanted contaminating flux



**Fig. 9.** Comparison of the peak [OII] flux (*top panel*) and S/N (*bottom panel*) from the spectra of the 153 MUSE-Wide DR1 [OII] emitters also analyzed in Fig. 5. The spectra were extracted with TDOSE based on multi-component Sérsic models of the HST F814W reference imaging generated with GALFIT (*x*-axes), and apertures with radii equal to 1, 2, and 3 times the major axis of each object provided in the photometric catalogs by Skelton et al. (2014) plotted in blue, green and red, respectively. In both panels the dashed line shows the one-to-one relation, and the solid lines indicate the best linear fits to the individual flux and S/N measurements with the  $3\sigma$  uncertainty on each fit indicated by the shaded regions. In all cases the average flux and S/N is higher for the model-based TDOSE extractions, even though the median flux increase is only  $\sim 10\%$  for the largest flux aperture.

to generate new corrected data cubes. This process is illustrated on the right-hand side of the TDOSE version 3.0 flowchart in Fig. 2. As mentioned, these modifications of the intrinsic IFS

data cubes can improve the S/N, extent, and reliability of spatial maps derived from the data cubes.

To illustrate this use of the TDOSE output we focus on the objects shown in Fig. 10 from the galaxy group COSMOS-Gr32 ( $z = 0.73$ ) presented by Knobel et al. (2012) and Boselli et al. (in prep.). These objects were observed for 5.25 h with MUSE as part of the GTO program focusing on the effect of environment on galaxy evolution processes (PI: T. Contini). At this depth the S/N is high enough to derive kinematic maps for the central galaxy in the HST ACS F814W image shown in the top left panel of Fig. 10. In the  $7''0 \times 7''0$  postage stamps four sources are contaminating the signal obtainable from the central object. By modeling all sources in the FoV with GALFIT (top central panel of Fig. 10) and generating the 3D source models with TDOSE, the flux from the contaminants can be removed from the original MUSE data cube, as illustrated by the original and contamination-corrected MUSE white light images shown in the central panels of Fig. 10. We derived the stellar velocity maps for both the original and the contamination-corrected MUSE data cubes using the pPXF method (Cappellari & Emsellem 2004). To ensure reliable fits to the data, we used Voronoi binning (Cappellari & Copin 2003) requiring that each bin had  $S/N > 10$ . The voxels in each Voronoi bin were collapsed into a single 1D spectrum and used to estimate the velocities. The resulting velocity maps are shown in the bottom panels of Fig. 10. Without correcting the data cube for contaminating flux the velocity map (bottom left panel) indicates rotation of the central galaxy around an approximate E-W axis, even though such a conclusion is naturally uncertain given the contamination. Instead, as the velocity field is not that of the central galaxy, but rather that of the system as a whole, we likely see the relative motions of the contaminating sources, which are all at the group redshift of  $z = 0.73$ .

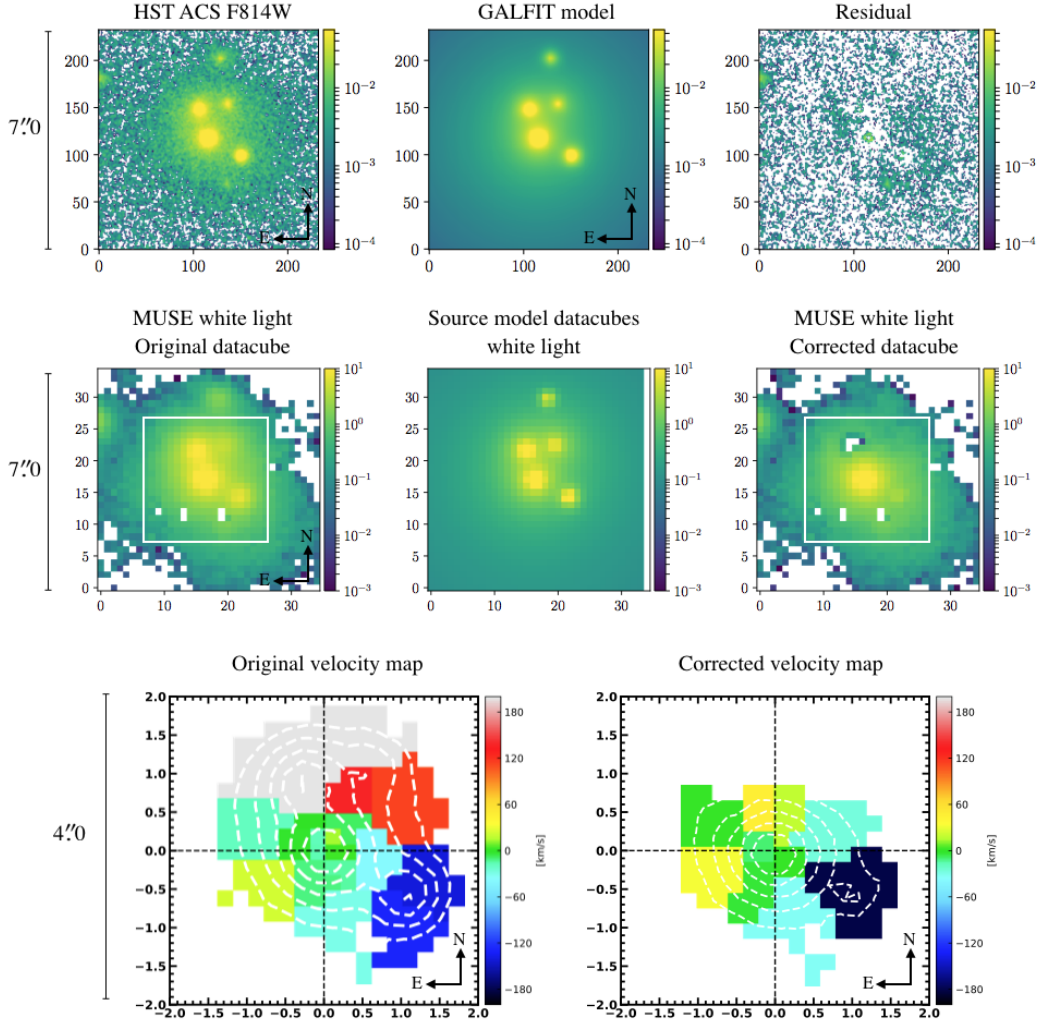
After removing the contaminating flux, the velocity map is significantly cleaner. The dark Voronoi bin to the southwest still shows signs of contamination residuals. These residuals potentially result from kinematic signatures in this galaxy that are poorly represented by the single-Gaussian model (see Sect. 4.7). Nevertheless, from the cleaned velocity map (bottom right panel of Fig. 10) it becomes clear that the central object likely has no significant rotation, and if any, the rotation appears to be around a SE-NW axis and not the E-W axis implied by the original map. Hence, using the TDOSE source models to correct the original IFS data cube removing the contaminants provides an alternative, but more reliable assessment of the 2D velocity map.

Improvements similar to what the velocity maps in Fig. 10 show are not restricted to kinematic maps. Source models like the ones produced by TDOSE can be used to modify and correct maps of metallicity, star formation rate, electron density etc. generated from IFS data cubes where 3D flux contamination has been corrected for.

#### 4.7. Extractions for objects with spatially varying flux

Even though spectral extractions based on morphological reference image models are generally flexible and versatile in their approach, they do offer some limitations. As mentioned, spectral extractions with TDOSE assume that the extent of the features to be extracted are closely represented by the input model. As explained above, this causes potential biases in representing extended nebular emission if the model reflects the stellar continuum of the galaxy. Another limitation of the model-based extraction approach occurs when the morphological model represents the source well, but there are spatial variations in the flux



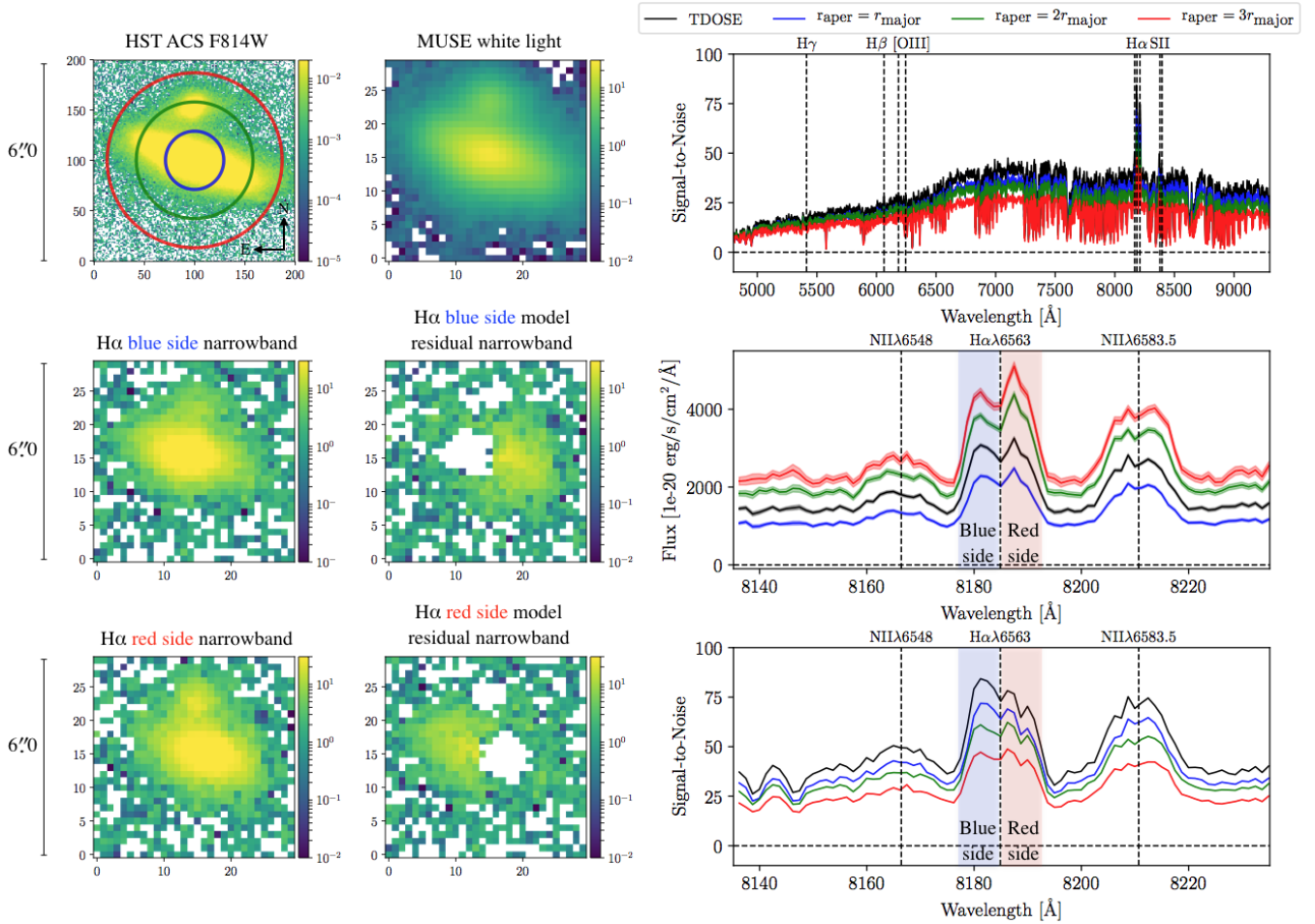


**Fig. 10.** Illustration of the improved velocity map obtained after removing the 3D source models of the contaminants generated by TDOSE from the original IFS data cube. *Top panels:*  $7'' \times 7''$  postage stamp of the HST F814W image, the HST multi-component GALFIT model, and the corresponding residual image. *Central panels:*  $7'' \times 7''$  postage stamp of the MUSE white light image, the HST model in the MUSE coordinate system, and the MUSE white light image after removing all but the central 3D source model generated by TDOSE. *Bottom left panel:* velocity map generated from the intrinsic MUSE data cube (see Sect. 4.6 for details). *Bottom right panel:* velocity map constructed the same way, but now based on the MUSE cube after the contamination has been removed using the TDOSE source models and indicates that the central galaxy has no significant rotation. The extent of the velocity maps, as well as their extent, are different as we are left with the flux belonging only to the galaxy of interest after the contaminating flux was removed (*bottom right panel*).

distribution as a function of wavelength. An example of such a spatially varying flux distribution could be strong emission line regions within an otherwise dormant galaxy. However, such regions are likely identifiable in the reference imaging and can be accounted for based on the reference image model (see, e.g., Fig. 4). A more challenging example of wavelength dependent spatial variations is the presence of velocity shifts of spectral features due to rotation. These effects cannot be identified in broadband imaging and can therefore not be corrected for using only reference image modeling. Figure 11 shows the HST F814W image and the MUSE white light image of  $ID_{\text{Guo}} = 16009$  ( $ID_{\text{MUSE-wide}} = 136002114$ ) in the top left panels. This object is rotating around an approximate N-S axis which introduces significant shifts of the  $H\alpha$  emission. In the bottom left panels  $6\text{ \AA}$  wide narrowband filters blueward and redward of the  $H\alpha$  line show the shift of the emission centroid caused by the galaxy's rotation. In the right panels three aperture spectra (red, green, blue) and a TDOSE spectrum (black) extracted based on

a single Gaussian model of the HST reference image are shown for comparison. The layers included on the blue side and red side  $H\alpha$  narrowband images are indicated on the spectra in the bottom right panels. For  $ID_{\text{Guo}} = 16009$  the TDOSE extraction is biased and is unable to correctly recover the line flux due to the galaxy's rotation. This is seen by considering the residual narrowbands where the optimized flux model has been subtracted from the blue and red side narrowbands in the bottom central panels. The single-component HST source model is clearly unable to represent the spatially varying flux distribution of the  $ID_{\text{Guo}} = 16009$  and causes under- and oversubtraction of the  $H\alpha$  line in the IFS flux around the N-S rotation axis of the galaxy. To improve the TDOSE model-based extraction of this object, it is not enough to only rely on the information from the reference imaging, where the rotational information is unavailable. Instead, building a multi-component source model based on the narrowband white light images would provide a much more reliable spectral extraction of  $ID_{\text{Guo}} = 16009$ . A similar





**Fig. 11.** Example of spectral extraction of spectra from an object with significant wavelength dependent spatial flux variation due to rotation. These objects provide limitations to the capabilities of extractions with TDOSE based on reference image source models. *Top left panels:*  $6'0 \times 6'0$  postage stamp of the HST ACS F814W image and the MUSE white light image of ID<sub>Guo</sub> = 16009 (ID<sub>MUSE-wide</sub> = 136002114). Below these, narrowband images of the blue and red sides of the H $\alpha$  emission in the MUSE data cubes (*left*) and the residual cube after the removal of the scaled source model (*right*) are shown. *Right panels:* aperture spectra, corresponding to 1 (blue), 2 (green), and 3 (red)  $\times r_{\text{major}}$ , together with the TDOSE spectrum based on a single-source Gaussian HST model (black). *Bottom right panels:* narrow zoom-in around the H $\alpha$ +NII lines, with the ranges included in the narrowband images in the *bottom left panels* indicated by the shaded regions.

approach is needed to correctly extract spectra with TDOSE of objects with significant wavelength-dependent spatial variations in the overall flux distribution that are not identifiable in the reference imaging. Hence, for spectral extractions with prominent spatial shifts, using multi-component models (see Fig. 4) based on IFS data cube information is recommendable.

## 5. Conclusion

The data volume from sensitive wide-field IFSs has grown considerably over the last few decades, and with future missions and planned instruments this appears to continue. Therefore, precise and efficient tools to handle these IFS data cubes, to efficiently de-blend flux in crowded exposures, and to automatically extract 1D spectra of large samples of objects for further analysis are needed now, more than ever. In an attempt to satisfy this demand, we have presented a flexible Python tool for Three-Dimensional Optimal Spectral Extraction (TDOSE).

The spectral extraction performed by TDOSE is based on 2D modeling of the object morphology based on (preferably high resolution) reference imaging. By default, all objects in the FoV are modeled with multivariate Gaussian sources, and are then

simultaneously scaled to minimize the differences between the model components and each of the wavelength layers in the IFS data cubes from which the spectra are extracted. This makes TDOSE fully analytical, and hence very efficient for large samples of objects. Alternatively, any numerical 2D models (e.g., morphological models from GALFIT or MGE models of individual objects) can be provided to TDOSE as the basis for the spectral extraction.

Using MUSE data cubes to illustrate the capability and limitations of TDOSE we show the following:

- The model-based TDOSE extractions are capable of recovering flux from spectra that are only partially covered by the IFS detector despite a minimal loss in S/N caused by the fewer voxels available in the data cube (Sect. 4.2).
- Representing objects by multi-source models improves the reliability of the extracted spectra. For a sample of  $\sim 150$  [OII] emitters the median increase in flux recovered basing the TDOSE extractions on object models composed of multiple sources as opposed to just one multivariate Gaussian object model is on the order of 5%. However, in extreme cases, this flux increase can be as high as 50% for certain wavelength ranges (Sect. 4.3).

- The simultaneous scaling of all sources in the FoV allows TDOSE to perform efficient and precise de-blending of sources in 3D. In this way, TDOSE can be used to remove contaminating sources (satellite, foreground, or background) when extracting 1D spectra (Sect. 4.4).
- Spectral extractions performed with TDOSE precisely recover object fluxes comparable to aperture extractions using larger aperture sizes. However, the TDOSE spectra simultaneously provide a high S/N, which in the case of aperture extractions is only possible for smaller apertures. Hence, TDOSE brings the best of two worlds. The TDOSE extractions for the  $\sim 150$  [OII] emitters improve the peak [OII] flux and S/N by 9% and 14%, respectively, compared to aperture extractions (Sect. 4.5).
- The 3D source models produced by TDOSE are also capable of correcting 2D maps generated from the IFS data cubes, for instance emission line, metallicity, or kinematic maps (Sect. 4.6).
- The main limitation of TDOSE is that the precision of the spectral extraction is only as good as the model. For instance, spectral extractions of extended emission based on compact continuum models will naturally be biased. In cases where the extent is well represented by the reference image model, but significant wavelength dependent spatial flux variations are present in the IFS data cube, spectral extractions with TDOSE will also be biased. However, the flexibility of the multiple source component modeling approach offers several ways to mitigate and account for these biases (Sect. 4.7).

Hence, TDOSE offers a flexible, efficient, and (by default) fully analytic tool to extract spectra and to account for undesirable flux contamination via efficient de-blending in three dimensions from IFS data cubes while optimizing the S/N of the extracted 1D spectra.

*Acknowledgements.* We would like to thank the MUSE consortium for constructive feedback during the development and testing of TDOSE. This work has been supported by the BMBF grant 05A14BAC and we acknowledge support from the Competitive Fund of the Leibniz Association through grant SAW-2015-AIP-2. This research made use of the following programs and open-source packages for Python and we are thankful to the developers: DS9 (Joye & Mandel 2003), Astropy (Astropy Collaboration 2013, 2018), APLpy (Robitaille & Bressert 2012), iPython (Pérez & Granger 2007), numpy (van der Walt et al. 2011), matplotlib (Hunter 2007), SciPy (Jones et al. 2001), and PyFITS, a product of the Space Telescope Science Institute, which is operated by AURA for NASA.

## References

- Allen, J. T., Croom, S. M., Konstantopoulos, I. S., et al. 2015, *MNRAS*, **446**, 1567
- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, **558**, A33
- Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, **156**, 123
- Bacon, R., Accardo, M., Adjali, L., et al. 2010, in *Centre de Recherche Astrophysique de Lyon, CNRS, Univ. Claude-Bernard Lyon I, France*, eds. I. S. McLean, S. K. Ramsay, H. Takami, et al. (SPIE), *Proc. SPIE*, **773508**
- Bacon, R., Vernet, J., Borisova, E., et al. 2014, *The Messenger*, **157**, 13
- Bacon, R., Brinchmann, J., Richard, J., et al. 2015, *A&A*, **575**, A75
- Bacon, R., Piqueras, L., Conseil, S., Richard, J., & Shepherd, M. 2016, *Astrophysics Source Code Library* [record ascl:1611.0093]
- Bacon, R., Conseil, S., Mary, D., et al. 2017, *A&A*, **608**, A1
- Bendinelli, O., & Parmeggiani, G. 1995, *AJ*, **109**, 572
- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bundy, K., Bershad, M. A., Law, D. R., et al. 2015, *ApJ*, **798**, 7
- Cappellari, M. 2002, *MNRAS*, **333**, 400
- 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
- Carton, D., Brinchmann, J., Contini, T., et al. 2018, *MNRAS*, **478**, 4293
- Contini, T., Epinat, B., Bouché, N., et al. 2016, *A&A*, **591**, A49
- de Zeeuw, P. T., Bureau, M., Emsellem, E., et al. 2002, *MNRAS*, **329**, 513
- Drake, A. B., Garel, T., Wisotzki, L., et al. 2017a, *A&A*, **608**, A6
- Drake, A. B., Guiderdoni, B., Blaizot, J., et al. 2017b, *MNRAS*, **471**, 267
- Emsellem, E., Monnet, G., & Bacon, R. 1994a, *A&A*, **285**, 723
- Emsellem, E., Monnet, G., Bacon, R., & Nieto, J. L. 1994b, *A&A*, **285**, 739
- Feltre, A., Bacon, R., Tresse, L., et al. 2018, *A&A*, **617**, A62
- Finley, H., Bouché, N., Contini, T., et al. 2017, *A&A*, **605**, A118
- Förster Schreiber, N. M., Genzel, R., Bouché, N., et al. 2009, *ApJ*, **706**, 1364
- García-Benito, R., Zibetti, S., Sánchez, S. F., et al. 2015, *A&A*, **576**, A135
- Green, A. W., Croom, S. M., Scott, N., et al. 2018, *MNRAS*, **475**, 716
- Griffiths, A., & Conselice, C. J. 2018, *ApJ*, **869**, 68
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35
- Guerou, A., Krajnović, D., Epinat, B., et al. 2017, *A&A*, **608**, A5
- Guo, Y., Ferguson, H. C., Gialalisco, M., et al. 2013, *ApJS*, **207**, 24
- Herenz, E. C., & Wisotzki, L. 2017, *A&A*, **602**, A111
- Herenz, E. C., Urrutia, T., Wisotzki, L., et al. 2017, *A&A*, **606**, A12
- Herenz, E. C., Wisotzki, L., Saust, R., et al. 2019, *A&A*, **621**, A107
- Hill, G. J., & HETDEX Consortium 2016, *Multi-Object Spectroscopy in the Next Decade: Big Questions*, **507**, 393
- Hill, G. J., Tuttle, S. E., Lee, H., et al. 2012 in *Ground-based and Airborne Instrumentation for Astronomy IV*, eds. I. S. McLean, S. K. Ramsay, H. Takami, et al., *Proc. SPIE* (Austin, US: McDonald Observatory, The Univ. of Texas)
- Hogg, D. W., & Lang, D. 2013, *PASP*, **125**, 719
- Horne, K. 1986, *ASP*, **98**, 609
- Hunter, J. D. 2007, *Comput. Sci. Eng.*
- Husmann, B., Jahnke, K., Sánchez, S. F., et al. 2013, *A&A*, **549**, A87
- Husser, T.-O., Kamann, S., Dreizler, S., et al. 2016, *A&A*, **588**, A148
- Inami, H., Bacon, R., Brinchmann, J., et al. 2017, *A&A*, **608**, A2
- Jones, E., Oliphant, T., & Peterson, P. 2001, *SciPy: Open Source Scientific Tools for Python*
- Joye, W. A., & Mandel, E. 2003, *ASP Conf. Ser.*, **295**, 489
- Kamann, S. 2018, *Astrophysics Source Code Library* [record ascl:1805.021]
- Kamann, S., Wisotzki, L., & Roth, M. M. 2013, *A&A*, **549**, A71
- Kamann, S., Husser, T.-O., Brinchmann, J., et al. 2016, *A&A*, **588**, A149
- Kamann, S., Husser, T.-O., Dreizler, S., et al. 2018, *MNRAS*, **473**, 5591
- Knobel, C., Lilly, S. J., Iovino, A., et al. 2012, *ApJ*, **753**, 121
- Kochanek, C. S., Falco, E. E., Impey, C. D., et al. 2000, *ApJ*, **535**, 692
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36
- Krajnović, D., Emsellem, E., den Brok, M., et al. 2018, *MNRAS*, **477**, 5327
- Lagattuta, D. J., Richard, J., Clement, B., et al. 2017, *MNRAS*, **469**, 3946
- Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, *A&A*, **608**, A8
- Mahler, G., Richard, J., Clément, B., et al. 2018, *MNRAS*, **473**, 663
- Maseda, M. V., Bacon, R., Franx, M., et al. 2018, *ApJ*, **865**, L1
- Meillier, C., Chatelain, F., Michel, O., et al. 2016, *A&A*, **588**, A140
- Monnet, G., Bacon, R., & Emsellem, E. 1992, *A&A*, **253**, 366
- Nelson, E. J., van Dokkum, P. G., Momcheva, I., et al. 2013, *ApJ*, **763**, L16
- Nelson, E. J., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2016a, *ApJ*, **828**, 27
- Nelson, E. J., van Dokkum, P. G., Momcheva, I. G., et al. 2016b, *ApJ*, **817**, L9
- Paalvast, M., Verhamme, A., Straka, L. A., et al. 2018, *A&A*, **618**, A40
- Patricio, V., Richard, J., Carton, D., et al. 2018, *MNRAS*, **477**, 18
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H. W. 2002, *AJ*, **124**, 266
- Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H. W. 2010, *AJ*, **139**, 2097
- Pérez, F., & Granger, B. E. 2007, *Comput. Sci. Eng.*, **9**, 21
- Poggianti, B. M., Moretti, A., Gullieuszk, M., et al. 2017, *ApJ*, **844**, 48
- Robertson, J. G. 1986, *ASP*, **98**, 1220
- Robitaille, T., & Bressert, E. 2012, *Astrophysics Source Code Library* [record ascl:1208.017]
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, *A&A*, **538**, A8
- Sánchez, S. F., García-Benito, R., Zibetti, S., et al. 2016, *A&A*, **594**, A36
- Schmidt, K. 2019, <https://doi.org/10.5281/zenodo.3243945>
- Schmidt, K. B., Rix, H. W., da Cunha, E., et al. 2013, *MNRAS*, **432**, 285
- Scott, N., Cappellari, M., Davies, R. L., et al. 2013, *MNRAS*, **432**, 1894
- Scott, N., van de Sande, J., Croom, S. M., et al. 2018, *MNRAS*, **481**, 2299
- Sérsic, J. L. 1963, *Boletín de la Asociación Argentina de Astronomía*, **6**, 41
- Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, *ApJS*, **214**, 24
- Swinbank, A. M., Harrison, C. M., Trayford, J., et al. 2017, *MNRAS*, **467**, 3140
- Urrutia, T., Wisotzki, L., Kerutt, J., et al. 2019, *A&A*, **624**, A141
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Comput. Sci. Eng.*, **13**, 22
- Vulcani, B., Treu, T., Schmidt, K. B., et al. 2015, *ApJ*, **814**, 161
- Vulcani, B., Treu, T., Schmidt, K. B., et al. 2016, *ApJ*, **833**, 178
- Vulcani, B., Treu, T., Nipoti, C., et al. 2017, *ApJ*, **837**, 126
- Wang, X., Jones, T. A., Treu, T., et al. 2017, *ApJ*, **837**, 89
- Wisnioski, E., Förster Schreiber, N. M., Wuyts, S., et al. 2015, *ApJ*, **799**, 209
- Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, *A&A*, **587**, A98
- Wisotzki, L., Bacon, R., Brinchmann, J., et al. 2018, *Nature*, **562**, 229

## Appendix A: Running TDOSE

TDOSE is run through a main setup file that contains the information for the extractions to perform. The TDOSE setup file contains pointers to the main inputs for TDOSE, namely the IFS flux data cube, the IFS variance data cube, the reference image of the FoV, and a source catalog defining the sources in the FoV. Among other things, the setup file also defines the extraction mode (aperture, Gaussian modeling, or pre-defined reference image model), the region of the FoV to consider, what objects to extract spectra for, the PSF model, and the location of initial guesses on the morphological parameters of each source to be modeled (if available). A completed setup file is parsed to the main wrapper of TDOSE to perform the spectral extractions. For extractions from multiple independent IFS data cubes, independent setup files can be generated, and TDOSE can be run in parallel on multiple cores to minimize computation time. In the following, details are provided on the TDOSE setup file, the main wrappers, and routines to use for standard runs of TDOSE and the corresponding outputs produced. Appendix B presents a few examples of calling sequences and commands for running TDOSE.

Spectral extractions with TDOSE are performed by running `tdose.perform_extraction(setupfile=setupfile)`. Here the setup file is the main interface for setting up and controlling the type of spectral extraction TDOSE will perform. We will describe the setup file in more detail in Appendix A.2 below. A range of keywords can be passed to `perform_extraction` to either skip individual steps or force certain additional features of the extraction. For details on these options, we refer the reader to the header of the function itself. A wrapper around `perform_extraction` that parallelizes the spectral extractions from multiple data cubes is provided in `tdose.perform_extractions_in_parallel()`.

A successful run of TDOSE produces a range of outputs (see Appendix A.4). The collection of 3D source models are particularly useful for modifications of the data cubes themselves. The intrinsic data cubes can be modified with the function `tdose_modify_cube.perform_modification()`, as described below.

In the following we present a short overview of the main routines and functions of TDOSE, which are called by the `perform_extraction` function; we then present an overview of the two setup files, and a short description of the outputs that can be generated. In Appendix B we provide a few minimal examples for performing standard tasks with TDOSE.

### A.1. Main TDOSE scripts

The file `tdose.py` contains the main wrapper for performing spectral extractions with TDOSE; it includes the main command to run, namely `tdose.perform_extraction()`. This function calls the main scripts available in the TDOSE repository to carry out the individual task of defining the region to consider, model the reference images, generate the source models, scale the models to recover the source fluxes at each wavelength range in the data cube, and finally generate the actual 1D spectra as shown in the TDOSE flowchart in Fig. 2. The scripts and functions handling these tasks are described in the following.

- `tdose.gen_cutouts()`. Before starting the modeling and spectral extraction, this function can be used to generate cutouts of the FoV of relevance from the IFS data cube and the reference imaging in case the full data cube is not modeled and used for the extraction. This limits the required memory needed and makes

the modeling and flux scaling more efficient. The size of each cutout is defined in the setup file (Appendix A.2).

- `tdose.model_refimage()`. This function calls `tdose_model_FoV.gen_fullmodel()`, which performs a multivariate Gaussian modeling of the morphology of all sources in the FoV. Alternatively, TDOSE loads a pre-defined reference image model provided by the user.

- `tdose.define_psf()`. Based on the input from the setup file, a 3D wavelength dependent model of the IFS PSF is defined. This PSF model (currently a symmetric 2D Gaussian to keep the PSF convolution analytic) is used to convert the morphological reference image source model to the 3D reference frame of the IFS data.

- `tdose.model_datacube()`. This function calls `tdose_model_cube.gen_fullmodel()`, which generates a full 3D model of each source based on the reference image model and the model of the IFS PSF. This is done by scaling each source component as described in Sect. 3.2. While scaling each of the individual sources in the model, a collection of the individual de-coupled source model cubes is generated by `tdose_model_cube.gen_source_model_cube()`. This main data structure is used for de-blending in both 1D and 2D, and for extracting the individual 1D spectra, and for modifying data cubes (Sects. 3.4 and 4.6).

- `tdose_extract_spectra.py`. This script contains the functions used for collapsing the source model cubes to extract the actual 1D spectra. The main function `extract_spectra()` extracts and stores the 1D spectra of individual objects produced by the TDOSE modeling in binary FITS tables. A single object spectrum can combine the flux models from multiple sources if desired, and any remaining sources are then treated as contaminants, and are accounted for in the spectral extraction. The extraction is controlled by a “source association dictionary”, which defines the sources to assign to each object. As shown in Appendix B.5, spectra can be generated independently of the main TDOSE extraction by manually generating a source association dictionary and passing it to the function `tdose_extract_spectra.extract_spectra()`.

- `tdose.plot_spectra()`. If requested in the setup file, the extracted spectra can be plotted in designated wavelength ranges for a quick way to assess the quality of the extractions. `tdose.plot_spectra()` is a wrapper calling `tdose_extract_spectra.plot_1Dspecs()`, which provides a flexible plotting and comparison tool that can be used to plot spectra as part of the post-processing. For example, all plotted spectra shown in this paper were generated with `tdose_extract_spectra.plot_1Dspecs()`. For a list of available keywords and plotting options we refer the reader to the header of the function itself. The generated outputs allows for further post-processing.

- `tdose_utilities.py`. Throughout TDOSE this script is called as it contains a collection of useful tools including, among other things, functions to generate template setup files, build 2D multivariate Gaussians, perform convolutions, prepare modeling with GALFIT, duplicate setup templates, generate overview plots, and extract sub-regions of images and data cubes.

- `tdose_build_mock_cube.py`. A small package of functions to generate mock data cubes is provided in this script, which is used for testing and trouble shooting.

- `tdose_modify_cube.py`. This is the last main script available in the TDOSE repository. As mentioned above, this is used to modify the input 3D data cube by subtracting the 3D source models generated by TDOSE to illuminate undesired flux contamination. This modification is controlled by



the setup file `tdose_setup_template_modify.txt` described in Appendix A.3 below. A template of this file can be generated with the function `tdose_utilities.generate_setup_template_modify()` and is provided with the TDOSE version 3.0 release.

## A.2. TDOSE setup file

The TDOSE setup file is the main way to interact and set up a spectral extraction with TDOSE and is an input for `tdose.perform_extraction()`. A template setup file (`tdose_setup_template.txt`) is provided in the TDOSE GitHub repository<sup>5</sup> (and the packed TDOSE version 3.0 release) and can be generated with the `tdose_utilities.generate_setup_template()` function. The setup file contains several sections with different information on both inputs and outputs. In the following, each of these are described. For further details, the template setup file itself provide comments on each of the input values.

- **DATA INPUT.** The data input section defines the input required for a minimal default TDOSE run. Here the location of the IFS data cube from which to extract the spectra and the corresponding variance cube, used to estimate the uncertainty on the extracted spectra, are provided. This section also indicates the reference image and the location and column names of the main source catalog. The source catalog defines which sources to model (assuming TDOSE is responsible for generating the reference image source model). It can be generated with standard source detection software, or be put together manually if more detailed modeling should be performed. A weight image of the reference image can also be provided. This is not used by TDOSE itself, but if only a subregion of the IFS data cube is considered, it can be useful to have the relevant FoV cut out of the reference weight image together with the reference image. Hence, any image can actually be provided here. For instance it can be useful to have TDOSE cut out from a “sigma image” if GALFIT modeling of the reference image cutout is performed.

- **OUTPUT DIRECTORIES.** This section defines the location of the outputs generated with TDOSE.

- **CUTOOUT SETUP.** Here the cutouts around each source in the source catalog to consider during the extraction and modeling are defined. It is recommended to always use a cutout for the extractions as this both limits the required memory available, and speeds up calculations. The cutout sizes can be defined to have the same size for all objects by providing the dimensions in RA and Dec given in arc seconds. Alternatively, the location of a file with source specific cutout sizes can be provide here.

- **SOURCE MODEL SETUP.** The source model setup defines what extraction method to use. The currently enabled modes (see also Fig. 2) are listed here:

- **gauss.** This is the default mode of TDOSE. Here, the reference image is modeled by positioning a multivariate Gaussian component at the location of each source in the cutout FoV. Using this reference image model makes the spectral extraction fully analytic. The optimization of these morphological models is performed using Scipy’s `curve_fit`<sup>6</sup> (Jones et al. 2001) function. The details of this extraction is provided in the **GAUSS MODEL SETUP** section of the setup file described below.

- **modelimg.** This option allows the user to provide an already existing model of the reference image, as detailed in the **MODEL IMAGE SETUP** section of the setup file, instead of

using the built-in Gaussian modeling of TDOSE. Such models can for instance be generated with GALFIT. This model will be treated numerically, as opposed to the Gaussian model generated by TDOSE. If only a model image is provided, TDOSE assumes that it represents the one object of interest in the FoV. To be able to de-blend sources when extracting spectra a model cube needs to be provided as specified under the **MODEL IMAGE SETUP** section (see below).

- **aperture.** This option performs aperture extractions as described in Sect. 3.5. The aperture sizes for this mode are provided in the **APERTURE MODEL SETUP** section of the setup file described below.

- **GAUSS MODEL SETUP.** If the default gauss extraction is performed, it is possible to provide initial estimates of the relative sizes and orientation of each source in the source catalog in this section of the setup file. This is done by providing a **SExtractor** (Bertin & Arnouts 1996) output with Gaussian morphological parameter estimates and flux scalings. This enables a quicker convergence of the reference image modeling of TDOSE. This section also provides a limit on how much the source catalog source position and the centroid of the corresponding multivariate Gaussian model are allowed to differ. This is particularly important when trying to model faint objects, as in these cases TDOSE might attempt to fit noise, and the model location might be fixed on noise peaks at random location in the FoV as opposed to the peak flux close to the location of the source.

- **MODEL IMAGE SETUP.** If a `modelimg` extraction is chosen, this part of the setup file provides the path to the directory containing the individual source models. TDOSE looks for a model named like the reference image (cutout) it is suppose to represent with “model\_” prepended. If no model is found with this name, the object is skipped. If a model appended `_cube` is found in the directory, it is assumed that this file contains a cube with the individual model components isolated in individual layers of the cube. A model cube always has priority over a model image. Based on the parent IDs in the source catalog, the model cube is used to define what sources belong to the object of interest (i.e., which sources to extract a spectrum for), and what sources should be considered contaminants. GALFIT models can be converted to TDOSE-suited model-cubes with `tdose_utilities.galfit_convertmodel2cube()`. If individual isolated source model images are available, they can be assembled into a TDOSE-suited model-cube with `tdose_utilities.build_modelcube_from_modelimages()` (see Appendix B.2 for examples).

- **APERTURE MODEL SETUP.** This section provides the sizes of the apertures to use for an aperture extraction. A single number, a list, or the location of a text file with source specific apertures can be provided.

- **PSF MODEL SETUP.** This section of the setup file defines the IFS PSF and its wavelength dependence. TDOSE version 3.0 focuses on analytic spectral extraction and therefore currently only allows a symmetric 2D Gaussian representation of the IFS PSF. The wavelength dependence of the PSF is described as a linear evolution on the form

$$FWHM(\lambda) = p_0[""] + p_1[""/\text{\AA}] \times (\lambda - p_2[\text{\AA}]), \quad (\text{A.1})$$

where each of the parameters  $p_0$ ,  $p_1$ , and  $p_2$  have to be provided. This follows the Gaussian PSF description of the MUSE-Wide fields described in Table 2 of Urrutia et al. (2019) where  $p_2 = 7050 \text{\AA}$ .

- **NON DETECTIONS.** If any of the sources listed in the source catalog (e.g., faint sources that are hard to model reliably or sources with emission lines detected in the IFS data without

<sup>5</sup> <https://github.com/kasperschmidt/TDOSE>

<sup>6</sup> <https://www.scipy.org>



clear continuum counterparts) should be treated as point sources, they can be specified in this section. Either a list of IDs or a text file listing the sources to treat as non-detections can be provided. If the source model is set to the default `gauss` option, the sources listed are replaced in the reference image models by a single point source (within the radius of ignorance, which is also provided here) before convolution with the IFS PSF. If, on the other hand, the source model mode is `modelimg`, TDOSE assumes that the provided reference image model (cube) already represents the desired extraction of the non-detection, and therefore ignores the information provided in the `NON DETECTIONS` section of the setup file

- `CUBE MODEL SETUP`. If the user is only interested in a specific part of the wavelength range of the IFS data cube, the layers to model and output can be specified here. By default, TDOSE models and extracts spectra from all wavelength layers. This part of the TDOSE setup file also provides the name scheme for the outputs to generate and defines the source model optimizer to use. The one described in Sect. 3.2, which is the standard of TDOSE, is selected as `matrix`. Optimization can also be done using `Scipy`'s non-negative least squares solver by selecting `nnls`. In this case the flux scales  $\mathbf{a}_m$  described in Sect. 3.2 are restricted to be  $\geq 0$ . As mentioned,  $\mathbf{a}_m$  describes the spectrum of a given source, hence for bright sources with continuum (and modest absorption), formally  $\mathbf{a}_m \geq 0$  should always be true. Due to noise in the IFS data, however, this is not true for fainter sources where scales can be (slightly) negative, in which case the more general `matrix` optimization is preferred.

- `SPECTRAL EXTRACTION`. If only a subset of the sources in the source catalog are of scientific interest (the remaining being considered contaminating sources), there is no reason to extract spectra of all the sources in the source catalog. Through a list of source catalog IDs (or a text file with IDs) the sources to extract spectra for are defined. This section also defines the prefix used in naming the output spectra.

- `PLOTTING`. The last part of the TDOSE setup file describes the wavelength, flux, and S/N ranges to plot for the extracted spectra. If plotting is requested, an overview plot for each object is also generated.

### A.3. TDOSE modification setup file

A key feature of the TDOSE source model cubes that can be output when extracting spectra is that they can be used to subtract undesired sources and the corresponding fluxes in full 3D from the input IFS data cube. This is beneficial for removing and/or de-blending flux of different sources, when generating 2D images and maps from the data cube as described in Sects. 3.4 and 4.6. The TDOSE modification setup file provides a simple way to perform an IFS data cube modification. All that is required to complete the modification setup file is the location and name of the data cube, the collection of 3D source models generated by TDOSE, a pointer to the output directory of the modified cube, and a list of the sources to remove in the original data cube. The setup file is then passed to `tdose_modify_cube.perform_modification()` to correct the data cube.

### A.4. TDOSE output

This section describes the main outputs of a TDOSE spectral extraction. Several of them are referred to in the description of the individual scripts and functions above. In square brackets, [ ], after each output listed below, we provide the parameter in the

TDOSE setup file (Appendix A.2) that defines the name extension of the output to ease the identification of the outputs. The main TDOSE outputs are the following:

- Reference Image Model [`model_image_ext`]. A 2D FITS Image containing a model of the reference image (cutout) if TDOSE is asked to model the sources. This model defines the morphology of the sources in the source catalog before convolution with the PSF is performed. If the source model mode is `gauss`, this model is generated by fitting multivariate Gaussians to each of the sources described in the source catalog.

- Source Model Parameters [`model_param_reg`]. A DS9 (Joye & Mandel 2003) region file marking the location and extent of the Gaussian model parameters describing the individual source models is used to generate the reference image model. This can be useful for assessing the quality of the reference image model and for locating each individual model in the FoV.

- Image Model in the IFS Frame [`model_image_cube_ext`]. A 2D Fits Image showing the reference image model in the IFS data cube reference frame, i.e., after converting the pixel scales of the reference image to match the IFS data cube.

- PSF Cube [`psf_savecube`]. A 3D FITS data cube, containing the IFS PSF model defined in the TDOSE setup file. The PSF is spatially centered and captures the wavelength evolution. If this cube is saved, it is named by prepending `_psfcube_` to the FITS file containing the source model cubes.

- Data Cube Model [`model_cube_ext`]. This output is again a 3D FITS data cube. It contains the model data cube of the IFS data cube after convolution of the reference image model with the IFS instrument PSF and after each of the individual source components in the model have been flux-optimized to match the IFS data cube flux levels at each wavelength layer in the data cube as described in Sect. 3.2.

- Residual Data Cube Model [`residual_cube_ext`]. A 3D FITS data cube providing the residual, i.e., the difference between the original IFS data cube and the 3D data cube model. The cube contains “IFS data cube” and “data cube model”.

- Source Model Cubes [`source_model_cube_ext`]. A FITS data structure containing the de-blended 3D source models for all  $n$  modeled sources in the FoV. This can be used to account for individual sources in full 3D and modify the input IFS data cube accordingly by removing individual source models with `tdose_modify_cube.py`. The dimensions of the FITS structure is [source number, wavelength, y-axis, x-axis]. Hence, the last three dimensions define the 3D model for the chosen source number.

- Extracted 1D Spectra [`spec1d_name`]. The extracted spectra are generated by collapsing the source model cubes of the sources contributing to the object of interest as described in Sect. 3.3. The individual spectra are provided as multi-extension FITS files containing a binary table with the fluxes, flux errors, S/N, and wavelength of the extracted spectrum, and the 3D object model used for the extraction. As described in Sect. 3.3, an object spectrum can be an arbitrary combination of any  $k$  sources in the FoV. However, by default each object corresponds to a single source in the source catalog, and the number of sources therefore equals the number of extracted object spectra if all spectra are extracted.

- Plots of the 1D Spectra [`plot_generate`]. If requested, TDOSE generates plots of the flux and S/N spectra extracted and an overview plot for each object extracted.

The command line output of TDOSE assembles copy-paste ready DS9 (Joye & Mandel 2003) commands that are useful for displaying and inspecting the various FITS data cubes and

images generated and used in a TDOSE extraction. To take advantage of these commands a functioning command line version of DS9 must be available.

## Appendix B: Examples for running TDOSE

In the following a few examples of calling sequences and commands to execute the spectral extractions and other tools provided as part of the TDOSE software package are presented. The shown examples of Python code can be copy-pasted, for instance into iPython after updating the inputs required. For keywords and parameters available beyond those described here, we refer to the docstring of the individual functions and procedures which are accessible, for instance by typing `? tdose.perform_extraction()` in iPython.

### B.1. TDOSE gauss extractions

This extraction mode is the default spectral extraction with TDOSE and can be performed with

```
import tdose
# --- INPUT ---
setupfile = '/Path/to/setupfile/tdose_setup_template.txt'
# --- COMMAND ---
tdose.perform_extraction(setupfile=setupfile,performcutout=
    ↪ True,generatesourcecat=True,verbose=True,verbosefull
    ↪ =True)
```

Here the `source_model` parameter in the setup file should be set to `gauss`

TDOSE positions multivariate Gaussians at the position of all sources in the source catalog when modeling the reference image. Such a source catalog can be generated with standard methods like SExtractor or can be generated by hand. In the latter case, a simple ascii file can be converted to a fits catalog with

```
import tdose_utilities as tu
# --- INPUT ---
outputdir = '/Path/to/output/directory/'
catfile = outputdir+'manually_generated_source_catalog.txt'
# --- COMMAND ---
outputfile = tu.ascii2fits(catfile,asciinames=True,
    ↪ skip_header=7,outputpath=outputdir,verbose=True)
```

It can be useful to point to a SExtractor photometric catalog with morphological estimates with the parameter `gauss_guess` in the setup file. These estimates will then be used as initial guesses for the Gaussian modeling of the FoV and will likely improve the precision of the resulting model. In the case that sources are added by hand to the source catalog, a default Gaussian point source is used unless they have shape measurements added to the SExtractor catalog.

### B.2. TDOSE modeling extractions

The extraction mode `modeling` can be used to base the TDOSE extraction on an existing model of the reference image instead of having TDOSE generate a model of multivariate Gaussians. Simply updating the keywords `sourcecatalog` (if different for the provided model), `source_model`, and `modeling_directory` and re-running the `tdose.perform_extraction()` command, as described above, extracts spectra based on the models found in the model directory. TDOSE expects to either find models of the reference image (named as the reference image prepended `model_`) or

a cube containing individual source models (named as the reference image model appended `_cube`). The latter model format is necessary for performing de-blending for modeling extractions.

#### B.2.1. Using GALFIT Models

The `tdose_utilities.py` script includes a selection of tools to handle GALFIT models consisting of Sersic and Gaussian components. More importantly, to perform de-blending using a GALFIT model a cube of the individual GALFIT components that is compatible with TDOSE can be generated:

```
import tdose_utilities as tu
# --- INPUT ---
import astropy.io.fits as afits
galfitmodel = '/Path/to/GALFIT/model/model_ref_image.fits'
sourcecat_compinfo = '/Path/to/source/component/info/
    ↪ model_componentinfo.txt'
PSFkernel = afits.open('/Path/to/reference/image/PSF/model/
    ↪ PSFmodel')[0].data
# --- COMMAND ---
tu.galfit_convertmodel2cube([galfitmodel],savecubesumimg=
    ↪ True,includewcs=True,convkernels=[PSFkernel],
    ↪ sourcecat_compinfo=sourcecat_compinfo)
```

Here the optional PSF kernel input is used to convolve the reference image model if not already accounted for.

The component infofile associates individual sources with objects in the model. TDOSE expects an ascii file containing the file model name, the object ID, and a designation of which sources in the 3D cube should belong to the object and which should be counted as contaminants. This is indicated by strings on the format “*X:Y*” where *X* counts the sources (starting from 1 corresponding to the `COMP_X` GALFIT model header keyword) and *Y* indicates whether a model component belongs to the object (*Y* = 1), is a contaminant (*Y* = 2) or represents the sky model (*Y* = 3). Hence, the following indicates a model with model components 1 and 3 belonging to the main object (ID = 55), whereas sources 2 and 4 are treated as contaminants when TDOSE de-blends the model components. The model component 5 represents the sky:

```
model_ref_image.fits 55 1:1 2:2 3:1 4:2 5:3
```

#### B.2.2. Using independent multi-Gaussian expansion source models

Alternative to using GALFIT models, individual 2D source models of individual sources can be combined to a cube of source models with the following commands:

```
import tdose_utilities as tu, glob, numpy as np
# --- INPUT ---
models2D = glob.glob('/Path/to/2D/models/*.fits')
modelsext = np.ones(len(models2D)) # list of FITS extensions
    ↪ with models
basename = '/Path/to/output/and/base/naming/combined_models'
# --- COMMAND ---
tu.build_modelcube_from_modelimages(models2D,modelsext,
    ↪ basename,savecubesumimg=True)
```

To perform de-blending for this approach, the individual de-blended spectra can be extracted based on the 3D source models generated by TDOSE using a source association dictionary, as described in Appendices A.1 and B.5.

### B.3. TDOSE aperture extractions

TDOSE is also capable of extracting aperture spectra. To do this, the setup file parameter `source_model` should be set to aperture and the aperture size(s) to use for each object should be provided via the `aperture_size` parameter. When these parameters are set the extraction can be performed by running the commands provided in Appendix B.1 above.

### B.4. Generating a large number of setup files for multiple extractions

Extractions from a large number of data cubes is handled by running multiple instances of the `tdose.perform_extraction()` function (potentially in parallel with `tdose.perform_extractions_in_parallel()`). Individual setup files can handle the individual extractions. Handling and editing a large number of TDOSE setup files can be done with the following commands:

```
import tdose_utilities as tu
# --- INPUT ---
outdir = '/Path/to/output/directory/'
infile = outdir+'tdose_setupfile_info.txt'
namebase = 'tdose_setupfile_namebase'
# --- COMMAND ---
tu.duplicate_setup_template(outdir,infile,namebase=
    ↳ namebase,loopcols='all')
```

Here the `infile` is a simple ascii file containing the values of the parameters to edit in the individual setup files. The first column named `setupname` indicates what to append to the `namebase` when naming the setup files. The rest of the columns provide the information for each of the setup file parameters (which should be used to name the columns) to replace in the template setup file (generated with `tdose_utilities.generate_setup_template()`). Hence, all setup files can be generated and edited with just a single file.

### B.5. Generating 1D spectra from 3D source models

The collection of 3D source models produced and output by TDOSE form the basis of the spectral extraction of individual objects. As described in Sect. 3.3, objects are extracted by combining and collapsing the scaled models of one or more sources from the source catalog. If multiple sources are extracted from the same data cube, TDOSE centers the cutout region on each source and performs the extraction. However, for some applications it can be useful to extract multiple objects from a

single model without re-centering the FoV. This can be done by running the extraction tool on an existing FITS structure containing the 3D source models cubes:

```
import tdose_extract_spectra as tes
# --- INPUT ---
data_cube_file = '/Path/to/data/datacube.fits'
data_cube_ext = 'DATA'
variance_cube_file = '/Path/to/data/variancecube.fits'
variance_cube_ext = 'VAR'
model_cube_file = '/Path/to/data/
    ↳ datacube_tdose_modelcube_gauss.fits'
model_cube_ext = data_cube_ext
sourcemodels_file = '/Path/to/data/
    ↳ datacube_tdose_source_modelcube_gauss.fits'
sourcemodels_ext = data_cube_ext
nameextension = 'tdose_spectrum_manual_extract'
outdir = '/Path/to/output/directory/'
SAD = {111:[0,4,5], 222:[2], 333:[1,6,7]}
# --- COMMAND ---
specfiles = tes.extract_spectra(model_cube_file,
    ↳ model_cube_ext=model_cube_ext,nameext=nameextension,
    ↳ source_association_dictionary=SAD, outdir=
    ↳ outdir, variance_cube_file=variance_cube_file,
    ↳ variance_cube_ext=variance_cube_ext,
    ↳ source_model_cube_file=sourcemodels_file,
    ↳ source_cube_ext=sourcemodels_ext, data_cube_file=
    ↳ data_cube_file, verbose=True)
```

Here SAD is the source association dictionary associating sources with individual objects (see also Appendix A.1). In the above example, object 111 is comprised of source 0, 4, and 5; object 222 corresponds to source 2; and object 333 is comprised of source 1, 6, and 7 from the reference image model.

### B.6. Modifying data cubes using 3D source models

The collection of 3D source models produced by TDOSE enable correcting the intrinsic flux data cubes for contamination in full 3D, as described in Sect. 3.4 and shown in Sect. 4.6. Modification of original IFS data cubes and thereby removing contaminating flux based on the TDOSE source models is controlled by the modification setup file described in Appendix A.3 above. The modification is performed as follows:

```
import tdose_modify_cube as tmc
# --- INPUT ---
setupfile = '/Path/to/modify/setupfile/
    ↳ tdose_setup_template_modify.txt'
# --- COMMAND ---
tmc.perform_modification(setupfile=setupfile)
```