University of Groningen

The very knotty lenser: Exploring the role of regularization in source and potential reconstructions using Gaussian process regression

Vernardos, G.; Koopmans, L. V. E.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

The very knotty lenser: Exploring the role of regularization in source and potential reconstructions using Gaussian process regression

G. Vernardos [1,2]★ and L. V. E. Koopmans[2]

[1]*Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[2]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands*

ABSTRACT
Reconstructing lens potentials and lensed sources can easily become an underconstrained problem, even when the degrees of freedom are low, due to degeneracies, particularly when potential perturbations superimposed on a smooth lens are included. Regularization has traditionally been used to constrain the solutions where the data failed to do so, e.g. in unlensed parts of the source. In this exploratory work, we go beyond the usual choices of regularization and adopt observationally motivated priors for the source brightness. We also perform a similar comparison when reconstructing lens potential perturbations, which are assumed to be stationary, i.e. permeate the entire field of view. We find that physically motivated priors lead to lower residuals, avoid overfitting, and are decisively preferred within a Bayesian quantitative framework in all the examples considered. For the perturbations, choosing the wrong regularization can have a detrimental effect that even high-quality data cannot correct for, while using a purely smooth lens model can absorb them to a very high degree and lead to biased solutions. Finally, our new implementation of the semi-linear inversion technique provides the first quantitative framework for measuring degeneracies between the source and the potential perturbations.

Key words: gravitational lensing: strong.

1 INTRODUCTION

The standard cosmological model, comprising the still unknown dark energy and dark matter, has been successful in describing the large-scale structure of the Universe and its properties (>1 Mpc; e.g. Komatsu et al. 2011; Planck Collaboration VI 2020). The dark matter component in particular, plays an important role throughout cosmic evolution by participating in the collapse of baryons via gravitational instability to form galaxies (White & Rees 1978). Verifying the validity of the current Cold Dark Matter paradigm down to sub-galactic scales, and what this implies for the microscopic properties of the dark matter particle, is masked by the onset of highly non-linear physical mechanisms attributed to baryons, e.g. stellar winds, supernovae, feedback from Active Galactic Nuclei, etc, that appear in such high density environments (Vogelsberger et al. 2014; Schaye et al. 2015).

The tension between dark matter theory and observations on galactic and sub-galactic scales (<1 Mpc; Bullock & Boylan-Kolchin 2017) has several manifestations, e.g. the 'missing satellites' (Klypin, Kravtsov & Valenzuela 1999; Moore et al. 1999), the 'cusp-core' (Moore 1994; Oh et al. 2015), the 'too-big-to-fail' (Boylan-Kolchin, Bullock & Kaplinghat 2011), and the 'bulge-halo conspiracy' (Dutton & Treu 2014) problems. Regardless of the role of baryons and their gravitational interactions with dark matter in each case, aspects of which constitute independent major research

fields (e.g. the efficiency of star formation; McKee & Ostriker 2007), measuring the overall shape and smoothness of the mass density in galaxies is critical. In the local Universe, this can be achieved by understanding the statistics (e.g. Papastergis et al. 2014), instrumental effects (Kim, Peter & Hargis 2018), and dynamics (e.g. Helmi et al. 2012) of dwarf galaxies and stellar streams (e.g. Carlberg, Grillmair & Hetherington 2012; Erkal et al. 2016).

As soon as one leaves the neighbourhood of the Milky Way, the only way to achieve such measurements is via gravitational lensing – the deflection of light from a distant source by the intervening mass of a galaxy. In this way, the overall shape of the total mass distribution has been measured for massive elliptical galaxies (Koopmans et al. 2006, 2009; Gavazzi et al. 2007; Auger et al. 2010; Barnabè et al. 2011; Sonnenfeld et al. 2013; Suyu et al. 2014; Oldham & Auger 2018) and massive substructures down to the order of 10^8 M$_\odot$ have been detected out to cosmological distances (Vegetti et al. 2010, 2012; Fadely & Keeton 2012; MacLeod et al. 2013; Nierenberg et al. 2014; Hezaveh et al. 2016b; Li et al. 2016; Birrer & Amara 2017). Strong lensing analysis has also been combined with other techniques, e.g. stellar kinematics (Barnabè et al. 2011; Yildirim, Suyu & Halkola 2020), stellar kinematics and weak lensing (Sonnenfeld et al. 2018), stellar population analysis (Barnabe et al. 2013; Smith, Lucey & Conroy 2015; Spiniello et al. 2015), and quasar microlensing (Oguri, Rusu & Falco 2014), in order to disentangle the baryonic and dark mass components.

The gravitational imaging technique (Koopmans 2005; Vegetti & Koopmans 2009) is a powerful method to study the non-smoothness of the lensing mass distribution, analysing perturbations of lensing

features, such as arcs and Einstein rings.[1] Based on the semi-linear inversion method of Warren & Dye (2003), which can reconstruct the light distribution of the lensed source on a grid once the lensing potential is given, Koopmans (2005) provided an extension that simultaneously obtains a grid-based reconstruction of potential perturbations to an overall smooth (parametric) lens potential: in the presence of substructure, dark or luminous, the smooth modelling residuals are remodelled in terms of lens potential perturbations using the smooth potential and its corresponding source as a starting point. Vegetti & Koopmans (2009) improved this technique in a number of ways, expanding the work of Suyu et al. (2006) by casting the problem in a Bayesian framework that includes the potential perturbations and using an adaptive grid for the source. With careful control over the regularization level of the solutions, the presence of substructure in a lens can be uncovered by accumulating small potential corrections within an iterative scheme. The detection is then justified by comparing the Bayesian evidence to the best purely smooth lensing model (Vegetti et al. 2010).

The regularization scheme plays a critical role in such a strong lensing Bayesian analysis approach, as it enables the matrix inversions to find a unique solution (MacKay 1992, 2003). Focusing only on the reconstruction of the source, there are several pixel-based methods[2] that employ a brightness, gradient, or curvature based regularization scheme, or a combination thereof (Dye & Warren 2005; Suyu et al. 2006; Vegetti & Koopmans 2009; Tagore & Keeton 2014; Nightingale & Dye 2015; Yildirim et al. 2020), i.e. they assume that each of these source properties is drawn from a normal distribution, whose variance is determined by the regularization parameter that itself can be optimized for, and whose correlation properties are set by a corresponding covariance matrix. However, a poor choice of the regularization parameter in each case is known to cause problems with over- and underfitting of the data in some cases, which in turn might affect the mass model parameters (Nightingale & Dye 2015). Suyu et al. (2006) solve exactly for the value of the regularization parameter that maximizes the evidence. To allow for more flexibility, Nightingale, Dye & Massey (2018) have introduced a non-constant (adaptive) regularization scheme, whose principle is to vary the strength of the regularization (width of the normal distribution) across the source, based on its surface brightness profile. Some form of regularization is necessary to be able to solve the equations, however all of these methods are equivalent to setting priors for the different source properties that are not necessarily astrophysically motivated.

Upon combining the source reconstruction with potential perturbations, which enter the equations in a very similar way to the source and require their own regularization scheme, an additional non-linear dependence of the perturbations on the source is introduced (Koopmans 2005). Again, the regularization of the two fields, the source and the perturbations, plays an important role in reaching a unique solution. Vegetti & Koopmans (2009) follow a line-search optimization, starting with finding the best

smooth lens-mass model and then proceeding with calculating potential corrections based on the corresponding source (see also equation 3 here). In their iterative scheme, the source and potential perturbations are solved for at each step and then updated: the new surface brightness derivatives are calculated across the source and the perturbations are added to the overall smooth potential in the form of corrections. The regularization parameter of the perturbations is carefully controlled, initially set to very high values (very smooth fields) and later reduced to allow for more structure. This is similar to a Gauss–Newton optimization scheme that is known to be sensitive to the step size; any spurious structure appearing in the solutions would be added to the overall lensing potential with the risk of irrecoverably drifting away from the true solution. Although this is a powerful approach, it is limited by two caveats: some manual fine-tuning is needed in setting up the algorithm to converge to a meaningful solution, and there is no obvious means to quantify degeneracies between the reconstructed source and the potential perturbations. The latter is inherent to the technique and has not been studied in depth before (see Chatterjee 2019, for an example).

In this paper, we more rigorously investigate the importance of new forms of regularization, introducing more realistic priors on the source surface brightness distribution that are more flexible in capturing higher order statistical properties, and a statistical approach to finding the best regularization parameters via sampling. The latter is based on the theory of Gaussian Process Regression (Rasmussen & Williams 2006) and is quite powerful as it provides a way to quantify degeneracies between the source and perturbation fields. In addition, this sampling approach is better suited to describe extended perturbations, which are not necessarily restricted to compact and well-localized perturbers that might be more accurately detected by an iterative and additive scheme (as in Vegetti & Koopmans 2009). The outcome is a statistical approach to generic perturbations of a smooth lensing potential, which can be directly linked to the underlying statistical properties of baryonic and dark matter (e.g. via the power spectrum), or to higher order structure in the lens potential, such as the presence of a galactic disc (as was recently found by Hsueh et al. 2017).

The structure of the paper is as follows. In Section 2 we set up the theoretical framework, provide the Bayesian evidence equation extending the work of Suyu et al. (2006) and Vegetti & Koopmans (2009), and demonstrate the use of this approach under various regularization schemes. Section 3 presents a set of selected applications of the method on mock lens systems, which are discussed further in Section 4. Our conclusions are summarized in Section 5.

2 METHOD

The Bayesian formalism applied to grid-based strong lensing analyses was introduced by Suyu et al. (2006) and Vegetti & Koopmans (2009). Here, we use the same framework and repeat some of the steps, while we point out the differences, particularly with respect to the regularization and our sampling approach. In addition, an explicit equation describing the Bayesian evidence is derived, which has not appeared in the literature so far (Suyu et al. 2006 give such an expression but including only the source).

First, we formulate the problem in terms of a lensing operator depending on a parametrized smooth lens potential and a source brightness distribution defined on a grid, and then we introduce potential perturbations. Solving the resulting equations directly is an ill-posed problem. We therefore need to look for solutions minimizing some form of penalty function that includes regularization. This leads to a set of linear equations with respect to the source and

---

[1] This can also be achieved by analysing flux ratios from lensed quasars (Dalal & Kochanek 2002), however, this requires carefully planned spectroscopic observations, taking into account the possible effect of microlensing (e.g. Nierenberg et al. 2014).

[2] The possibility of using basis sets to reconstruct the source has been explored by Birrer, Amara & Refregier (2015), Joseph et al. (2019), Galan et al. (2021) and the use of deep neural networks was investigated by Morningstar et al. (2019). Both methods do not explicitly require regularization, but rely on the number of independent basis vectors and a descriptive training set, respectively, to model higher order statistics of the source.

the potential perturbations that has an exact solution. The problem is then re-cast using a Bayesian formalism and the expression of the evidence is derived. The general treatment is independent of any assumption on the particular type of regularization, however, several physically motivated schemes are examined in more detail. Finally, we present our sampling approach that allows us to (1) compare different choices of model components (e.g. the regularization scheme) based on their Bayesian evidence, and (2) for any given model, obtain the probability distribution of all of its non-linear parameters and determine statistical uncertainties, systematic biases, and degeneracies.

2.1 The lensing operator and the source grid

The problem at hand is finding how the lensed images relate to the brightness of the background source via gravitational lensing, and can be cast in the following way (similarly to Warren & Dye 2003; Koopmans 2005; Vegetti & Koopmans 2009):

$$d = BL(\psi)s + n, \tag{1}$$

where d and n are the vectors of brightness measurements (the 'data') and the associated noise (the 'noise') of the image pixels, s is the vector of the (unknown) source brightness (the 'source'), B is the blurring operator that is linked to the point spread function (PSF), and L is the lensing operator that depends on the lensing potential ψ. The data and noise vectors correspond to a rectangular $M \times N$ grid of N_d pixels in total on the image plane, which delineates the part of the pixel array of the optical detector covering the lensed images. The blurring operator ($N_d \times N_d$) is assumed constant[3] and mimics the effect of the PSF; it acts on (blurs) the resulting image plane pixels with a fixed weighting scheme, after the source has been lensed. Assuming that the source can also be described by a pixelated grid of N_s pixels and arbitrary form on the source plane, then the lensing operator ($N_d \times N_s$) couples each data pixel position to the source grid via the lens equation (Vegetti & Koopmans 2009). This can introduce multiplicity because different image pixels can be associated with the same source location, thus creating multiple images. Equation (1) is a linear transformation between the image and source planes that depends on the gradient of the lensing potential ψ. We note that the lensing potential is typically a non-linear function of the lens plane coordinates, x, and some parameters, η, that can vary in complexity.

Once the positions of the data pixels are traced back to the source plane, they are matched to pixels on the source grid via an interpolation scheme that guarantees the conservation of surface brightness (see fig. 1 in Koopmans 2005). The source grid can have any arbitrary structure, e.g. fixed or free-floating regular grids, irregular, adaptive, etc. On a regular grid, bi-linear interpolation is sufficient, while higher order schemes could also be used (e.g. bi-cubic, natural neighbour, etc). An irregular grid has a unique Delaunay triangulation and a corresponding dual Voronoi tesselation, whose triangles and cells, respectively, can both be considered as source 'pixels' (Gallier 2011). Data pixels that are cast back on to the source plane land inside a Delaunay triangle and their value is interpolated linearly between the triangle's vertices (the centres of the irregular Voronoi cells). Hence, the brightness values inside any such triangle lie on a tilted plane defined by the values at the triangle vertices. Barycentric coordinates are used to perform these triangular

interpolations, which is equivalent to the procedure described in Vegetti & Koopmans (2009, figs 1 and 2).

An irregular source grid can also be constructed randomly (e.g. Nightingale & Dye 2015) or by a recipe designed to facilitate the source reconstruction. An example is a so-called adaptive grid that is reconstructed every time the lens potential $\psi(\eta)$ changes. Here, we create such adaptive grids by casting back one out of every $n \times n$ block of the data pixels, with $1 \leq n < 6$ (fixed throughout the reconstruction). Alternative gridding techniques are known to affect the 'discreteness-noise' in the computed Bayesian evidence and χ^2 terms (Tagore & Keeton 2014; Nightingale & Dye 2015). However, exploring different grids is out of this paper's scope and left for future improvements to our method. For very large values of n the resulting grid will be too coarse to successfully describe a detailed lensed image brightness distribution. For $n = 1$, there is no need for any interpolation as all the data pixels have been used to create the source grid ($N_s = N_d$). However, in this case the system of equations to solve is underconstrained and heavily relies on the regularization (i.e. assumed prior on the source surface brightness).

Applying this procedure for any given lens potential $\psi(\eta)$ results in a set of N_s points on the source plane representing the positions of the source brightness values s and a $N_d \times N_s$ operator L, whose rows contain the interpolation weights on the source grid for each data pixel. The procedure is repeated each time the lens potential ψ changes (Vegetti & Koopmans 2009).

2.2 Lens potential corrections

Often, an elliptical power-law mass model is assumed for the lens (Kassiola & Kovner 1993; Barkana 1998). However, such smooth lens potential models may well be too simplified to capture more detailed structure of real lenses. Deviations from smoothness could be the result of dark matter substructure or higher order moments in the mass distribution of the lens galaxy itself, originating from its morphology (e.g. Hsueh et al. 2017, find a non-negligible disc component) or evolution history (e.g. mergers). If such deviations exist in an observed system, they will manifest themselves as residuals, δd, left behind after modelling the lens with a smooth potential:

$$\delta d = Ms_p - d, \tag{2}$$

where $M \equiv M(\eta) = BL(\eta)$, and s_p is the solution for the source after inverting the smooth model as described in Section 2.3 (the subscript p indicates a previously obtained quantity). Such residuals will persist regardless of the smooth potential used to describe the lens, although they may be absorbed to some degree into the source surface brightness or by modifying the values of the parameters η (e.g. Bayer 2021).

If the residuals from the smooth modelling are not noise-like, then the inclusion of a new lens potential component may be warranted in order for $\delta d \rightarrow 0$ (or, more precisely, δd reaching the properties of the noise). The most general treatment of such a component is assuming a potential perturbations field, $\delta\psi$, which to first order can de described by (Koopmans 2005):

$$\delta d = -BD_s(s_p)D_{\delta\psi}\delta\psi, \tag{3}$$

where $D_s(s_p)$ is a matrix containing the gradient of the previously known source s_p, and $D_{\delta\psi}$ is the gradient operator of the potential perturbations that yield $\delta\alpha$, the perturbative deflection angle vector (see appendix A of Koopmans 2005, for a derivation of this equation). This equation describes how potential perturbations induce additional deflections, causing the positions in the image plane to become associated with a different position in the source plane, and hence with a different source brightness. These deflections are assumed to be small enough for the source to be well approximated

---

[3]The PSF can in fact vary for each pixel based on the spectral energy distribution of the source for that specific pixel, or due to atmospheric effects if we are dealing with ground-based observations.

by a first-order Taylor expansion around the original unperturbed locations. In this way, the residual image plane brightness of the smooth model can be associated with the gradient of the source brightness and some small potential perturbation field.

Equations (1), (2), and (3) can be combined to reformulate the lensing problem as (Koopmans 2005; Vegetti & Koopmans 2009):

$$d = M_r r + n, \qquad (4)$$

where M_r is the block matrix:

$$M_r \equiv M_r(\psi_p, s_p) = B[L(\psi_p)| - D_s(s_p)D_{\delta\psi}], \qquad (5)$$

and:

$$r \equiv \begin{pmatrix} s \\ \delta\psi \end{pmatrix}. \qquad (6)$$

The similarity with equation (1) is striking, however, there is one important difference: some prior knowledge of the source brightness is necessary to construct the matrix $D_s(s_p)$. The lens potential ψ_p can either depend solely on η (in which case there is no need for the subscript), as is the case in equation (1), or it can include accumulated corrections $\delta\psi_p$ derived at previous stages – similarly to a Gauss–Newton scheme where a small update to the previous solution is calculated via a linear extrapolation.

The $\delta\psi$ field can be approximated by $N_{\delta\psi}$ pixels on the image plane, which we here assume to be on a fixed regular P×Q grid (as opposed to, for example, being adaptive). The $D_s(s_p)$ matrix entries are calculated at the locations of the deflected data pixels on the source grid. Similarly, the $D_{\delta\psi}$ operator determines the derivatives of $\delta\psi$ at the locations of the data pixels on the image plane. The product $D_s(s_p)D_{\delta\psi}\delta\psi$ is a $N_d \times N_{\delta\psi}$ matrix, whose rows contain the terms:

$$[D_s(s_p)D_{\delta\psi}\delta\psi]_i = \frac{\partial s_p(y_i)}{\partial y_1}\frac{\partial \delta\psi(x_i)}{\partial x_1} + \frac{\partial s_p(y_i)}{\partial y_2}\frac{\partial \delta\psi(x_i)}{\partial x_2}, \qquad (7)$$

where x are the data pixel coordinates on the image plane and y their corresponding source plane positions. If the data and perturbation grids coincide this matrix is diagonal, but usually the $\delta\psi$ grid has a lower resolution such that each matrix row will contain the terms and corresponding weights resulting from a bilinear (in our case) interpolation on the $\delta\psi$ grid (i.e. $\delta\psi(x_i) = \sum_{j=1}^4 w_{i,j}\delta\psi_{i,j}$, where the j-th index goes over the four vertices of the $\delta\psi$ pixel encompassing the i-th data pixel).

2.3 Model inversion

The observed data result from the physical and instrumental processes of lensing and blurring, described as operators acting on a gridded source (their order is important), the finite detector pixel size, and the inclusion of noise with some properties (e.g. statistical Poisson noise of photon counts, correlated noise introduced at data reduction, cosmic rays, etc). Even in the absence of noise, inverting equation (1) for the source is generally an ill-posed problem that does not have a unique or exact solution. One way to proceed is by searching for a source solution that minimizes a regularized penalty function. First, we define the penalty function, which is a likelihood function under the assumption of Gaussian errors in the data, excluding the perturbations $\delta\psi$, to be the sum of a generalized χ^2 and a regularization term:

$$G(s) \equiv G(s|d, \eta, g_s, \lambda_s)$$
$$= \frac{1}{2}(Ms - d)^T C_d^{-1}(Ms - d) + \frac{1}{2}\lambda_s s^T C_s^{-1}(g_s)s, \qquad (8)$$

where M is the operator used in equation (2), and C_d and C_s are the covariance matrices of the data and the source. In the case of

the source, C_s may in general be a function of another set of non-linear regularization parameters, g_s – we take out the regularization parameter λ_s to separate the effect of the overall level of source brightness from the shape of its correlations and make its effect more explicit. The parameter λ_s sets the level of contribution to the overall penalty of the regularization term with respect to the value of χ^2. In the following, the covariance matrix C_s is always assumed to be a function of g_s, while specific regularization schemes are discussed in Section 2.5.

The source property used for regularization (gradient, curvature, etc) is assumed to be distributed normally, i.e. a quadratic form in equation (8), similarly to the χ^2 term, guaranteeing that the source for which $\nabla_s G = 0$ minimizes the penalty function (Suyu et al. 2006). Using this condition, after some basic algebraic manipulations, we get:

$$(M^T C_d^{-1} M + \lambda_s C_s^{-1})s = M^T C_d^{-1} d, \qquad (9)$$

where the matrix $M^T C_d^{-1} M + \lambda_s C_s^{-1}$ is now positive-definite and can be inverted using standard techniques. The source that minimizes the penalty function is found in this way for each set of fixed η, g_s, and λ_s. This solution implicitly assumes that the Gaussian random field describing the source has a zero mean. Although this is not formally correct because of the finite dimensions of the source grid, this offset is in general easily absorbed by the shape of the covariance matrix, and as our tests later will show, this assumption holds to very good approximation.

Often, masking the data is required to isolate and model only the lensed image features. This can be achieved by an operator S, acting on the image plane and excluding all the pixels outside the mask from the modelling. This is simply a diagonal matrix with values of 1 or 0 for included and excluded pixels, respectively. In equations (8) and (9), this can be incorporated into a 'masked' covariance matrix $S^T C_d^{-1} S$, all rest being the same. In the remaining treatment, C_d^{-1} and $S^T C_d^{-1} S$ can be used interchangeably.

Reformulating the problem to include the potential perturbations is straightforward due to the similarity of equations (1) and (4). As before, in general equation (4) cannot be directly inverted and we have to proceed by minimizing some penalty function. Here we define such a function similarly to equation (8), including an additional regularization term for the potential perturbations in the same way as for the source:

$$G(r) \equiv G(r|d, s_p, \psi_p, g_s, \lambda_s, g_{\delta\psi}, \lambda_{\delta\psi})$$
$$= \frac{1}{2}(M_r r - d)^T C_d^{-1}(M_r r - d) + \frac{1}{2}r^T R r, \qquad (10)$$

where:

$$R = \begin{pmatrix} \lambda_s C_s^{-1} & 0 \\ 0 & \lambda_{\delta\psi} C_{\delta\psi}^{-1} \end{pmatrix}. \qquad (11)$$

We underline again that the important difference with equation (8) is the dependence on a previously known source, s_p (through M_r). This equation has the same form as equation (8), and the condition $\nabla_r G = 0$ leads to:

$$(M_r^T C_d^{-1} M_r + R)r = M_r^T C_d^{-1} d, \qquad (12)$$

which can be solved for r by inverting the positive-definite matrix on the left-hand side.

2.4 Bayesian framework

The number of free parameters involved in the lens potential and source reconstruction may vary between different models. For example, one may choose different parametric models for the smooth

mass distribution, with or without additional perturbations, and regularization schemes (see Section 2.5). As in Suyu et al. (2006) and Vegetti & Koopmans (2009), we use a Bayesian approach to quantitatively justify the inclusion of extra free parameters and compare models to find the one most consistent with the data - assuming all quantities are Gaussian processes. By recasting the problem in Bayesian terms, the evidence term necessary to compare models can be computed. In addition, the solutions for the source and the perturbations obtained in the previous section, which minimize the penalty function, coincide with the most probable solutions that maximize the posterior probability. A similar treatment is followed in Suyu et al. (2006) and Vegetti & Koopmans (2009), however, here we explicitly derive the expression for the evidence.

Bayes' theorem states that the posterior probability density of the source and potential perturbations given the data, lensing operator, and some form of prior (regularization) described by parameters $\boldsymbol{g}$ and λ is:

$$
\begin{aligned}
P(\boldsymbol{r}) &\equiv P(\boldsymbol{r}|\boldsymbol{d}, \boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\mathrm{s}}, \lambda_{\delta\psi}) \\
&= \frac{P(\boldsymbol{d}|\boldsymbol{r}, \boldsymbol{\eta}) \, P(\boldsymbol{s}|\boldsymbol{g}_{\mathrm{s}}, \lambda_{\mathrm{s}}) \, P(\delta\boldsymbol{\psi}|\boldsymbol{g}_{\delta\psi}, \lambda_{\delta\psi})}{P(\boldsymbol{d}|\boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\mathrm{s}}, \lambda_{\delta\psi})},
\end{aligned} \quad (13)
$$

where the numerator terms are the likelihood, source prior, and potential perturbations prior, respectively, and the denominator is the evidence for this specific problem. Assuming the likelihood and priors are normal distributions and associating them with the previously introduced χ^2 and regularization terms, their individual probability densities can be written as:

$$
\begin{aligned}
P(\boldsymbol{d}|\boldsymbol{r}, \boldsymbol{\eta}) &= \frac{1}{Z_{\mathrm{d}}} \exp\left[-\frac{1}{2}(M_{\mathrm{r}}\boldsymbol{r} - \boldsymbol{d})^T C_{\mathrm{d}}^{-1}(M_{\mathrm{r}}\boldsymbol{r} - \boldsymbol{d})\right], \\
P(\boldsymbol{s}|\boldsymbol{g}_{\mathrm{s}}, \lambda_{\mathrm{s}}) &= \frac{1}{Z_{\mathrm{s}}} \exp\left[-\frac{1}{2}\lambda_{\mathrm{s}}\boldsymbol{s}^T C_{\mathrm{s}}^{-1}\boldsymbol{s}\right], \\
P(\delta\boldsymbol{\psi}|\boldsymbol{g}_{\delta\psi}, \lambda_{\delta\psi}) &= \frac{1}{Z_{\delta\psi}} \exp\left[-\frac{1}{2}\lambda_{\delta\psi}\delta\boldsymbol{\psi}^T C_{\delta\psi}^{-1}\delta\boldsymbol{\psi}\right],
\end{aligned} \quad (14)
$$

where the normalization factors are given by:

$$
\begin{aligned}
Z_{\mathrm{d}} &= (2\pi)^{N_{\mathrm{d}}/2}(\det C_{\mathrm{d}})^{1/2}, \\
Z_{\mathrm{s}}(\boldsymbol{g}_{\mathrm{s}}, \lambda_{\mathrm{s}}) &= \left(\frac{2\pi}{\lambda_{\mathrm{s}}}\right)^{N_{\mathrm{s}}/2}(\det C_{\mathrm{s}})^{1/2}, \\
Z_{\delta\psi}(\boldsymbol{g}_{\delta\psi}, \lambda_{\delta\psi}) &= \left(\frac{2\pi}{\lambda_{\delta\psi}}\right)^{N_{\delta\psi}/2}(\det C_{\delta\psi})^{1/2}.
\end{aligned} \quad (15)
$$

The above set of equations assumes that we already have a decent solution for the source, $\boldsymbol{s}_{\mathrm{p}}$, in order to derive M (see equation 5), which could come, for example, by solving the smooth version of the problem (see Koopmans 2005). The most probable solution for the source and the perturbations – the one that maximizes the posterior probability – can be derived by requiring $\nabla_{\mathrm{r}} P = 0$ in equation (13). This can be calculated independently of the evidence term, which is a constant factor in this case. This is the solution that also minimizes the penalty function in equation (10), which has already been given in equation (12).

The posterior in equation (13) is the product of equations (14), hence it is itself a normal distribution and can be written as:

$$
P(\boldsymbol{r}) = \frac{1}{Z_{\mathrm{G}}} \exp[-G(\boldsymbol{r})], \quad (16)
$$

where $G(\boldsymbol{r})$ is given in equation (10). Taking a Taylor expansion of G around the most probable solution $\boldsymbol{r}_{\mathrm{MP}}$, which satisfies $\nabla_{\mathrm{r}} G = 0$ (equation 12), we get:

$$
G(\boldsymbol{r}) = G(\boldsymbol{r}_{\mathrm{MP}}) + \frac{1}{2}(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{MP}})^T H(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{MP}}), \quad (17)
$$

where H is the Hessian of G:

$$
H \equiv \nabla_{\mathrm{r}}^2 G = M_{\mathrm{r}}^T C_{\mathrm{d}}^{-1} M_{\mathrm{r}} + R. \quad (18)
$$

Equation (17) is in fact exact – assuming we already know $\boldsymbol{s}_{\mathrm{p}}$ – because all terms in equation (10) are quadratic in $\boldsymbol{r}$. Equation (16) can now be rewritten as:

$$
P(\boldsymbol{r}) = \frac{1}{Z_{\mathrm{G}}} \exp\left[-G(\boldsymbol{r}_{\mathrm{MP}}) - \frac{1}{2}(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{MP}})^T H(\boldsymbol{r} - \boldsymbol{r}_{\mathrm{MP}})\right], \quad (19)
$$

where:

$$
\begin{aligned}
Z_{\mathrm{G}} &\equiv Z_{\mathrm{G}}(\boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \lambda_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\delta\psi}) \\
&= e^{-G(\boldsymbol{r}_{\mathrm{MP}})}(2\pi)^{(N_{\mathrm{s}}+N_{\delta\psi})/2}(\det H)^{-1/2}.
\end{aligned} \quad (20)
$$

Combining equations (10), (14), (17), and (19) the evidence term from equation (13) can be computed for the most probable solution $\boldsymbol{r}_{\mathrm{MP}}$:

$$
P(\boldsymbol{d}|\boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\mathrm{s}}, \lambda_{\delta\psi}) = \frac{Z_{\mathrm{G}}(\boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \lambda_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\delta\psi})}{Z_{\mathrm{d}} Z_{\mathrm{s}}(\boldsymbol{g}_{\mathrm{s}}, \lambda_{\mathrm{s}}) Z_{\delta\psi}(\boldsymbol{g}_{\delta\psi}, \lambda_{\delta\psi})}. \quad (21)
$$

Substituting the normalization factors from equations (15) and (20), and taking the logarithm of the evidence we get:

$$
\begin{aligned}
\log P = {} &-\frac{N_{\mathrm{d}}}{2}\log(2\pi) + \frac{N_{\mathrm{s}}}{2}\log(\lambda_{\mathrm{s}}) + \frac{N_{\delta\psi}}{2}\log(\lambda_{\delta\psi}) - \frac{1}{2}\log(\det C_{\mathrm{d}}) \\
&-\frac{1}{2}\log(\det C_{\mathrm{s}}) - \frac{1}{2}\log(\det C_{\delta\psi}) - G(\boldsymbol{r}_{\mathrm{MP}}) - \frac{1}{2}\log(\det H).
\end{aligned} \quad (22)
$$

Deriving an analytic expression for this term is necessary in order to compute the posterior probability distribution of the non-linear parameters of the problem, i.e. the smooth lens potential and regularization parameters. The latter is given by:

$$
\begin{aligned}
&P(\boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\mathrm{s}}, \lambda_{\delta\psi}|\boldsymbol{d}) = \\
&\quad \frac{P(\boldsymbol{d}|\boldsymbol{\eta}, \boldsymbol{g}_{\mathrm{s}}, \boldsymbol{g}_{\delta\psi}, \lambda_{\mathrm{s}}, \lambda_{\delta\psi}) \, P(\boldsymbol{\eta}) \, P(\boldsymbol{g}_{\mathrm{s}}) \, P(\boldsymbol{g}_{\delta\psi}) \, P(\lambda_{\mathrm{s}}) \, P(\lambda_{\delta\psi})}{\mathcal{E}(\boldsymbol{d})},
\end{aligned} \quad (23)
$$

where the first term in the numerator – the likelihood – is precisely the previous evidence term from equation (13) that has been derived in equation (22), and the remaining terms are individual parameter priors. It is the evidence term in this last equation, $\mathcal{E}(\boldsymbol{d}) \equiv \mathcal{E}(\boldsymbol{d}|H_{\psi}, H_{\mathrm{s}}, H_{\delta\psi})$, where H are specific choices (hypotheses) of a smooth lens potential model and regularization schemes for the source and the potential perturbations, that eventually allows quantitative model comparisons. Similarly to Suyu et al. (2006, but see also MacKay 1992), the posterior distribution of our model choices is:

$$
P(H_{\psi}, H_{\mathrm{s}}, H_{\delta\psi}|\boldsymbol{d}) \propto \mathcal{E}(\boldsymbol{d}|H_{\psi}, H_{\mathrm{s}}, H_{\delta\psi}) \, P(H_{\psi}) \, P(H_{\mathrm{s}}) \, P(H_{\delta\psi}). \quad (24)
$$

In this work, we calculate the term $\mathcal{E}$ by calculating the integral of the posterior (see Section 2.6). This allows us to rank different regularization schemes and find the one most consistent with the data (the one with the largest value; MacKay 2003). Here we do not compare different smooth lens potential models, but exactly the same procedure can be used to achieve this.

2.5 Regularization schemes

Adding regularization terms to the penalty function (equation 10), or equivalently using priors in the posterior probability density (equation 13), is necessary in order to find a solution for the source and the potential perturbations by inverting the matrices in equations (9) and (12). Quadratic terms such as the ones used

here (Gaussian priors), as opposed to other forms of regularization,[4] have the advantage of leading to linear equations that have exact and efficient to calculate analytic solutions (equations 9 and 12), and put the problem in the context of Gaussian process regression (Rasmussen & Williams 2006).

The effect of the regularization on the source and perturbation fields is captured in the detailed structure of the generic covariance matrices C_s and $C_{\delta\psi}$, while the overall contribution to the penalty function (posterior probability) is moderated by the λ_s and $\lambda_{\delta\psi}$ parameters. Here we examine different physically motivated forms of the covariance matrices C_s and $C_{\delta\psi}$, and because the treatment is the same for both source and perturbations, we simply use C and λ in the following.

The usual forms of regularization impose some sort of 'smoothness' condition on the solution (see Press et al. 1992). Choices in the literature are based on derivatives of some order (e.g. Dye & Warren 2005; Suyu et al. 2006; Vegetti & Koopmans 2009; Tagore & Keeton 2014; Nightingale & Dye 2015; Yildirim et al. 2020). For example, a zero-th order derivative of the source (the usual Tikhonov regularization, or ridge regression; Tikhonov 1963) means that C is the identity matrix and brightness values are derived from a normal distribution centred on zero with standard deviation $\lambda^{-1/2}$. Similarly, the gradient and curvature regularizations constrain the corresponding source derivatives, imposing a varying degree of smoothness to the solution. However, although such schemes are useful to find a solution to the problem, they are not physically motivated (there is no reason for the gradient or curvature of a galaxy's brightness profile to follow a normal distribution centred at zero or any other value), may introduce spurious properties to the solutions, and cause degeneracies between the source and the lens potential. In other words, the assumed covariance matrix resulting from these choices imposes a correlation function (or power spectrum) on the source or potential perturbations that might not reflect reality.

A more realistic and general approach can involve covariance matrices C that do not correspond to any particular derivative and impose a physically motivated structure, via its covariance, directly on the solutions. Here we examine two forms of such covariance kernels:

$$C(\mathbf{y}_i, \mathbf{y}_j, l) = \exp\left(-\frac{d_{i,j}}{l}\right), \text{ (exponential)} \quad (25)$$

$$C(\mathbf{y}_i, \mathbf{y}_j, l) = \exp\left(-\frac{d_{i,j}^2}{2\,l^2}\right), \text{ (Gaussian)} \quad (26)$$

where $\mathbf{y}$ are pixel coordinates, $d_{i,j}$ the Euclidean distance between pixels i and j, and l is the characteristic correlation length of the kernels (Rasmussen & Williams 2006). These two choices (also known as Ornstein–Uhlenbeck and squared exponential kernels, respectively) constitute two opposite extremes of the more general Matérn kernel (e.g. Mertens, Ghosh & Koopmans 2017), and have a single free parameter, l (which belongs to the $\mathbf{g}$ set of non-linear regularization parameters). This gives more freedom for additional structure in the covariance matrices C beyond the fixed-form derivative-based regularization. Also, these covariance kernels appear in better agreement with various source brightness profiles, as it will be shown by two examples later on (see also Vernardos, Tsagkatakis & Pantazis 2020). The variance level (i.e. the diagonal of the covariance matrix) is set by λ, and hence we assume here that the diagonal values of C are by definition equal to unity.

The potential perturbations given in equation (12) provide a measure of sub-galactic scale mass density fluctuations. The covariance matrix $C_{\delta\psi}$ is equivalent to the correlation function (or two-point correlation function), which is related to the power spectrum via the Fourier transform. Hence, measuring the covariance of $\delta\psi$ can probe the sub-galactic matter power spectrum. Although there have been theoretical and applied studies on this connection (Hezaveh et al. 2016a; Bayer et al. 2018; Chatterjee & Koopmans 2018; Diaz Rivero, Cyr-Racine & Dvorkin 2018), this work is the first consistent approach using the gravitational imaging technique. The derived power spectrum/covariance of $\delta\psi$ can then be associated to higher order moments in the lens mass distribution (e.g. Hsueh et al. 2017; Gilman et al. 2018) or dark matter substructure (e.g. Hezaveh et al. 2016a). For the latter, a more realistic approach to disentangle the effect of baryons would be to compare the observed sub-galactic scale perturbations to predictions from hydrodynamical simulations (e.g. Vogelsberger et al. 2014; Schaye et al. 2015).

2.6 Optimization

There are three main components in our approach to modelling gravitational lenses: the parametrized smooth lens potential, ψ, the grid-based potential perturbations, $\delta\psi$, and the grid-based source brightness, $\mathbf{s}$. The task is two-fold: (1) maximize the evidence for different choices of these components and compare them to find the combination that is most consistent with the data, and (2) within any given choice of model components, obtain the probability distributions of all the non-linear parameters in order to determine statistical uncertainties, systematic biases, and degeneracies between them. We also seek to evaluate how well our best $\mathbf{s}$ and $\delta\psi$ solutions, obtained at the most probable values of the non-linear parameters (the Maximum A Posteriori solutions, MAP), compare to the underlying truth. The linear part of the problem provides an exact solution for $\mathbf{s}$ and $\delta\psi$ – assuming that we already know $\mathbf{s}_p$ – that minimizes the penalty function and maximizes the posterior (equation 12), for fixed non-linear parameters. Here we describe our treatment of the non-linear parameters, namely, the smooth potential parameters $\boldsymbol{\eta}$, and the regularization parameters $\mathbf{g}$ and λ for the source and the potential perturbations.

First, we emphasize that the lens potential is dominated by the smooth model and any resulting perturbations are required to be small in order for equation (3) to be valid. This is also motivated by decent agreement between data and smooth models (e.g. Koopmans et al. 2009; Auger et al. 2010; Suyu et al. 2014; Oldham & Auger 2018), as well as evidence for lens perturbations (e.g. Vegetti et al. 2012; MacLeod et al. 2013; Nierenberg et al. 2014; Hezaveh et al. 2016b; Birrer & Amara 2017). Solving simultaneously for $\boldsymbol{\eta}$ and $\delta\psi$, however, is degenerate[5] and very inefficient; if $\boldsymbol{\eta}$ is far from the truth then $\delta\psi$ will try to make up for the correct sum of the smooth potential and the perturbations, leading away from a realistic solution. Hence, as a first step we optimize for the parameters $\boldsymbol{\eta}$ assuming $\delta\psi = 0$, while simultaneously solving the linear equations for the source (equation 9). The parameter space of $\boldsymbol{\eta}$ is explored using a nested-sampling approach (Skilling 2004), which provides several benefits: it computes the evidence term in equation (23) with the $\delta\psi = 0$ assumption, finds the most probable parameters, provides confidence intervals, and explores a large part of the parameter space

---

[4]Wayth et al. (2005) used maximum entropy regularization that has the benefit to prevent negative values for the source at the cost of not having quadratic penalty functions anymore. The solution minimizing the penalty function has to be found numerically at a higher computational cost.

[5]The $\delta\psi$ can in principle mimic almost any potential $\psi(\boldsymbol{\eta})$ and hence only the sum of the total potential is relevant. In practice however, $\psi(\boldsymbol{\eta})$ is set by general processes of galaxy formation and phase-space mixing and is expected to be smooth, while $\delta\psi$ describes any remaining structure such as sub-haloes, streams, etc.
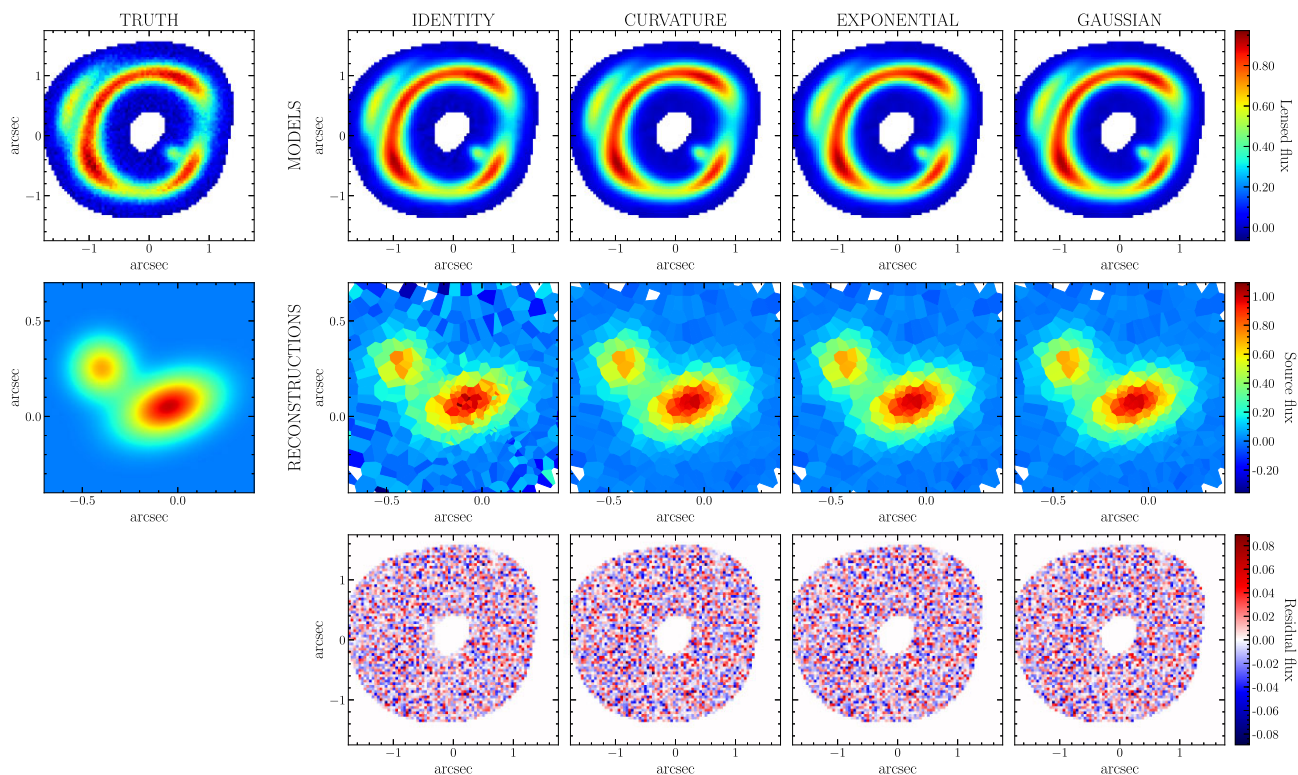
Figure 1. Lensed images (top), source (middle), and residuals (bottom) for the mock data (truth) and the reconstructions with different source regularization kernels: identity, curvature, exponential, and Gaussian. The source brightness, shown as Voronoi cells of an adaptive grid (see Section 2.1), has been reconstructed using $n = 3$. The corresponding parameters for the lens potential and the source regularization are shown in Table 1.

with a limited chance of getting stuck in local extrema. There is the additional option to start a Monte Carlo Markov Chain exploration of the parameter space near the most probable solution to obtain smoother posterior probability distributions.

Once the smooth model is determined, the parameters η are kept fixed to their maximum a posteriori values and solving for r (the potential perturbations and the source) is conducted. The varying non-linear parameters are now the regularization parameters g and λ, together describing the source and perturbation covariance matrices. Although it is possible to solve approximately for the λ parameters, at least in the case with $\delta\psi = 0$ (Koopmans 2005; Suyu et al. 2006), here we incorporate them in the full non-linear treatment. This allows one to infer confidence intervals and, most importantly, degeneracies between the source and the potential solutions.

The perturbations investigated here are assumed to be small and originate from an extended field of mass-density fluctuations permeating the lens, as opposed to specific and localized massive substructures, such as dark sub-haloes. For such prominent and confined perturbers, an iterative approach[6] would indeed be expected to perform better in locating and measuring the mass of putative massive substructures, carefully controlling the regularization parameters in the process (e.g. Suyu et al. 2006; Vegetti & Koopmans 2009; Nightingale & Dye 2015; Hezaveh et al. 2016b). In this work, however, we do not assume any restriction on the regularization parameters and solve for r for each set of sampled non-linear

parameters without updating the lens potential and the source. This approach is sufficient to capture the statistical properties of the perturbations field, provided that its amplitude is small. Mixing the two approaches, i.e. sampling the regularization parameters and then iterating up to a given number of steps for each combination, could be another possibility, especially when the extended field of perturbations also includes prominent mass concentrations such as subhaloes, but this is out of the scope of this paper.

3 RESULTS

In order to demonstrate the capabilities of our method, we examine modelling aspects of mock lenses combining smooth and complex lens potentials and source light profiles. In all cases, we use a point spread function (PSF) simulated for the *Hubble Space Telescope* (*HST*) using the TINY-TIM[7] software (Krist, Hook & Stoehr 2010), which is assumed to be the same in creating and modelling the mock data, uniform Gaussian random noise with a signal-to-noise ratio of ≈ 40 at peak brightness, and a mask to exclude regions of the image without lensing features (also the central part of the image that may hold residuals after removing the lens galaxy light, which we do not include or model). We generate the mocks using the MOLET[8] software package (Vernardos 2022).

The smooth parametric model used for the lens potential is a singular isothermal ellipsoid (SIE; Kassiola & Kovner 1993; Kormann, Schneider & Bartelmann 1994). We follow the notation

---

[6]At the end of each iteration, the lens potential is updated by adding the newly determined $\delta\psi$ and the $D_s(s_p)$ matrix in equation (5) is recalculated based on the derivatives of the newly determined source.

[7]http://www.stsci.edu/hst/observatory/focus/TinyTim
[8]https://github.com/gvernard/molet

of Schneider, Kochanek & Wambsganss ([2006](#)), with convergence given by:

$$\kappa(\omega) = \frac{b}{2\omega}, \tag{27}$$

where $\omega = \sqrt{q^2 x^2 + y^2}$, q is the minor to major axial ratio, and b (in arcsec) describes the overall potential strength.[9] This relation holds in the reference system whose x-axis is aligned with the ellipsoid's major axis, rotated by the position angle, θ, and whose origin coincides with the lens centre (x_0, y_0). External shear with magnitude γ and direction ϕ is included, leading to seven free parameters in total, hereafter denoted as $\boldsymbol{\eta}$. All angles are measured east-of-north, in order to remain consistent with the standard celestial definition.

3.1 Smooth lens and smooth source

A simulated lens system is created with a single massive lensing galaxy having $(b, q, \theta, x_0, y_0, \gamma, \phi) = (0.9, 0.8, -135°, 0, 0, 0.03, -40°)$. The source brightness distribution consists of two Gaussian components: the first is located at $x, y = (-0.05, 0.05)$ arcsec on the source plane, has an axial ratio of 0.6, position angle of $-70°$, and standard deviation on the x axis of $\sigma_x = 0.1$ arcsec, while the second component is at $x, y = (-0.4, 0.25)$ arcsec and has $\sigma_x = \sigma_y = 0.1$ arcsec (circular). The two components are scaled to have a peak brightness ratio of 0.7, with the first one being the brighter. The data are simulated on a square 3.5-arcsec 80-pixel field of view, having a pixel size somewhat bigger than 0.04 arcsec. The corresponding source and resulting lensed images are shown in the left-hand column in Fig. [1](#).

We model the system as a purely parametric smooth lens, without including any grid-based correction to the potential, using $n = 3$ for constructing the adaptive source plane grid (selecting 1 out of every $n \times n$ pixels). In addition to the lens potential parameters, the set of non-linear free parameters includes the regularization of the source, i.e. λ_s and $\boldsymbol{g}_s$. We use four different source regularization schemes with different associated parameters: identity (λ_s), curvature (λ_s), exponential kernel (λ_s, l_s), and Gaussian kernel (λ_s, l_s). The covariance between source pixels for the latter two schemes is given by equations ([25](#)) and ([26](#)) respectively; we note that the l_s are different parameters in these two cases, indicating the length where the correlation drops to roughly half its maximum. The value of the regularization parameter, λ_s, sets the overall level of regularization and is inversely proportional to the source variance, e.g. smaller values allow for more freedom in the source model. This parameter is expected to vary between different schemes because of the fundamentally different covariance matrices and cannot be straightforwardly compared. Instead, one can compare the evidence values to determine which choice of regularization is more justified by the data. We use the alternative curvature definition for adaptive grids provided in Vegetti & Koopmans ([2009](#)), which has a fixed regularization pattern/correlation length for a given grid. In this case, if H is a matrix holding the numerical coefficients for the local curvature of the source then $C_s = (H^T H)^{-1}$.

Fig. [1](#) shows the reconstructed sources, lensed images, and residuals. To further examine the model residuals and compare them to the noise, we compute their Fourier power spectrum that
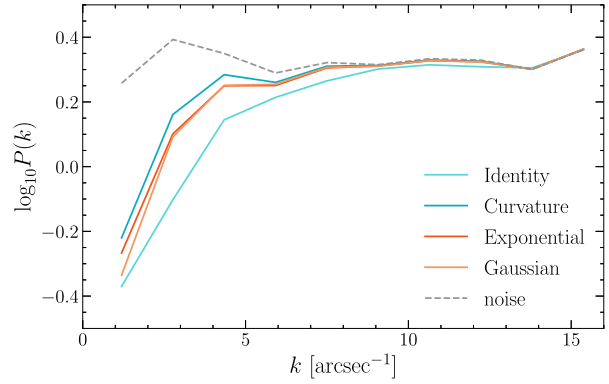

Figure 2. Fourier power spectrum of the model residuals shown in the bottom row of Fig. [1](#).

we show in Fig. [2](#). In such plots, we ideally want the residual power spectrum to be flat, i.e. matching the noise. Any deviations above (more power) or below (less power) a flat line introduce more correlations and overfitting (absorbing the noise), respectively, both indicating shortcomings of the employed model on the corresponding k scales. Table [1](#) (top) lists the Maximum A Posteriori (MAP) model parameters and the corresponding posterior probability terms from equation ([22](#)), for the four different regularization schemes, while Table [2](#) lists the mean parameter values, the 68 per cent confidence intervals, and the evidence, $\mathcal{E}$ in equation ([23](#)), for each model. Taking these results into account, we can see that the identity regularization allows the solution to vary wildly, similarly to white noise, resulting in an unrealistically grainy source. This is to be expected as the corresponding prior covariance matrix is the identity matrix, which has a flat power spectrum[10] and loosely constrains values of neighbouring pixels, allowing them to absorb more of the noise and lead to overfitting. Despite having the lowest likelihood (i.e. χ^2 term in Table [1](#)), and thus the lowest residuals[11] as shown in Fig. [2](#), the identity regularization also has the lowest evidence value. All other three regularization schemes perform better in recovering the source and the model parameters and give considerably higher evidence values. However, the Gaussian kernel is decisively preferred over the curvature and exponential kernels, having a Bayes factor of $\log_{10} K = 7.98$ and 17.85, respectively (Jeffreys [1998](#), assuming all models have the same prior probability). Although this is not the best possible kernel, it is still a sufficient approximation to describe the source brightness (see fig. 3 of Vernardos et al. [2020](#)). As a final note, it can be seen that in all cases there is some overfitting, most prominently for the identity regularization, that suppresses the noise in the large scales ($k < 5$ in Fig. [2](#)). This can also be seen in the reconstructed sources in Fig. [1](#), where the adaptive grid voronoi cells become noisy and do not drop to zero as we move away from the brightest pixels.

The full non-linear parameter probability densities for the reconstruction with the Gaussian kernel are shown in Fig. [3](#). The Nested Sampling method (Skilling [2004](#)), whose MultiNest implementation (Feroz, Hobson & Bridges [2009](#)) is used here, is designed to compute the Bayesian evidence but can still sample the probability

---

[9]We set $b = \sqrt{q}\,\theta_{\rm Ein}$, where the Einstein radius, $\theta_{\rm Ein}$, is defined as the radius within which the integral of equation ([27](#)) becomes equal to unity.

[10]Or a delta function two-point correlation function, which is the inverse Fourier transform of the power spectrum (i.e. the Wiener–Khinchin theorem).
[11]These residuals result from $n = 3$ for the adaptive grid and are expected to be reduced by increasing the number of pixels used to describe the source, i.e. $n = 2$ or $n = 1$.

Table 1. Values of the lens potential (η) and source regularization parameters (λ_s, g_s) that maximize the posterior probability density, i.e. Maximum A Posteriori (MAP) values, and corresponding terms from equation (22). The smooth source (top part of the table) is described in Section 3.1 and the lensed images corresponding to the parameters listed here are shown in Fig. 1. Similarly, NGC 3982 and NGC 2623 (middle and bottom parts) are described in Section 3.2 and shown in Figs 4 and 5.

	Name	Units	Truth	Identity	Curvature	Exponential	Gaussian
Smooth source	b	arcsec	0.9	0.894	0.897	0.898	0.897
	q	–	0.8	0.790	0.795	0.796	0.795
	θ	°	−135	−136.208	−135.031	−135.176	−135.135
	x_0	arcsec	0	−0.022	−0.021	−0.021	−0.021
	y_0	arcsec	0	0.022	0.022	0.022	0.022
	γ	–	0.03	0.028	0.029	0.029	0.029
	ϕ	°	−40	−37.278	−39.436	−39.352	−39.340
	λ_s	–	–	7.945	0.121	29.379	86.581
	l_s	arcsec	–	–	–	0.675	0.128
			$-\frac{N_d}{2}\log(2\pi)^\dagger$	−3536.08	−3536.08	−3536.08	−3536.08
			$\frac{N_s}{2}\log(\lambda_s)$	755.44	−769.81	1232.11	1626.06
			$-\frac{1}{2}\log(\det C_d)^\dagger$	24020.04	24020.04	24020.04	24020.04
			$-\frac{1}{2}\log(\det C_s)$	0	3175.55	907.62	583.66
			$-\frac{1}{2}\chi^2$	−1648.13	−1797.81	−1765.54	−1765.10
			$-\frac{1}{2}\lambda_s s^T C_s^{-1} s$	−257.99	−109.26	−133.33	−127.33
			$-\frac{1}{2}\log(\det H)$	−1916.13	−2757.57	−2525.20	−2556.19
			$\log P$	17417.16	18225.07	18199.64	18245.07
NGC 3982	b	arcsec	0.9	0.896	0.891	0.895	0.895
	q	–	0.8	0.793	0.785	0.791	0.791
	θ	°	−135	−134.743	−133.491	−133.837	−134.473
	x_0	arcsec	0	−0.023	−0.021	−0.022	−0.022
	y_0	arcsec	0	0.023	0.025	0.024	0.024
	γ	–	0.03	0.029	0.027	0.028	0.029
	ϕ	°	−40	−40.203	−41.950	−41.619	−40.533
	λ_s	–	–	16.172	0.126	69.969	56.937
	l_s	arcsec	–	–	–	0.385	0.194
			$-\frac{N_d}{2}\log(2\pi)^\dagger$	−4307.06	−4307.06	−4307.06	−4307.06
			$\frac{N_s}{2}\log(\lambda_s)$	1014.51	−755.05	1548.42	1473.29
			$-\frac{1}{2}\log(\det C_d)^\dagger$	25672.50	25672.50	25672.50	25672.50
			$-\frac{1}{2}\log(\det C_s)$	0	3181.83	711.78	684.29
			$-\frac{1}{2}\chi^2$	−2083.00	−2383.13	−2271.33	−2207.01
			$-\frac{1}{2}\lambda_s s^T C_s^{-1} s$	−291.20	−151.58	−172.64	−194.16
			$-\frac{1}{2}\log(\det H)$	−2335.42	−2929.98	−2766.41	−2685.90
			$\log P$	17670.32	18327.53	18415.24	18435.95
NGC 2623	b	arcsec	0.9	0.900	0.903	0.901	0.901
	q	–	0.8	0.802	0.808	0.804	0.803
	θ	°	−135	−135.151	−132.769	−134.165	−134.464
	x_0	arcsec	0	−0.020	−0.018	−0.020	−0.020
	y_0	arcsec	0	0.020	0.019	0.021	0.021
	γ	–	0.03	0.033	0.033	0.033	0.033
	ϕ	°	−40	−39.932	−44.236	−41.881	−41.180
	λ_s	–	–	25.623	0.016	64.835	37.979
	l_s	arcsec	–	–	–	0.078	0.065
			$-\frac{N_d}{2}\log(2\pi)^\dagger$	−3145.53	−3145.53	−3145.53	−3145.53
			$\frac{N_s}{2}\log(\lambda_s)$	1182.25	−1507.27	1520.64	1325.70
			$-\frac{1}{2}\log(\det C_d)^\dagger$	25983.49	25983.49	25983.49	25983.49
			$-\frac{1}{2}\log(\det C_s)$	0	3179.82	223.29	328.69
			$-\frac{1}{2}\chi^2$	−1705.31	−1846.70	−1735.70	−1762.97
			$-\frac{1}{2}\lambda_s s^T C_s^{-1} s$	−213.42	−162.16	−169.34	−175.53
			$-\frac{1}{2}\log(\det H)$	−2063.98	−2292.50	−2309.65	−2250.27
			$\log P$	20037.50	20209.16	20367.20	20303.59

Note. †constant

distributions at their peak almost as well as an MCMC algorithm. However, if such a method is chosen from start it would neither guarantee convergence to the global maximum nor compute the evidence (or be extremely inefficient in doing so). The recovered probability distributions for the lens model parameters b, q, θ, γ,

and ϕ contain the true values within confidence intervals of 1 to 2σ. The lens centre is systematically offset by approx. half a pixel in the negative x and positive y directions, which is due to a corresponding shift in the PSF's brightest pixel. There are no degeneracies observed between the parameters, other than the expected $b - q$ correlation

Table 2. Mean values and 68 per cent confidence intervals for the lens potential (η) and source regularization parameters (λ_s, g_s), and corresponding evidence values, $\mathcal{E}$ in equation (23). The smooth source (top part of the table) is described in Section 3.1, while NGC 3982 and NGC 2623 (middle and bottom parts) are described in Section 3.2. The lens centre appears shifted by about half a pixel in the x and y directions due to a corresponding shift in the PSF. The full probability densities for the Gaussian regularization model of the smooth source (top part of the table) are shown in Fig. 3.

	Name	Units	Truth	Identity	Curvature	Exponential	Gaussian
Smooth source	b	arcsec	0.9	$0.894^{+0.001}_{-0.001}$	$0.897^{+0.002}_{-0.003}$	$0.898^{+0.002}_{-0.003}$	$0.897^{+0.002}_{-0.002}$
	q	–	0.8	$0.790^{+0.002}_{-0.002}$	$0.795^{+0.004}_{-0.005}$	$0.796^{+0.003}_{-0.005}$	$0.795^{+0.003}_{-0.004}$
	θ	°	−135	$-136.208^{+0.484}_{-0.750}$	$-135.031^{+0.514}_{-0.504}$	$-135.176^{+0.591}_{-0.598}$	$-135.135^{+0.556}_{-0.633}$
	x_0	arcsec	0	$-0.022^{+0.001}_{-0.001}$	$-0.021^{+0.001}_{-0.001}$	$-0.021^{+0.001}_{-0.001}$	$-0.021^{+0.001}_{-0.001}$
	y_0	arcsec	0	$0.022^{+0.001}_{-0.001}$	$0.022^{+0.001}_{-0.001}$	$0.022^{+0.001}_{-0.001}$	$0.022^{+0.001}_{-0.001}$
	γ	–	0.03	$0.028^{+0.001}_{-0.001}$	$0.029^{+0.001}_{-0.001}$	$0.029^{+0.001}_{-0.001}$	$0.029^{+0.001}_{-0.001}$
	ϕ	°	−40	$-37.278^{+1.480}_{-0.995}$	$-39.436^{+1.101}_{-0.993}$	$-39.352^{+1.171}_{-1.238}$	$-39.340^{+1.183}_{-1.184}$
	λ_s	–	–	$7.958^{+0.801}_{-0.753}$	$0.122^{+0.005}_{-0.042}$	$30.355^{+5.243}_{-9.475}$	$86.992^{+8.063}_{-9.020}$
	l_s	arcsec	–	–	–	$0.695^{+0.169}_{-0.182}$	$0.129^{+0.004}_{-0.004}$
	log $\mathcal{E}$			17388.46 ± 0.58	18190.35 ± 0.55	18167.62 ± 0.55	18208.74 ± 0.57
NGC 3982	b	arcsec	0.9	$0.896^{+0.001}_{-0.004}$	$0.891^{+0.002}_{-0.002}$	$0.895^{+0.004}_{-0.004}$	$0.895^{+0.005}_{-0.005}$
	q	–	0.8	$0.793^{+0.001}_{-0.008}$	$0.785^{+0.003}_{-0.003}$	$0.791^{+0.007}_{-0.007}$	$0.791^{+0.008}_{-0.009}$
	θ	°	−135	$-134.743^{+0.499}_{-0.445}$	$-133.491^{+0.425}_{-0.443}$	$-133.837^{+0.709}_{-0.758}$	$-134.473^{+0.599}_{-0.653}$
	x_0	arcsec	0	$-0.023^{+0.001}_{-0.001}$	$-0.021^{+0.001}_{-0.001}$	$-0.022^{+0.001}_{-0.001}$	$-0.022^{+0.001}_{-0.001}$
	y_0	arcsec	0	$0.023^{+0.001}_{-0.001}$	$0.025^{+0.001}_{-0.001}$	$0.024^{+0.001}_{-0.001}$	$0.024^{+0.001}_{-0.001}$
	γ	–	0.03	$0.029^{+0.000}_{-0.002}$	$0.027^{+0.001}_{-0.001}$	$0.028^{+0.002}_{-0.002}$	$0.029^{+0.002}_{-0.002}$
	ϕ	°	−40	$-40.203^{+0.935}_{-0.942}$	$-41.950^{+1.134}_{-1.089}$	$-41.619^{+1.544}_{-1.256}$	$-40.533^{+1.435}_{-1.007}$
	λ_s	–	–	$16.197^{+0.871}_{-0.881}$	$0.126^{+0.005}_{-0.046}$	$73.260^{+21.781}_{-23.072}$	$57.157^{+4.580}_{-5.430}$
	l_s	arcsec	–	–	–	$0.408^{+0.070}_{-0.167}$	$0.194^{+0.008}_{-0.007}$
	log $\mathcal{E}$			17635.43 ± 0.57	18296.66 ± 0.56	18382.97 ± 0.56	18407.46 ± 0.58
NGC 2923	b	arcsec	0.9	$0.900^{+0.001}_{-0.001}$	$0.903^{+0.002}_{-0.002}$	$0.901^{+0.001}_{-0.001}$	$0.901^{+0.001}_{-0.001}$
	q	–	0.8	$0.802^{+0.002}_{-0.002}$	$0.808^{+0.004}_{-0.004}$	$0.804^{+0.002}_{-0.002}$	$0.803^{+0.002}_{-0.002}$
	θ	°	−135	$-135.151^{+0.302}_{-0.376}$	$-132.769^{+0.439}_{-0.521}$	$-134.165^{+0.516}_{-0.747}$	$-134.464^{+0.452}_{-0.614}$
	x_0	arcsec	0	$-0.020^{+0.001}_{-0.001}$	$-0.018^{+0.001}_{-0.000}$	$-0.020^{+0.001}_{-0.001}$	$-0.020^{+0.001}_{-0.001}$
	y_0	arcsec	0	$0.020^{+0.001}_{-0.001}$	$0.019^{+0.001}_{-0.001}$	$0.021^{+0.001}_{-0.001}$	$0.021^{+0.001}_{-0.001}$
	γ	–	0.03	$0.033^{+0.001}_{-0.001}$	$0.033^{+0.001}_{-0.001}$	$0.033^{+0.001}_{-0.001}$	$0.033^{+0.001}_{-0.001}$
	ϕ	°	−40	$-39.932^{+0.747}_{-0.532}$	$-44.236^{+0.953}_{-0.771}$	$-41.881^{+1.327}_{-0.906}$	$-41.180^{+1.096}_{-0.804}$
	λ_s	–	–	$25.678^{+1.626}_{-1.717}$	$0.016^{+0.001}_{-0.005}$	$65.296^{+7.285}_{-7.873}$	$38.388^{+5.278}_{-5.801}$
	l_s	arcsec	–	–	–	$0.079^{+0.009}_{-0.013}$	$0.065^{+0.005}_{-0.006}$
	log $\mathcal{E}$			20028.44 ± 0.59	20221.07 ± 0.58	20330.95 ± 0.59	20296.56 ± 0.59

from equation (27) and those between $b - \gamma$, $q - \gamma$, and $\theta - \phi$, which reflect the known degeneracy between the strength and orientation of the SIE and the external shear (e.g. see part 2 of Schneider et al. 2006). The joint probability distribution of λ_s and l_s allows for useful conclusions on the behaviour of the source regularization. Here, there is a very weak anticorrelation between λ_s and l_s, which is somewhat expected: increasing the overall regularization parameter λ_s smooths out the reconstructed source, as does increasing the correlation length l_s in the covariance kernel. This anticorrelation will become more prominent in the following, but it is worth pointing it out already at this smooth example. Such information will be increasingly helpful in quantifying the degree of degeneracy between more complex sources and perturbed lens potentials in subsequent examples.

3.2 Smooth lens and complex source

Setting the lens potential to the same smooth parametric model as before, we now change the source brightness profile to more realistic ones taken from observed galaxies. We use high resolution *HST* archival observations of NGC 3982 (a spiral galaxy) and NGC 2623 (a merger) taken with the ACS instrument, selected to represent a wider range of possible strongly lensed sources. We scale the source angular size arbitrarily to around 1 arcsec, roughly the same as for the analytic source used in the previous section. The *HST* images are scaled down dramatically in size and are significantly oversampled compared to the sampling of the final mock data. We take this subpixel structure into account by heavily oversampling the mock data by a factor of 10, producing very high resolution lensed images, applying an oversampled PSF, and finally averaging to the
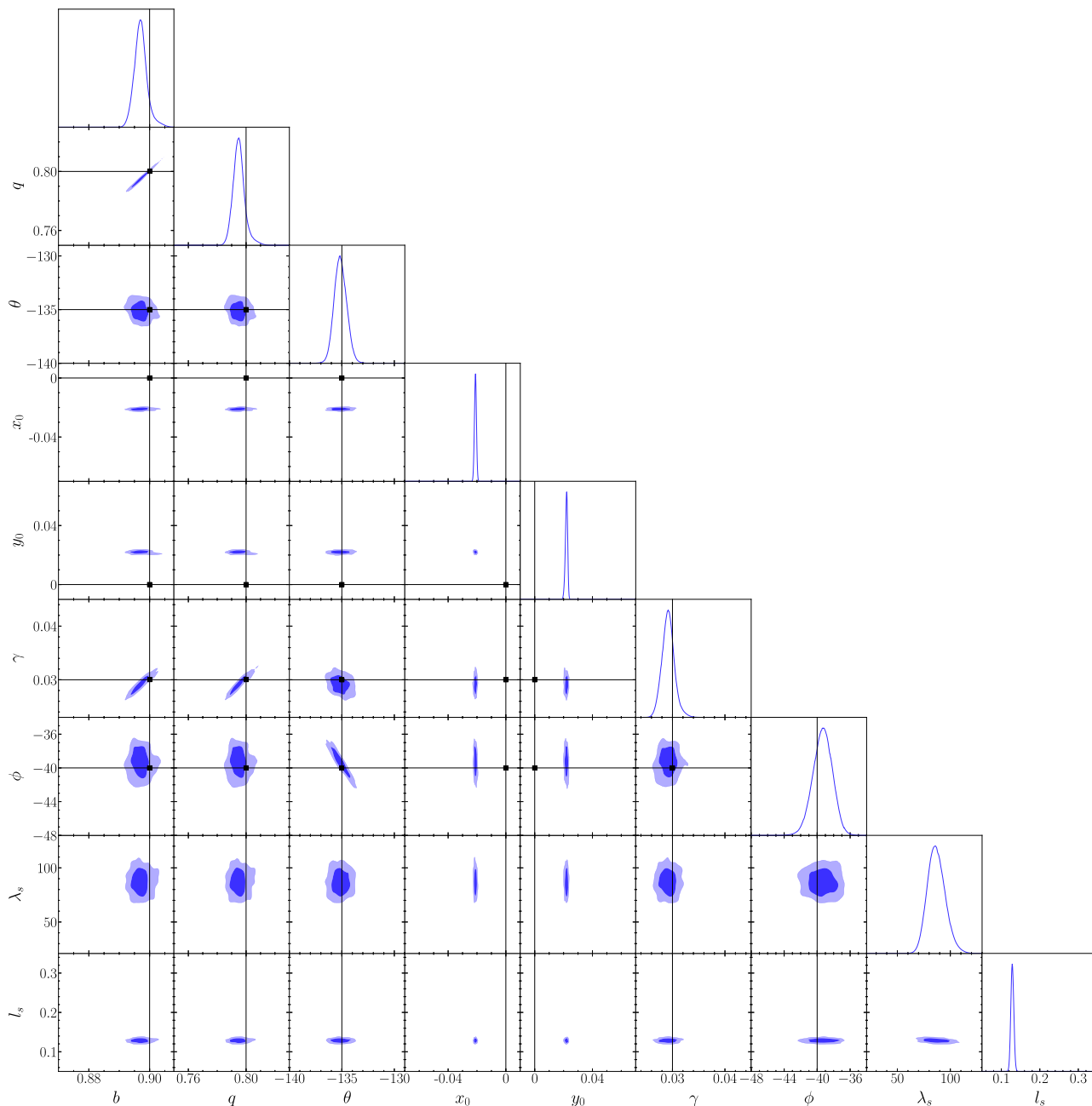
Figure 3. Marginalized probability densities and histograms for the lens potential (η) and regularization (λ_s, g_s) parameters for the Gaussian kernel reconstruction of the smooth source described in Section 3.1. The parameter ranges are set to match Fig. 12 and facilitate comparisons with the results described in Sections 3.3 and 3.4 – a zoomed-in version of this plot that shows the shape of the 2D distributions better is shown in Fig. A1. The true values of the smooth potential parameters (η) are indicated by the vertical and horizontal black lines and the points (squares). Contours are drawn at the 68 and 95 per cent confidence intervals. The corresponding mean values and 68 per cent confidence intervals are listed in Table 2.

final pixel scale: the same square 3.5-arcsec 80-pixel field of view as before. The resulting mock lensed images are shown in Figs 4 and 5.

Fig. 6 shows the two-point correlation function[12] of the *HST* observations, as well as visual fits of the exponential and Gaussian kernels

---

[12]We compute the two-point correlation function from the non-zero covariance matrix elements and the distance between the corresponding pixels. The covariance matrices can be constructed directly from the images, i.e. by calculating the correlation coefficients between all the pixels, or by the kernels discussed in Section 2.5.

and the corresponding best-fitting correlation length parameters (top panel). It can be clearly seen that the true underlying covariance properties of these two objects can, in principle, be captured very well by each regularization scheme. Using these schemes in solving equation (9) imposes a realistic prior on the reconstructed source that is motivated by real observations, as opposed to, for example, curvature regularization, which implicitly imposes a correlation that is unlikely to match the truth.

We model the two systems exactly in the same way as in the previous section, i.e. using $n = 3$ and the same four regularization schemes. The reconstructed sources, lensed images, and residuals
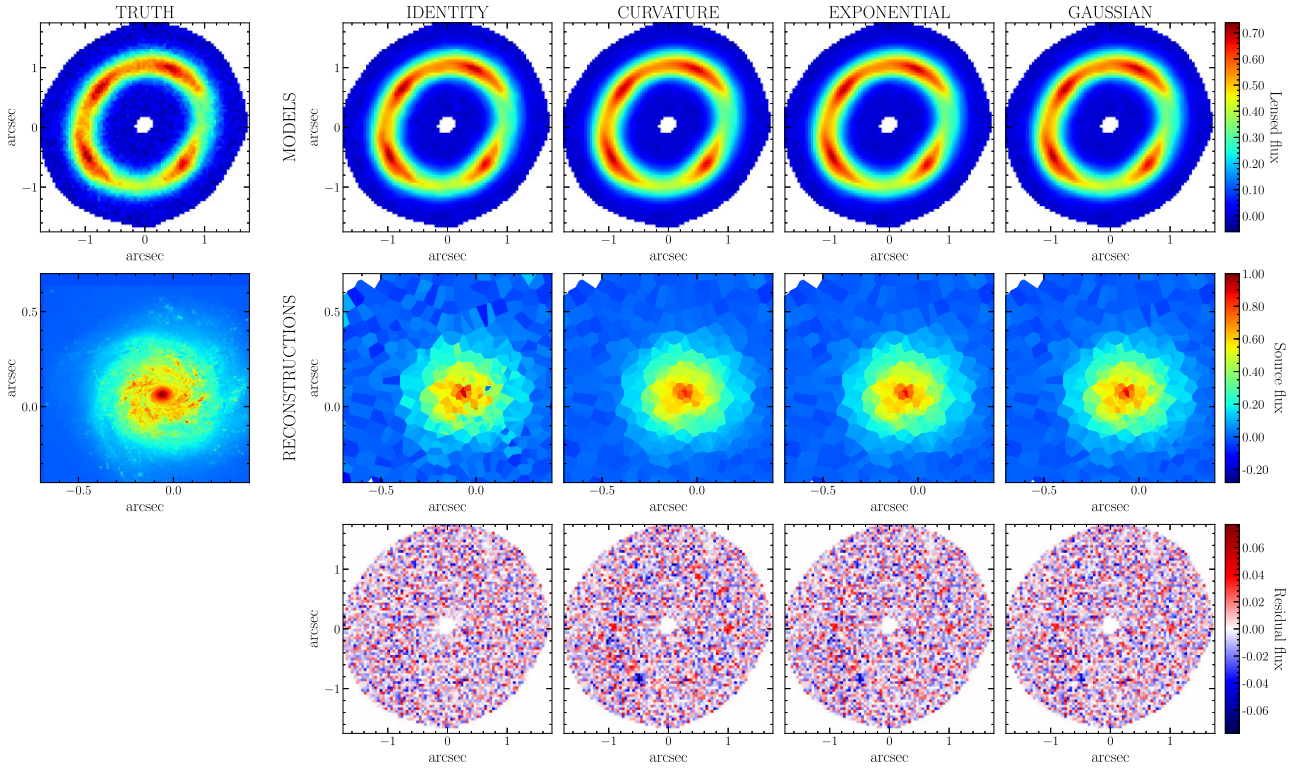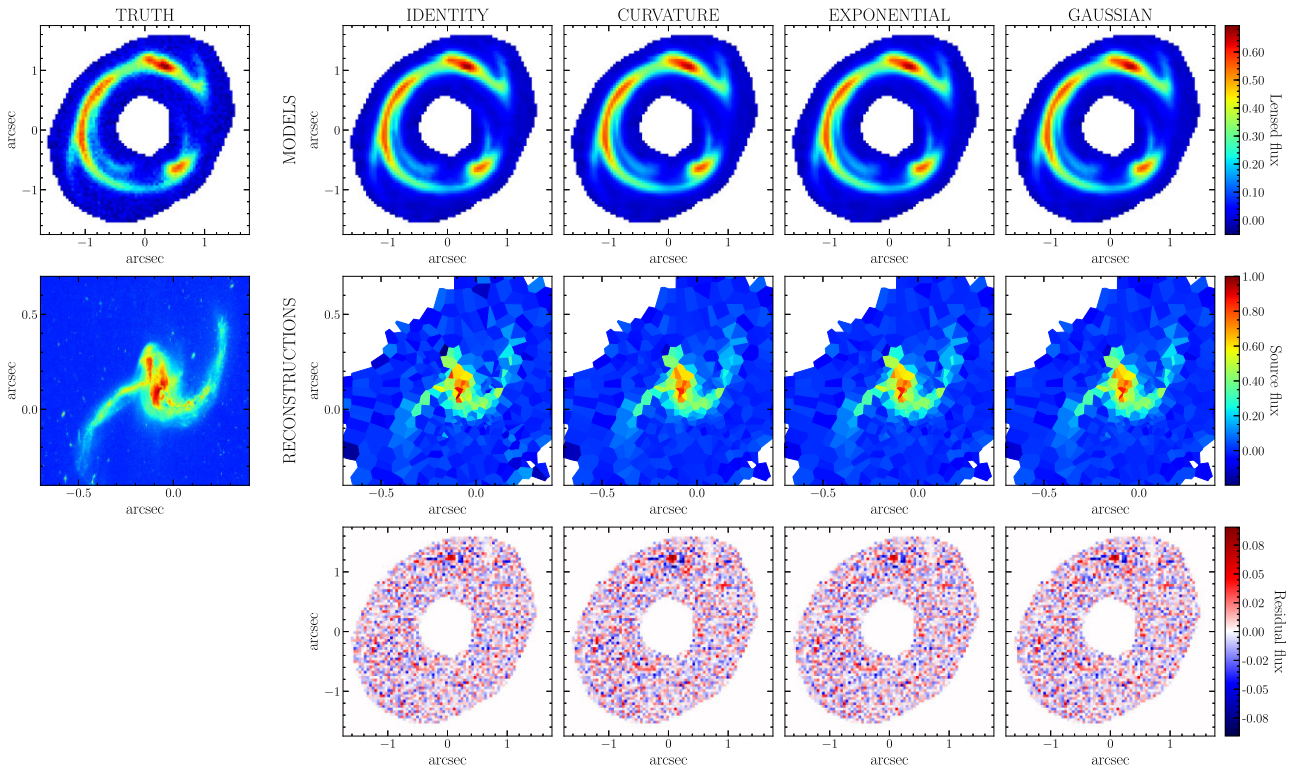
Figure 4. Same as Fig. 1 for NGC 3982.

Figure 5. Same as Fig. 1 for NGC 2623.

are shown in Figs 4 and 5, while the MAP and mean parameters and evidence terms are listed in Tables 1 and 2. In Fig. 6 we compare the radially averaged two-point correlation functions of the unlensed (*HST*-observed) sources and their reconstructions with the

priors imposed by the covariance matrix C_s. Correlations imposed by curvature regularization have a fixed length (no free parameters) and are quite different from the truth: pixels that are far from each other are much more correlated than, for example, in the case of
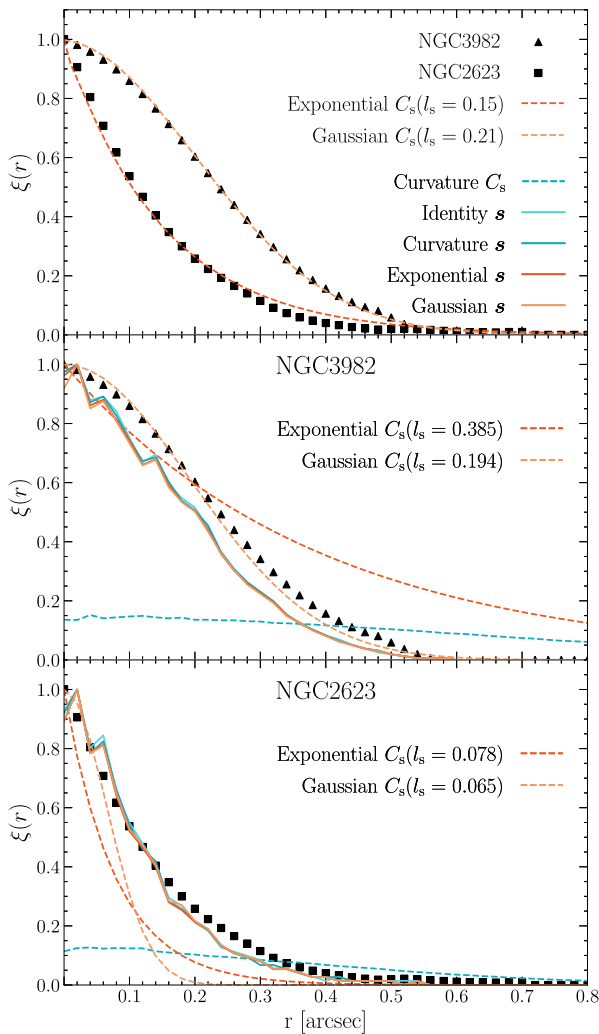
Figure 6. Radially averaged two-point correlation functions, defined directly from the covariance matrix and the distance between the pixels, for the unlensed (*HST*-observed) images of NGC 3982 and NGC 2623 (triangles and squares, respectively) and their corresponding reconstructions (solid lines) and priors (dashed lines). The l_s parameter for the exponential and Gaussian priors changes in each panel as indicated. We have assumed the angular size of the unlensed sources to be ≈ 1 arcsec, therefore the values on the horizontal axis are scaled accordingly. The functions have been normalized to unity to factor out the effect of λ_s and the pixel resolution. Top: the exponential and Gaussian theoretical covariance kernels from equations (25) and (26) are shown for values of l_s selected to visually match the data. Middle and bottom: the l_s parameters for the exponential and Gaussian covariance kernel priors are set to their MAP values (see Table 1). The two-point correlation function of an identity regularization prior would be a delta function centred at zero.

an exponential kernel, reflecting the implicit smoothness prior. This is a direct consequence of C_s being a quite dense matrix: if H is a matrix holding the numerical coefficients for the local curvature of the source, then $C_s = (H^T H)^{-1}$, and although H, H^T, and $H^T H$ are relatively sparse matrices, $(H^T H)^{-1}$ is not. However, the quality of the data is high enough to drive the solution close to the truth regardless of the regularization scheme/assumed prior – the two-point correlation functions for all the reconstructions lie on top of each other[13] in Fig. 6. Even the reconstruction using the least physically motivated
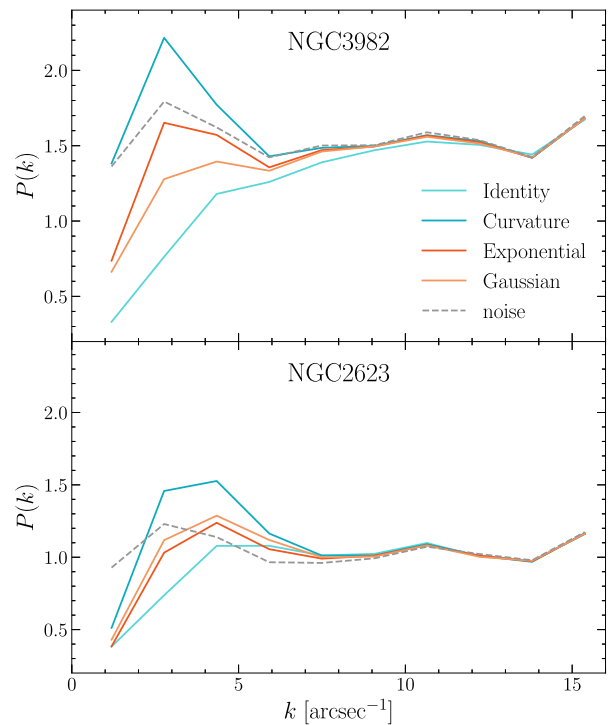
Figure 7. Fourier power spectra of the model residuals shown in the bottom rows of Figs 4 (top panel) and 5 (bottom panel).

identity regularization manages to recover the correct correlations of the source, suggesting that the solution is driven by the data and not the prior and therefore is not very degenerate. Nevertheless, the evidence values (see Table 2) are maximized by the correct regularization scheme in each case, viz. Gaussian for NGC 3982 and exponential for NGC 2623. Comparing the mean values and confidence intervals of the correlation length parameter, l_s, to the truth, i.e. those obtained from the observed images (see Fig. 6), we find a good agreement for both cases, despite the MAP value for NGC 2623 being quite low (see Fig. 6).

Fig. 7 shows the Fourier power spectra of the model residuals for the two systems, shown in the bottom rows of Figs 4 and 5. Comparing these power spectra we see that the identity regularization leads to overfitting, as is the case for the smooth source examined in Section 3.1. Curvature regularization produces residuals on the large scales (small wavenumber k), while the more physically motivated exponential and Gaussian regularizations result residuals closest to the noise and at the same time avoid overfitting. Despite the successful modelling of the smooth lens potential and finding the correct source prior, there is still some unmodelled flux in the residuals (at SSE in the residuals in Fig. 4 and N in Fig. 5), which results from using $n = 3$ to construct the adaptive grid, a value too high to account for the complex small-scale source structure. Such residuals could erroneously be interpreted as spurious lens potential perturbations when modelling real data – this is examined more closely in Section 3.4.

3.3 Modelling potential perturbations

A lens potential fully described by a parametrized smooth lens model, as examined so far, might be an idealized scenario. Therefore, in this

---

[13] The reconstructions become completely smooth and match almost perfectly the truth and the recovered covariance matrix if the reconstructed sources are

interpolated from the adaptive Delaunay grid on to a regular grid with similar resolution.
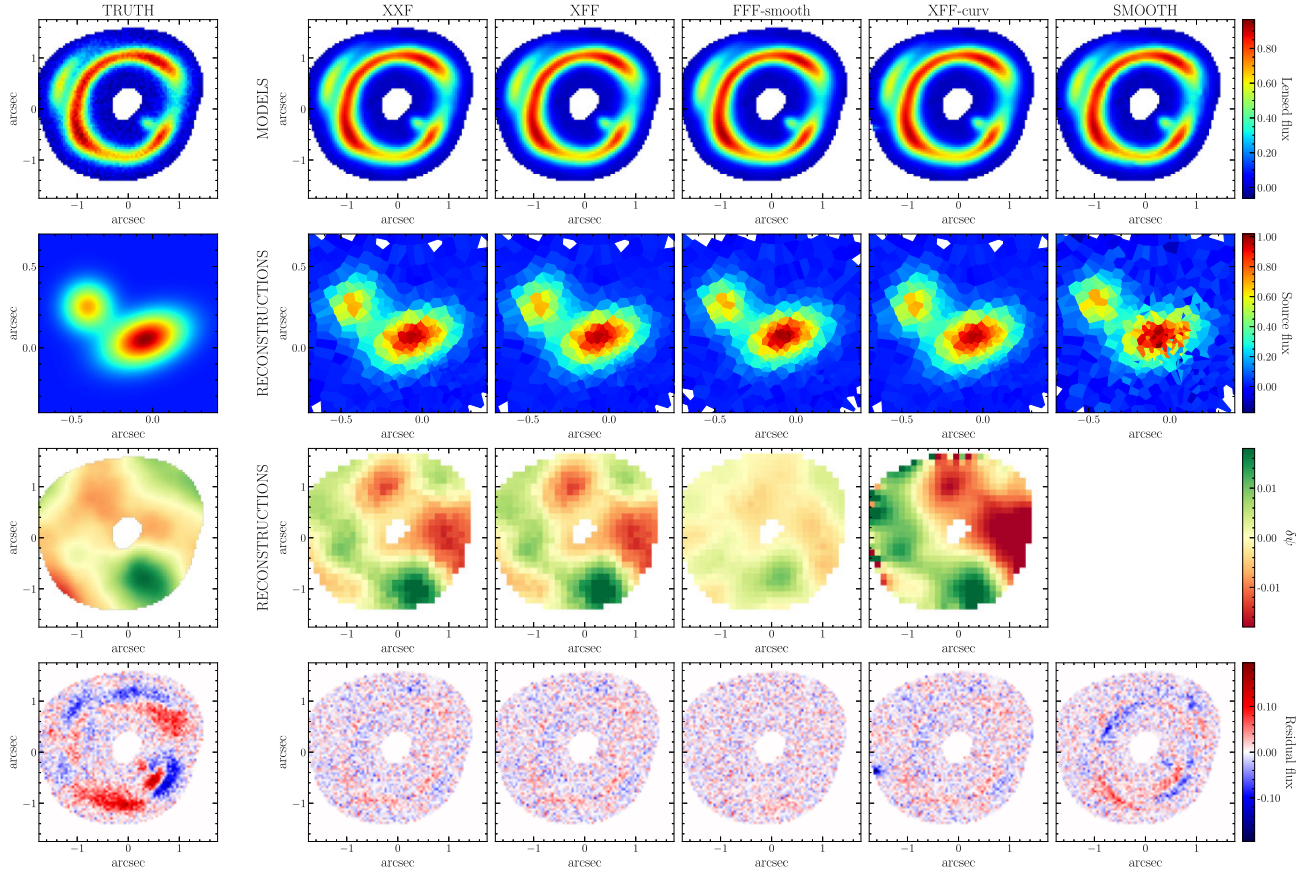
Figure 8. Same as Fig. 1, with the addition of the true and reconstructed perturbations $\delta\psi$ (third row, in the same dimensionless units as the SIE lens potential corresponding to equation 27). The different models shown in each column are described in Section 3.3. The bottom left-hand panel shows the difference between the perturbed (top left-hand panel) and unperturbed systems (top left-hand panel of Fig. 1).

section we introduce and model potential perturbations. We adopt the same smooth lens potential used in Sections 3.1 and 3.2, which we perturb using a Gaussian Random Field (GRF) of perturbations $\delta\psi$. GRF perturbations are defined by their power spectrum, which, in this case, we assume to be a power law:

$$P(k) = A\,k^\beta, \tag{28}$$

where A is the amplitude, associated to the variance of the zero-mean $\delta\psi$ field (for more details see Bayer et al. 2018; Chatterjee & Koopmans 2018; Chatterjee 2019), β is the slope, and k is the wavenumber of the Fourier harmonics. Regardless of our particular choice of GRF perturbations, the generality of the analysis presented here is not affected – in fact, any form of potential perturbations could be used and modelled.

We generate a single realization of $\delta\psi$ from a GRF having $\log_{10}(A)$, $\beta = (-7.8, -5.5)$, in the same 80×80 pixel grid as the mock image. Within the masked region of the field of view, the GRF field has slightly different A and β parameters (see Table 5 and Fig. 9). The resulting perturbations vary in magnitude between roughly ± 13 per cent of the average smooth lens potential (within the mask). The source (the same as the one used in Section 3.1), the perturbations, and the corresponding lensed image are shown in the left-hand column in Fig. 8. The difference[14] between the mock data with

the purely smooth underlying lens model used in Section 3.1 (top left-hand panel in Fig. 1) and its perturbed version used here (top left-hand panel in Fig. 8) is shown in Fig. 8, bottom left-hand panel.

An important and basic observation we need to make here is that in order to be able to reconstruct any perturbing $\delta\psi$, there needs to be some lensed light locally around it. This can be understood by examining matrix M_r (equation 5), which extends the smooth lens modelling framework presented in Section 2 to include potential perturbations: if there is no source light (strictly speaking, if the source light is constant, i.e. its derivative is zero) then the terms $D_s(s_p)$ introduced in equation (3), and consequently the entire perturbing part of M_r, vanish. The $\delta\psi$ are then reconstructed based mainly on the regularization prior. As a result, in general, the further a reconstructed $\delta\psi$ value is from pixels with some lensed light in them the less accurate its estimate based on the data becomes. In the following, we do not attempt to mitigate this and our reconstructed $\delta\psi$ away from pixels with brightness should be viewed as an extrapolation regularized by the prior. A similar argument holds for the smooth potential as well.

The covariance matrix of a GRF field is derived from its two-point correlation function, which is simply the inverse Fourier transform of its power spectrum. For a GRF with a power-law power spectrum, like the one given in equation (28), the two-point correlation function is:

$$\xi(r) = 2\pi A\,J_0(k_{max}\,r)\,k_{max}^{\beta+2}, \tag{29}$$

where J_0 is the zeroth-order Bessel function of the first kind, and k_{max} the maximum wavenumber. However, the mask truncates the GRF and changes its covariance properties so that the above relation

[14]We first subtract the perturbed and unperturbed mock lens images without any noise, and then add an artificial white noise realization with the same signal-to-noise ratio as the unperturbed case.
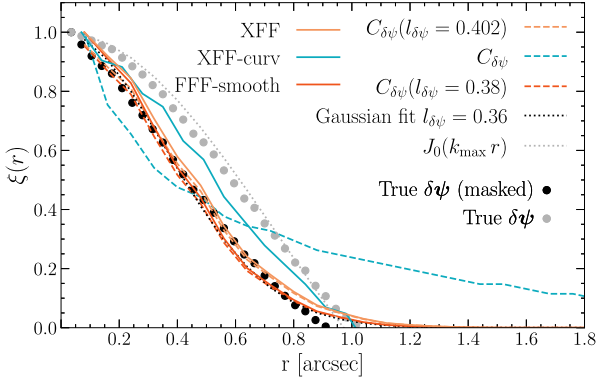
Figure 9. Radially averaged two-point correlation functions of the true $\delta\psi$ field (circles), the reconstructions from the models shown in Fig. 8 (solid lines, see Section 3.3 for details), and different $C_{\delta\psi}$ priors (dashed lines). The priors for the XFF and FFF-smooth models are Gaussian with the $l_{\delta\psi}$ parameter MAP value indicated in the parentheses (see Table 3). The black dotted line is a Gaussian fit to the correlation function of the masked $\delta\psi$ with $l_{\delta\psi} = 0.36$ (using equation 26). The grey dotted line is directly plotted from equation (29), i.e. not a fit, with k_{max} set to the diagonal of the 3.5 arcsec field of view.

cannot be used to construct a regularization kernel anymore. In this case, the Gaussian kernel provides a sufficiently good approximation for the two-point correlation function, as shown in Fig. 9.

To model the perturbed system, we use a Gaussian regularization kernel for both s and $\delta\psi$, and $n = 3$ for reconstructing the adaptive source grid. The size of the pixel grid to reconstruct the perturbations $\delta\psi$ on and n set the number of free parameters of any model and can be selected by maximizing the Bayesian evidence (Vegetti et al. 2012). However, this is outside the scope of this work – and a computationally very demanding task. We use a 30×30 pixel grid for $\delta\psi$, which has enough resolution to capture the details of the true underlying GRF perturbations while still leading to tractable computations (such a grid has been also used in the case of a single

perturbing substructure, e.g. Koopmans 2005; Vegetti et al. 2012). For each of the models presented in Fig. 13 we either fix (X) or set free (F) each of the three parameter sets η, (λ_s, l_s), $(\lambda_{\delta\psi}, l_{\delta\psi})$ and name it accordingly, e.g. model XFX has only (λ_s, l_s) free to vary. We model the perturbed lens in three different set-ups: (i) we fix the smooth lens model to the truth and the source regularization parameters to the mean values of the Gaussian kernel model obtained in Section 3.1 (see Table 2) and we sample only $\lambda_{\delta\psi}$, $\boldsymbol{g}_{\delta\psi}$ (model XXF), (ii) we fix the smooth lens model to the truth and sample both λ_s, $\boldsymbol{g}_s$ and $\lambda_{\delta\psi}$, $\boldsymbol{g}_{\delta\psi}$ (model XFF), and (iii) we sample η, λ_s, $\boldsymbol{g}_s$, $\lambda_{\delta\psi}$, and $\boldsymbol{g}_{\delta\psi}$ simultaneously (model FFF-smooth, not to confuse with the FFF-complex model presented in the following section). Fig. 8 shows the resulting lensed images, reconstructed s and $\delta\psi$, and residuals, Table 3 lists the MAP model parameters and the posterior probability terms from equation (22), and Table 4 lists the mean parameter values, their 68 per cent confidence intervals, and the evidence for each set-up. Models XXF and XFF give almost identical results. Models XFF and FFF-smooth recover a similar correlation length for the source, in very good agreement with the unperturbed case presented in Section 3.1 – this is also true for the parameters η recovered by the FFF-smooth model. The correlation length of the perturbations, $l_{\delta\psi}$, has a very similar value for all the models; the values from XXF and XFF and the corresponding covariance matrices, $C_{\delta\psi}$, are in fact so close that their determinants differ by very little (see Table 3).

To further investigate the effect of the prior on the lens potential perturbations, we evaluate a model using curvature regularization for $\delta\psi$. To do this, we fix η to their true values and sample λ_s, $\boldsymbol{g}_s$, and $\lambda_{\delta\psi}$ (the XFF-curv model – there are no $\boldsymbol{g}_{\delta\psi}$ parameters in this case). First, we notice that the values of λ_s and l_s are almost identical with the XFF model, however, the evidence has a much lower value, despite the latter model having an additional free parameter. In Fig. 9, we show the two-point correlation function from this model and compare it with the one from the true underlying $\delta\psi$ field and the reconstructions from the XFF and FFF-smooth models. It is evident that in this case the data and not the prior is driving the $\delta\psi$ reconstruction. In

Table 3. MAP parameter values and corresponding probability terms (from equation 22, same as Table 1). Models XXF, XFF, XFF-curv, and FFF-smooth are described in Section 3.3 and model FFF-complex in 3.4. Notice that the dimensions of the parameter space are not the same between the models.

Name	Units	Truth	XXF	XFF	XFF-curv	FFF-smooth	FFF-complex
b	arcsec	0.9	–	–	–	0.895	0.881
q	–	0.8	–	–	–	0.799	0.772
θ	$^\circ$	-135	–	–	–	-134.351	-133.769
x_0	arcsec	0	–	–	–	-0.044	-0.054
y_0	arcsec	0	–	–	–	0.017	0.026
γ	–	0.03	–	–	–	0.032	0.036
ϕ	$^\circ$	-40	–	–	–	-40.767	-42.472
λ_s	–	-	(88.068 fixed)	110.271	110.983	83.306	75.358
l_s	arcsec	-	(0.128 fixed)	0.130	0.130	0.129	0.154
$\lambda_{\delta\psi}$	–	–	30796.077	32806.464	18.418	117961.620	20345.966
$l_{\delta\psi}$	arcsec	–	0.427	0.402	–	0.380	0.285
$-\frac{N_d}{2}\log(2\pi)^\dagger$			-3571.00	-3571.00	-3571.00	-3571.00	-3145.53
$\frac{N_s}{2}\log(\lambda_s)$			1534.75	1714.22	1716.57	1612.01	1668.14
$\frac{N_{\delta\psi}}{2}\log(\lambda_{\delta\psi})$			4650.81	4679.27	1311.00	5255.15	4579.85
$-\frac{1}{2}\log(\det C_d)^\dagger$			24021.55	24021.55	24021.55	24021.55	26011.99
$-\frac{1}{2}\log(\det C_s)$			591.56	586.57	587.30	584.15	394.75
$-\frac{1}{2}\log(\det C_{\delta\psi})$			884.23	872.31	4297.36	859.87	710.49
$-\frac{1}{2}\chi^2$			-1764.76	-1777.64	-1804.76	-1726.69	-1616.15
$-\frac{1}{2}\lambda_s s^T C_s^{-1} s - \frac{1}{2}\lambda_{\delta\psi}\delta\psi^T C_{\delta\psi}^{-1}\delta\psi$			-185.53	-215.73	-208.68	-170.42	-246.74
$-\frac{1}{2}\log(\det H)$			-8199.20	-8332.19	-8424.01	-8749.70	-7981.98
$\log P$			17962.42	17977.36	17925.32	18114.93	20374.82

Note. †constant

Table 4. Mean parameter values, 68 per cent confidence intervals, and evidence terms (same as Table 2). Models XXF, XFF, XFF-curv, and FFF-smooth are described in Section 3.3 and model FFF-complex in 3.4. The full probability densities for models FFF-smooth and FFF-complex are shown in Fig. 12. Notice that although the dimensions of the parameter space differ between the models, this is taken into account while integrating to calculate the evidence. We do not compare model FFF-complex to any other model, hence its evidence value is omitted.

| Name	Units	Truth	XXF	XFF	XFF-curv	FFF-smooth	FFF-complex
b	arcsec	0.9	–	–	–	$0.895^{+0.005}_{-0.005}$	$0.881^{+0.001}_{-0.002}$
q	–	0.8	–	–	–	$0.799^{+0.011}_{-0.010}$	$0.772^{+0.002}_{-0.005}$
θ	$^\circ$	-135	–	–	–	$-134.349^{+2.085}_{-2.171}$	$-133.769^{+0.643}_{-0.460}$
x_0	arcsec	0	–	–	–	$-0.044^{+0.003}_{-0.004}$	$-0.054^{+0.001}_{-0.000}$
y_0	arcsec	0	–	–	–	$0.017^{+0.003}_{-0.003}$	$0.026^{+0.001}_{-0.001}$
γ	–	0.03	–	–	–	$0.032^{+0.003}_{-0.003}$	$0.036^{+0.001}_{-0.002}$
ϕ	$^\circ$	-40	–	–	–	$-40.768^{+4.199}_{-3.640}$	$-42.472^{+1.076}_{-1.063}$
λ_s	–	–	–	$110.703^{+9.289}_{-10.443}$	$111.384^{+9.815}_{-9.443}$	$83.742^{+8.318}_{-9.443}$	$77.163^{+15.816}_{-16.099}$
l_s	arcsec	–	–	$0.130^{+0.004}_{-0.004}$	$0.130^{+0.004}_{-0.004}$	$0.129^{+0.004}_{-0.004}$	$0.159^{+0.022}_{-0.047}$
$\lambda_{\delta\psi}$	–	–	$31208.361^{+4786.931}_{-6465.456}$	$33295.353^{+4828.483}_{-6464.454}$	$18.628^{+2.487}_{-3.068}$	$120906.122^{+20792.873}_{-30053.728}$	$21494.286^{+5691.486}_{-7949.972}$
$l_{\delta\psi}$	arcsec	–	$0.431^{+0.053}_{-0.047}$	$0.405^{+0.043}_{-0.054}$	–	$0.381^{+0.038}_{-0.038}$	$0.292^{+0.035}_{-0.071}$
		$\log\mathcal{E}$	17956.22 ± 0.23	17963.48 ± 0.35	17913.27 ± 0.33	18082.77 ± 0.56	–
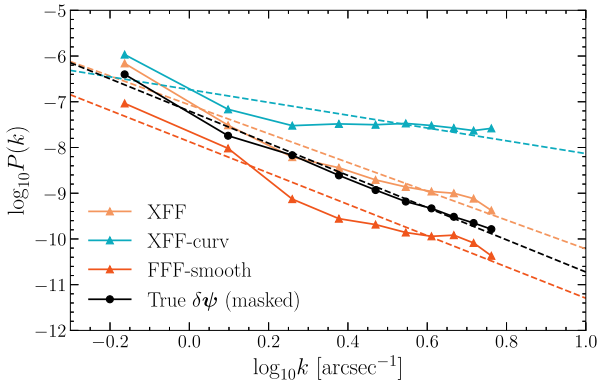
Figure 10. Fourier power spectrum of the perturbations shown in the third row of Fig. 8. The dashed lines are fits using equation (28) with the corresponding parameters listed in Table 5. The power spectra are computed within the mask.

Table 5. Power-law fits from equation (28) to the power spectra of the true and reconstructed $\delta\psi$ shown in Figs 10 (top part, Section 3.3) and 15 (bottom part, Section 3.4). The XXF and XFF models in the top part of the table give identical fits.

	$\log_{10}A$	β
True $\delta\psi$ (masked)	-7.20 ± 0.01	-3.52 ± 0.02
XFF-curv	-6.73 ± 0.03	-1.40 ± 0.03
XXF/XFF	-7.07 ± 0.02	-3.15 ± 0.07
FFF-smooth	-7.87 ± 0.03	-3.42 ± 0.10
True $\delta\psi$ (masked)	-7.21 ± 0.01	-3.43 ± 0.03
XXX $n = 3$	-7.19 ± 0.02	-3.29 ± 0.09
XXF-curv	-5.18 ± 0.02	-1.69 ± 0.07
FFF-complex	-7.16 ± 0.02	-3.26 ± 0.09
FFF-MAP $n = 2$	-7.46 ± 0.02	-2.75 ± 0.07

the slope decreases as the correlation function becomes narrower (or $l_{\delta\psi}$ becomes smaller), first for the XFF and then for the FFF-smooth model. We note, however, that although XFF gives the value for the amplitude closest to the truth, its parameters η are fixed to the true underlying smooth model (a quite unrealistic scenario), which means that the dimensions of the parameter space to explore are significantly fewer compared to the FFF-smooth model.

The FFF-smooth model has the smooth potential parameters η free, which in principle could absorb part of the perturbations. However, as discussed in Appendix B, this is not the case. The fitted smooth potential model is very close to the truth, meaning that any differences between the true total and reconstructed potentials is mostly due to the $\delta\psi$.

A parametric-only, purely smooth model is also evaluated, which is obviously insufficient to correctly model the lens, leading to biased values of η (see Table C1) and reconstructed s, and prominent residuals above the noise level (bottom right-hand panel in Fig. 8). These residuals are lower in amplitude and different from the (unmodelled) residuals between the smooth and perturbed data (bottom left-hand panel in Fig. 8), having a correlation coefficient of 0.26. This means that the perturbations are absorbed into the smooth model parameters and the source to some extent, but not fully (see Bayer 2021, for a thorough exploration of this effect). This can be seen in the residual power spectrum, shown in Fig. 11, where the 'unmodelled' residuals that appear on the large scales have significant power (above the noise) for $k < 4$ and the smooth model residuals have 2 to 7 times less power in the same range, yet still also 2 to 7 times more than the noise.

In Fig. 12, we show the full non-linear parameter probability densities for the FFF-smooth model. In general, the parameters η are distributed similarly to Fig. 3 but with larger statistical uncertainty. A systematic bias is introduced in b, whose lower values become more probable, because the inclusion of perturbations $\delta\psi$ can now absorb some of the overall strength of the lens potential. Similarly, the presence of the perturbations causes x_0 to be offset by one pixel instead of half, which was the case in Section 3.1. The same degeneracies are observed as in Fig. 3 between the parameters $b - q$, $b - \gamma$, $q - \gamma$, and $\theta - \phi$. The latter two have a bi-modal distribution with an extent of roughly ± 5 deg. Such small angular offsets between the SIE and the external shear can be understood in terms of the smoothness of the source, which allows for the perturbing field
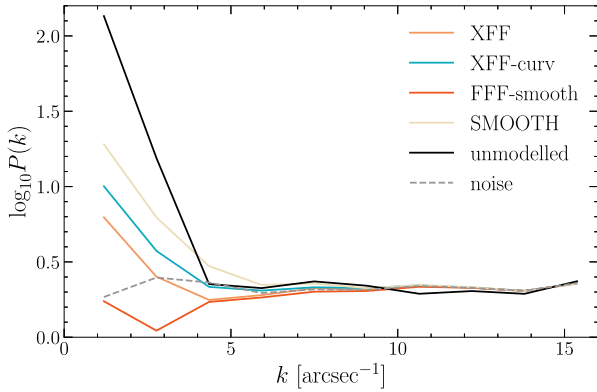
Fig. 10 we show the power spectra of the reconstructions and in Table 5 list the coefficients of the corresponding fits using equation (28). The connection between the slope of the power spectrum and stronger large-scale correlations is evident: the flattest power spectrum belongs to the model with curvature regularization, while

Figure 11. Fourier power spectrum of the model residuals shown at the bottom row of Fig. 8. The 'unmodelled' residuals correspond to the bottom left-hand panel of Fig. 8 and quickly drop to the noise level for $k > 4$.

$\delta\psi$ to make up for the difference and still provide solutions with high probability (low residuals). There are no correlations between η and the regularization parameters for the source or the potential perturbations, neither between the latter two. However, we observe again the expected anticorrelation between λ_s and l_s and a similar one between $\lambda_{\delta\psi}$ and $l_{\delta\psi}$ (better shown in Fig. A2), i.e. increasing the overall regularization parameters λ smooths out the reconstructed fields, as does increasing the correlation length l in the covariance kernels.

3.4 Perturbed lenses and complex sources

In reality, we expect complex sources to be lensed by non-smooth lens potentials. Here we combine the perturbed lens potential from the previous section with the complex brightness profile of NGC 2623 (a merger) used as source in Section 3.2. The resulting lensed images are shown in the left-hand column of Fig. 13. Although such a lensing scenario could be unrealistically complex – lensing of merging galaxies is not very probable – it serves as an extreme scenario for degeneracies to emerge as a result of the non-linear behaviour approximated by matrix M_r; from equation (7), perturbed deflection angles are associated with incoming rays from a highly structured source, and this information can be lost within the finite resolution of the mock data considered in our examples.

We model the system fixing the regularization kernels to the best-performing ones, i.e. an exponential kernel for the source (see Section 3.2) and a Gaussian for the perturbations (see Section 3.3). We reconstruct $\delta\psi$ in the same 30×30-pixel grid as before, and use $n = 3$ for the adaptive source grid, unless otherwise stated. For each of the models presented in Fig. 13 we either fix (X) or set free (F) each of the three parameter sets η, (λ_s, l_s), $(\lambda_{\delta\psi}, l_{\delta\psi})$ and name it accordingly, in the same way as in Section 3.3. For the fixed values of the parameters we have: the true values for η (e.g. see Section 3.1 or Table 1), $l_s = 0.15$ and $l_{\delta\psi} = 0.36$, which are the values fitted to the true source and perturbations as shown in Figs 6 and 9, $\lambda_s = 44.031$, the mean value from Section 3.2 (see Table 2), and $\lambda_{\delta\psi} = 86780.1$, the mean value from the FFF-smooth model presented in Section 3.3 (see Table 4).

In the first part of Fig. 13, we show two models with all the parameters fixed to the truth, one with $n = 3$ and one with $n = 2$, and three models with only one parameter set allowed to vary. We first note that there is very little difference in the residuals and the reconstructed $\delta\psi$ between the fixed models (despite the

many more source pixels for the case with $n = 2$) and the one with the source parameters free (XFX). However, allowing the $\delta\psi$ regularization parameters to vary leads to a worse reconstruction and the residuals increase. This is even more prominent if we change the $\delta\psi$ regularization from a Gaussian to a curvature kernel. These $\delta\psi$ solutions have too much structure (low regularization) because they may be actually overcompensating for a low resolution adaptive source grid. In the second part of Fig. 13, we see that the residuals and the $\delta\psi$ reconstruction do not improve if we set both the perturbation and source regularization parameters free (i.e. compare models XXF and XFF). As soon as we allow η to vary then the residuals do decrease at the cost of a less smooth $\delta\psi$ reconstruction. This is regardless of fixing the source regularization parameters – models FXF and FFF-complex give very similar results. However, the adaptive grid resolution affects the residuals: after fixing all parameters to the MAP values from the FFF-complex model, we set $n = 2$ and although the $\delta\psi$ reconstruction does not improve too much, the residuals do (see also Fig. 16), in particular, the prominent positive residuals due north with respect to the lens in the models FXF and FFF-complex considerably decrease. This is most likely due to the more degrees of freedom available for the source, which is further supported by a smooth model with $n = 1$ that absorbs the perturbations almost down to the noise level.

Looking at the two-point correlation functions of the reconstructed $\delta\psi$ shown in Fig. 14, we note that the prior and the data lie close to each other, which accordingly drives the FFF-complex model. Increasing the adaptive grid resolution leads to somewhat stronger correlations on the larger scales and brings the reconstructed $\delta\psi$ even closer to both the prior and the data. However, in Fig. 15, and from the fitted coefficients listed in Table 5, there is a remarkable agreement between the power spectrum of the FFF-complex model (all the parameters free) and the true $\delta\psi$. The same holds for the reconstructed $\delta\psi$ of the XXX $n = 3$ model that has all the parameters fixed to their true values. Hence, despite their different appearance (see the third row of panels in Fig. 13) the reconstructed $\delta\psi$ of the FFF-complex (and XXX $n = 3$) model have an almost identical power spectrum to the truth. We also note that the residual power spectrum of the FFF-complex and the XXX $n = 3$ models, shown in Fig. 16, is very similar, with both models being above the noise in the small scales ($k < 5$). Curvature regularization is clearly a bad prior for the GRF $\delta\psi$ as it leads to prominent residuals, even more than the difference between the unmodelled perturbed and unperturbed mock systems (see Fig. 16), and more extreme values of the reconstructed $\delta\psi$ (see Fig. 15 and Table 5). Completely ignoring the existence of any perturbations and modelling the system with a purely smooth model with $n = 1$ can reach the noise level (see Fig. 16). This is clearly a biased solution that could model away substructure or deviations from the smooth potential.

In Fig. 12 we compare the full non-linear parameter probability densities of the FFF-complex model presented here to the FFF-smooth model presented in Section 3.3. Its MAP and mean parameter values, and the 68 per cent confidence intervals are listed in Tables 3 and 4. The two models are actually the same but applied to different data, i.e. with a difference source light profile. We can observe three main characteristics of the distributions: (i) smaller statistical uncertainties, (ii) larger systematic biases, and (iii) fragmentation of the probability surfaces, with various local maxima separated by valleys and saddles, given rise to a complex parameter space configuration. The latter reflects the complex and degenerate underlying lens potential perturbations and source brightness profile. The smooth lens potential parameters η are correlated in the same way as before but the biases are more significant. The SIE potential strength
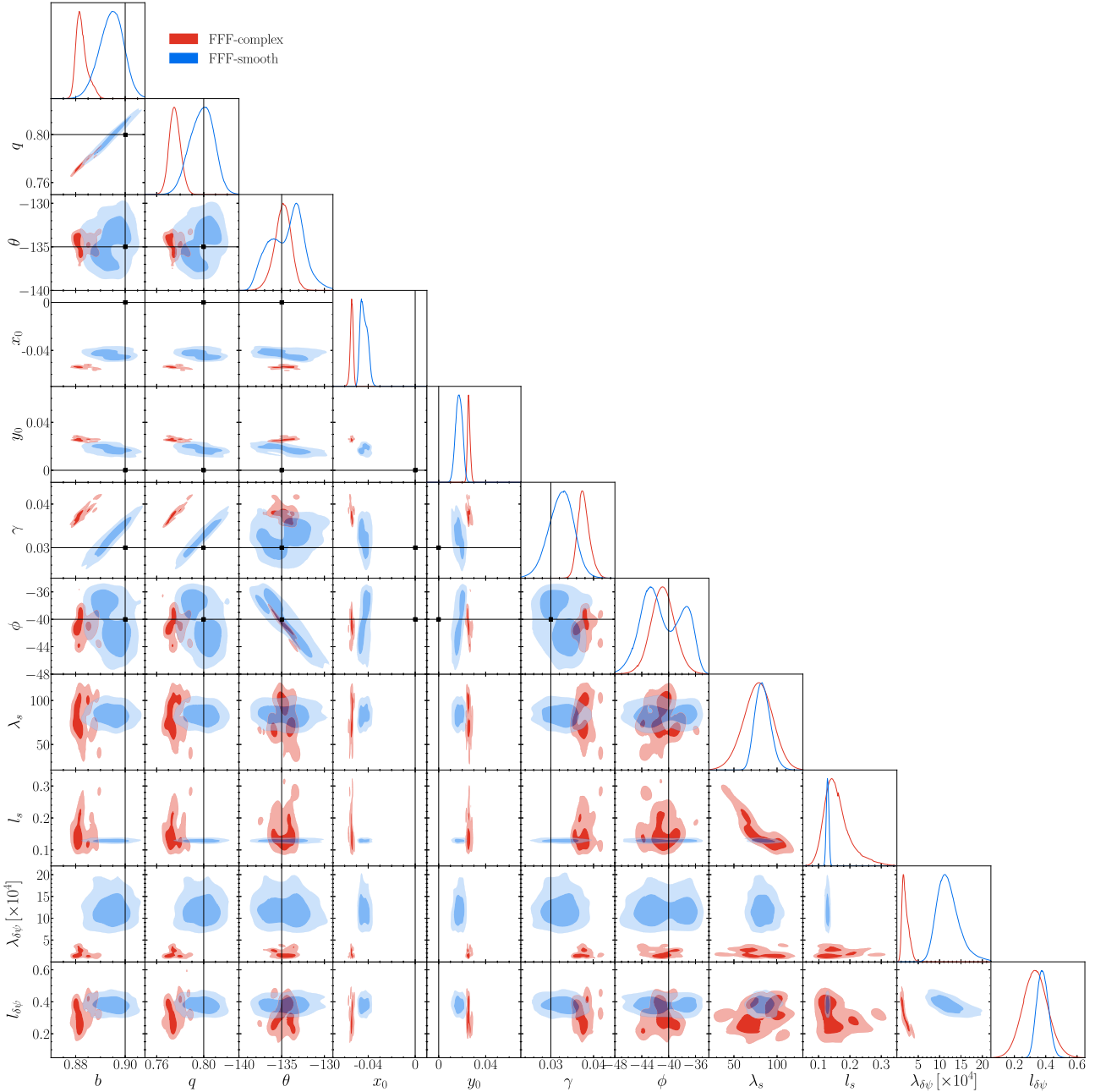
Figure 12. Same as Fig. 3, including the perturbation parameters $\lambda_{\delta\psi}$, $g_{\delta\psi}$, for the FFF-smooth (blue) and FFF-complex (red) models, described in Sections 3.3 and 3.4, respectively. The two models are actually the same and have the same free parameters, i.e. the smooth potential and regularization parameters for the source and the potential perturbations, but applied to mock data with different source light profiles. The corresponding mean values and 68 per cent confidence intervals are given in Table 4. A separate, zoomed-in version of this plot for each model is shown in Figs A2 and A3, which show the shape of the 2D distributions better.

b is pushed to even lower values as the $\delta\psi$ are now stronger (e.g. compare the reconstructed MAP perturbations between the FFF-smooth and the FFF-complex models in Figs 8 and 13, respectively), x_0 and y_0 are offset by approx. 1 pixel, and q and γ lie several σ further than their true values. Only the angles θ and ϕ are not biased and are in fact less degenerate than the FFF-smooth model, i.e. their distributions are not bi-modal anymore. This is because of the more detailed structure in the source that cannot be accounted for well by the perturbing field $\delta\psi$ for tilted smooth potentials. All of the regularization parameters have broader distributions except $\lambda_{\delta\psi}$ that is more narrowly distributed around values 3–4 times smaller than

the FFF-smooth model. This means that more structured and larger in amplitude $\delta\psi$ reconstructions are expected, which is indeed the case as shown in Fig. 13. A very strong anticorrelation is observed between the regularization strengths, λ, and correlation lengths, l, in the covariance kernels for both the source and the perturbations. Finally, the complex probability surfaces between the source and potential perturbation regularization parameters (see also Fig. A3) mean that the two are quite degenerate. The smaller values of $\lambda_{\delta\psi}$ in combination with the broader l_s distribution towards higher values indicate that the complexity of the source brightness is absorbed by the potential perturbations.
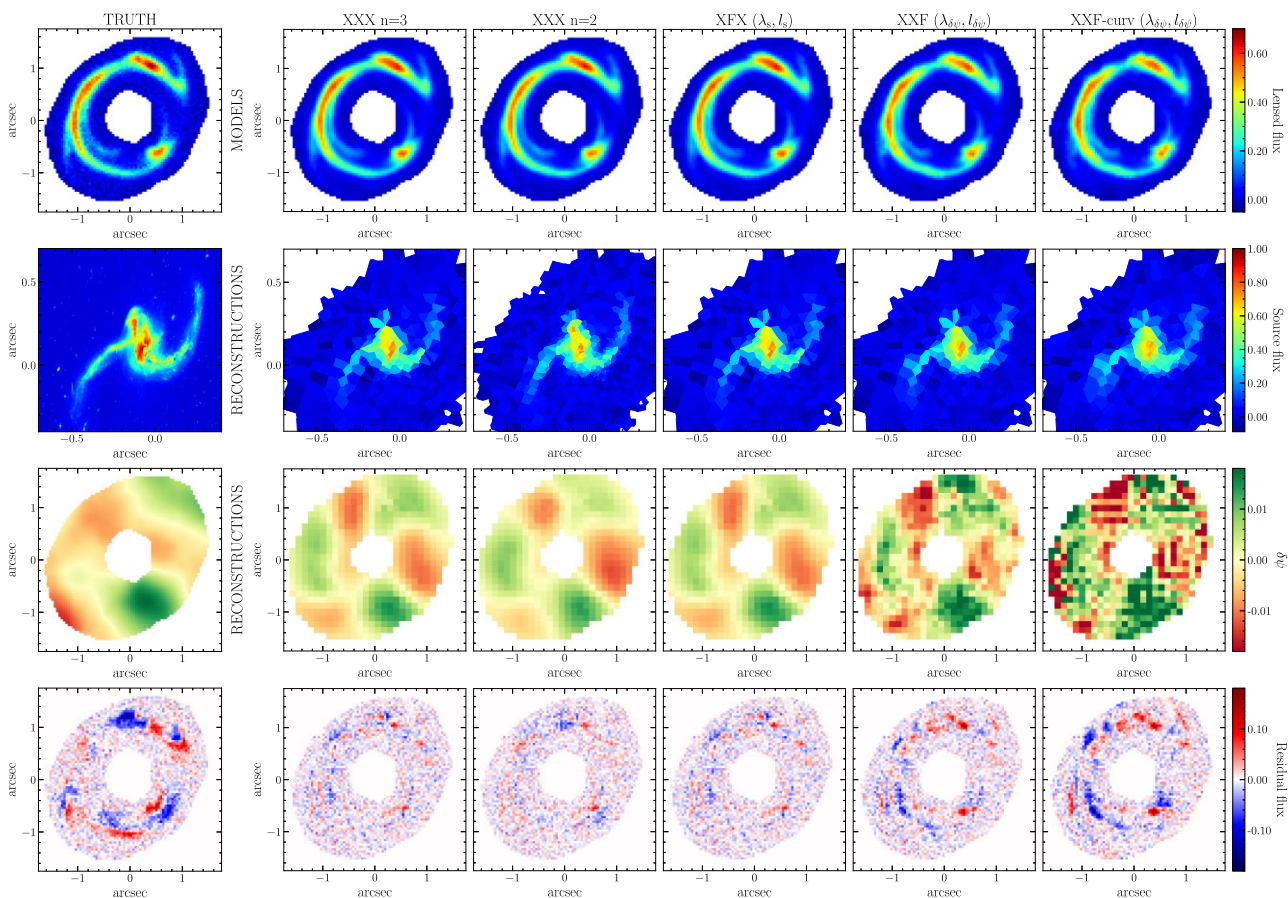
Figure 13. Same as Fig. 8 for NGC 2623. The bottom left-hand panel shows the difference between the perturbed (top left-hand panel) and unperturbed systems (top left-hand panel of Fig. 5). We list the free parameters of each model in the parenthesis next to its name at the top (see Section 3.4 for details).

4 DISCUSSION

Higher order statistical properties of the brightness profiles of gravitationally lensed galaxies can be incorporated in the semi-linear inversion technique through regularization priors based on physically motivated covariance kernels. In this work, we created mock gravitational lenses using NGC 3982 (a spiral) and NGC 2623 (a merger) as sources, whose covariance is well-described by a Gaussian and exponential covariance kernel, respectively. We found that these physically motivated priors outperform other traditionally used regularization schemes, such as identity and curvature, and we can model each system down to the noise level in almost all cases while simultaneously avoiding overfitting (some residuals remain in the case of perturbed potentials).

Using generic covariance priors comes at the cost of introducing additional non-linear parameters (in this case, the correlation length l_s; see equations 25 and 26). Our modelling framework can handle these new parameters and determine their full probability distribution jointly with the other non-linear parameters (e.g. the smooth mass model parameters, η) at the cost of a now denser source covariance matrix, C_s, that needs to be inverted (e.g. see equation 12), and slower convergence due to increasing the dimensions of the non-linear parameter space that needs to be explored. However, here we used logarithmic priors on a wide range of l_s, which might be a conservative choice. One could use observationally driven estimates of l_s (or other covariance kernel parameters) derived from populations of putative lensed sources, e.g. constructed from samples of observed lenses, in order to narrow-down the parameter space and speed up

the modelling process. In fact, we performed such a test by fixing l_s = 0.21 for NGC 3982, a value well-justified by the observations (see Fig. 6), and remodelling the corresponding mock lens, achieving a much faster convergence to the same result.

The quality of the data, viz. high signal-to-noise and resolution, plays a major role in finding an acceptable solution for the source, regardless of the choice of prior, observationally motivated or not, on the source brightness profile. In the cases examined in Section 3.2, the data are of sufficiently good quality to drive the solution close to the truth for all tested regularization schemes. For NGC 2623, the recovered l_s parameter for the case with an exponential covariance kernel – the one matching the true source – lies further than 3σ from the truth, despite having the highest evidence. The reverse statement, viz. whether the use of a (correct) prior becomes more important in the case of degraded/noisy data, is yet to be systematically explored. This is particularly relevant for upcoming surveys, such as *Euclid* and *LSST*, which are expected to have lower angular resolution than what we examined here. However, our method does prefer the models with the correct priors based on the Bayesian evidence, for the adopted observational setup.

Once perturbations to the lensing potential are introduced, we need to approach the problem in a different way. We demonstrated that the effect of $\delta\psi$ can be absorbed in the reconstructed source, especially if the adaptive grid resolution is set to the highest (n = 1, a common choice), and lead to wrong results on the model parameters, η, and the source, s. This, in turn, leads to spurious structures in the model residuals, unrelated to the original $\delta\psi$, which can be misinterpreted as the effect of a perturbing field of mass
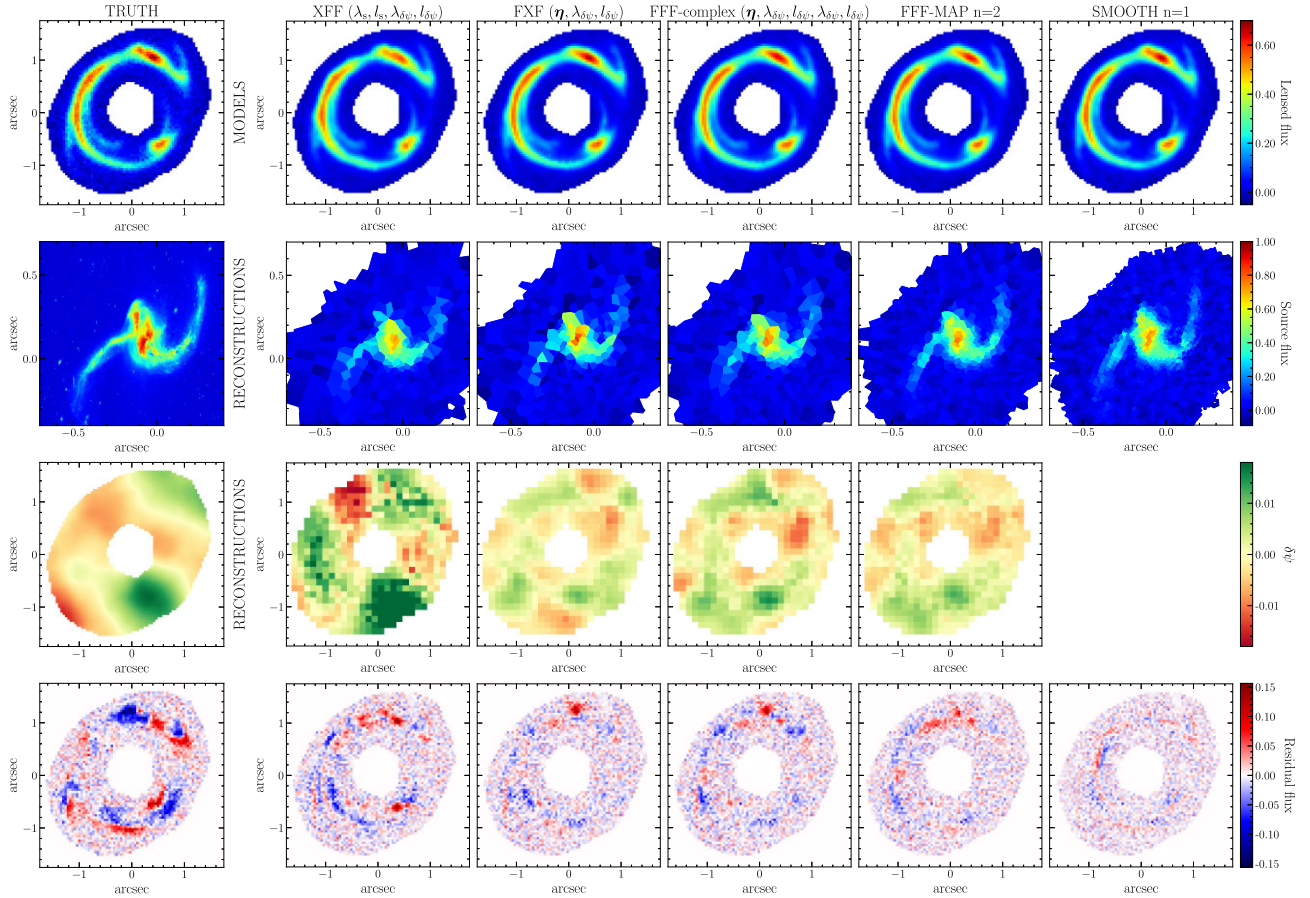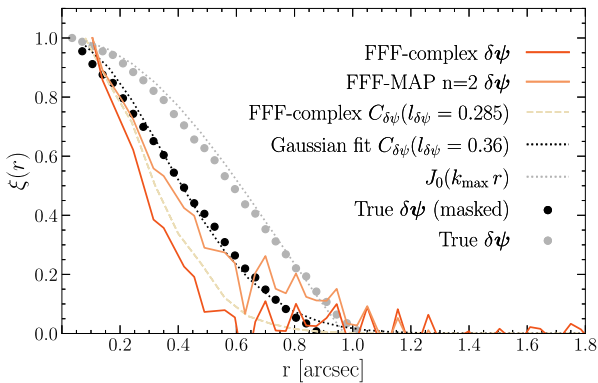
Figure 13. *– continued*

Figure 14. Radially averaged two-point correlation functions of the $\delta\psi$ reconstructions from the FFF-complex and FFF-MAP $n = 2$ models (see Section 3.4). We include the prior (dashed lines), with the $l_{\delta\psi}$ parameter for the Gaussian covariance kernel set to its MAP value, i.e. 0.285 (see Table 3). The true two-point correlation functions of the full GRF (grey circles) and the one within the mask (black cirles) are shown, together with a Gaussian fit to the latter with $l_{\delta\psi} = 0.36$ (using equation 26, dotted black line) and equation (29) with k_{max} set to the diagonal of the 3.5 arcsec-wide image (dotted grey line, not a fit).

Figure 15. Fourier power spectrum of some of the $\delta\psi$ reconstructions shown in the third row of Fig. 13. The dashed lines are fits using equation (28) with the corresponding parameters listed in Table 5. The power spectra are computed within the mask.

substructure (see also Chatterjee 2019, for another study on this). Hence, a two-step approach of first running a parametric smooth model to constrain η and then modelling the perturbations $\delta\psi$ would be unreliable (unless lower choices for n are used, e.g. see Bayer

2021). The extent of the above statement for perturbed lenses with varying $\delta\psi$ properties, as well as concentrated massive substructures, remains to be explored.

Attempting to reconstruct the perturbing $\delta\psi$ requires a regularizing term (prior) in addition to the one for the source. In contrast to the case of smooth potentials, where the data quality is good enough to drive the source reconstructions to solutions with the desired statistical properties regardless of which regularization scheme is used (see Fig. 6), the data alone are not sufficient and the form of
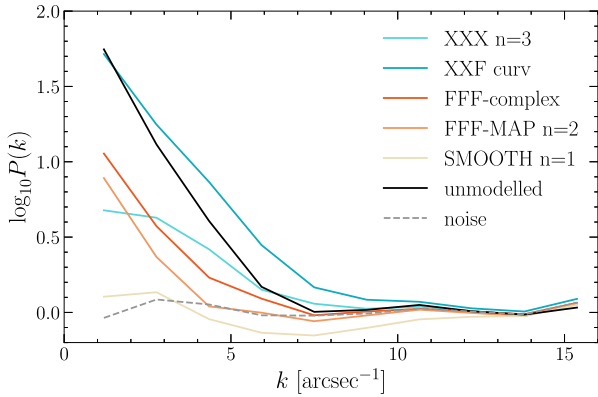
Figure 16. Fourier power spectrum of the residuals shown at the bottom row of Fig. 13 for the models used in Fig. 15. The 'unmodelled' residuals correspond to the bottom left-hand panel of Fig. 13 and drop to the noise level for $k > 7$.

regularization/prior seems to play a major role in reconstructing $\delta\psi$. Here we examined specifically the curvature and Gaussian covariance kernels, in connection to our choice of a GRF as the true underlying $\delta\psi$. The traditionally used curvature regularization is less flexible as it imposes fixed, long range correlations (see Fig. 9), which are in fact stronger than they should and irrecoverably lead to unphysically smooth solutions, seemingly regardless of the quality of the data. The covariance of our assumed GRF, however, can be well approximated by a Gaussian kernel (see Fig. 9), but in real galaxies the true covariance of potential perturbations is unknown. More flexibility could be achieved by assuming a covariance kernel described by a number of free parameters, e.g. a Matérn kernel (e.g. Mertens et al. 2017; Vernardos et al. 2020), or even a free form two-point correlation function. In addition, theoretically justified $\delta\psi$ priors could be derived based on dark matter models or *N*-body hydrodynamical simulations. Our method allows for a thorough and quantitative exploration of how different regularization schemes on the $\delta\psi$, as well as on the source, can affect the quality of the reconstructions, eventually ranking them by their Bayesian factors.

In Sections 3.3 and 3.4 we fully model the smooth potential, source, and perturbations in two example cases whose only difference is the brightness profile of the source, i.e. the smooth lens potential and the perturbative field of $\delta\psi$ remain the same. Our optimization strategy (described in Section 2.6) works quite well, but the extent of statistical uncertainty and systematic biases in the recovered parameters η depend on the complexity of the source brightness profile. In addition, we find a complicated degeneracy with the strength and smoothness of the regularization for the perturbations as well. In the case of the smooth source presented in Section 3.3, the recovered $\delta\psi$ from the FFF-smooth model match really well the shape of the true underlying perturbations but their amplitude is underestimated almost by an order of magnitude in comparison to the XFF model. This is clearly seen in the power spectra of the two $\delta\psi$ solutions shown in Fig. 15, however, it should be noted that they still have a slope that agrees very well with the truth. In the case of the complex source presented in Section 3.4, the power spectrum of the $\delta\psi$ solution of the FFF-complex model is recovered remarkably well, however, the shape of the actual $\delta\psi$ does not match well the underlying truth. In this case, the entire parameter space becomes more structured and degenerate (see Fig. A3) and systematic biases increase (see Fig. 12), which is especially true for the smooth potential strength, *b*, that is pushed to lower values. Smoother sources

become more compatible with the data and the freedom of the perturbing $\delta\psi$ is increased (i.e. its smoothness reduced), which leads to the latter absorbing the structure of the source. This explains why the reconstructed $\delta\psi$ from the FFF-complex model in Fig. 13 do not visually match the true GRF very well although their power spectrum does (see Figs 10, 15, and Table 5).

The visual differences of the reconstructed $\delta\psi$ compared to the truth (see the FFF-complex and FFF-smooth reconstructions in Figs 13 and 8, respectively), could be understood in terms of the 'light-constrains-mass' effect, which we explain here. Within the framework of our method, but also more generally, it is important to clarify how is $\delta\psi$ constrained where the lensed source brightness, and/or, more precisely, the gradient of the source is low or zero. Obviously, in such areas using equation (3) to model brightness residuals becomes problematic; the D_s operator, which holds the derivatives of the source at the source plane (deflected) location of the given image pixel(s), becomes zero. Hence, in order to obtain a reconstruction across the entire field of view (or even within a mask) it now becomes obvious that the regularization will be important, particularly where there is low/no source flux. This is analogous – but not exactly – to reconstructing the source brightness on pixels that are not constrained by the data, as could be the case in a fixed grid model. Taking the realization of the GRF $\delta\psi$ field that we used as an example (third-row panel in the left of Fig. 8), the success of our reconstructions depends on how much of the source flux eventually end ups in those crucial areas of the lens plane that have the largest gradients (largest deflection angles). This could play a role in the more degenerate results of the FFF-complex model and its $\delta\psi$ power spectrum amplitude difference with the FFF-smooth model (see Figs 10, 15, and Table 5). This could be mitigated by reconstructing the $\delta\psi$ within a carefully selected region of the lens plane around the lensed source brightness, possibly weighed by the values of the operator D_s. However, determining the extent of this 'light-constrains-mass' area may introduce another possible source of degeneracy: the gradient of $\delta\psi$, which is in fact the deflection angle, also enters equation (3), and for any pixel with some given lensed source brightness, regions having the same gradient, e.g. large density differences that lie further away or smaller density differences being closer, can have the same effect.

5 CONCLUSIONS

We explored the effect of regularization while reconstructing both the source and potential perturbations using the semi-linear inversion technique. Below we summarize the conclusions from this work and outline future directions of application and improvement.

(i) Physically motivated priors for the source galaxies, such as Gaussian and exponential kernels, lead to better results than traditional choices, such as identity and curvature regularization.

(ii) Curvature regularization, a traditionally popular choice, is fundamentally unsuitable as a prior for the GRF $\delta\psi$ perturbations that we examined here.

(iii) The source alone can absorb the structure created by $\delta\psi$ almost down to the noise, especially if a high resolution adaptive grid is used (low value of *n*). This leads to biased source reconstructions and parameters for the smooth potential (see also also Bayer et al. 2018; Chatterjee 2019; Bayer 2021).

(iv) The statistical properties of the $\delta\psi$, particularly the power spectrum, are recovered remarkably well, both for smooth and more complex sources.

Our study constitutes an initial exploration and test of our new code implementation, and as such we restricted ourselves to the four distinct and incrementally more complex examples presented in Section 3. The successful outcome of this study enables further and more in depth investigations of potential perturbation reconstructions in lensed systems. We propose, but not limit ourselves to, the following directions of future research:

(i) Here we used a specific GRF as the perturbing field, with specific amplitude (≈ 13 per cent of the smooth potential) and slope, which we believe is an extreme case, pushing the validity of the approximation of equation (3) to its limit. The type (GRF or other), as well as the associated parameter space of the perturbing field can be now explored more in depth, for different smooth potentials and sources.

(ii) One such case of particular interest would be using isolated massive perturbers as the perturbing $\delta\psi$, and determining how the conclusions of this work apply to it, e.g. comparing to the work of Vegetti & Koopmans (2009).

(iii) We have identified an interplay between data quality and priors in determining the best model, which needs to be explored in both directions: at which level of resolution and/or signal-to-noise ratio the data are driving the solution and the prior begins to play a secondary role, and inversely.

(iv) Our $\delta\psi$ reconstructions away from pixels that contain most of the lensed source flux are constrained mostly by the prior – what we described as the 'light-constrains-mass' effect. A weighed scheme – similar to adaptive regularization – could be devised to suppress terms in the D_s appearing in equation (3) that are very low or zero.

Finally, our new implementation of the method, the Very Knotty Lenser code, is made publicly available.[15]

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data that support the findings of this study are openly available in github at https://github.com/gvernard/verykool.

REFERENCES

Auger M. W., Treu T., Bolton A. S., Gavazzi R., Koopmans L. V., Marshall P. J., Moustakas L. A., Burles S., 2010, ApJ, 724, 511
Barkana R., 1998, ApJ, 502, 531
Barnabè M., Czoske O., Koopmans L. V., Treu T., Bolton A. S., 2011, MNRAS, 415, 2215
Barnabe M., Spiniello C., Koopmans V. E., Trager S. C., Czoske O., Treu T., 2013, MNRAS, 436, 253
Bayer D., 2021, PhD thesis, University of Groningen
Bayer D., Chatterjee S., Koopmans L. V. E., Vegetti S., McKean J. P., Treu T., Fassnacht C. D., 2018, preprint (arXiv:1803.05952)
Birrer S., Amara A., 2017, J. Cosmol. Astropart. Phys., 2017, 037
Birrer S., Amara A., Refregier A., 2015, ApJ, 813, 102
Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, MNRAS, 415, L40
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Carlberg R. G., Grillmair C. J., Hetherington N., 2012, ApJ, 760, 75
Chatterjee S., 2019, PhD thesis, University of Groningen
Chatterjee S., Koopmans L. V. E., 2018, MNRAS, 474, 1762
Dalal N., Kochanek C. S., 2002, ApJ, 572, 25
Diaz Rivero A., Cyr-Racine F. Y., Dvorkin C., 2018, Phys. Rev. D, 97, 23001
Dutton A. A., Treu T., 2014, MNRAS, 438, 3594
Dye S., Warren S. J., 2005, ApJ, 623, 31
Erkal D., Belokurov V., Bovy J., Sanders J. L., 2016, MNRAS, 463, 102
Fadely R., Keeton C. R., 2012, MNRAS, 419, 936
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Galan A., Peel A., Joseph R., Courbin F., Starck J. L., 2021, A&A, 647, A176
Gallier J., 2011, Texts in Applied Mathematics, vol. 38, Geometric Methods and Applications. Springer, New York, NY, Chapt. 11
Gavazzi R., Treu T., Rhodes J. D., Koopmans V. E., Bolton A. S., Burles S., Massey R. J., Moustakas L. A., 2007, AJ, 667, 176
Gilman D., Birrer S., Treu T., Keeton C. R., Nierenberg A., 2018, MNRAS, 481, 819
Helmi A., Sales L. V., Starkenburg E., Starkenburg T. K., Vera-Ciro C. A., De Lucia G., Li Y. S., 2012, ApJ, 758, 1
Hezaveh Y. D. et al., 2016b, ApJ, 823, 1
Hezaveh Y. D., Dalal N., Holder G., Kisner T., Kuhlen M., Levasseur L. P., 2016a, J. Cosmol. Astropart. Phys., 11, 48
Hsueh J. W. et al., 2017, MNRAS, 469, 3713
Jeffreys H., 1998, The Theory of Probability, 3ed edn. Oxford, England
Joseph R., Courbin F., Starck J. L., Birrer S., 2019, A&A, 623, A14
Kassiola A., Kovner I., 1993, ApJ, 417, 450
Kim S. Y., Peter A. H., Hargis J. R., 2018, Phys. Rev. Lett., 121, 211302
Klypin A., Kravtsov A., Valenzuela O., 1999, ApJ, 522, 82
Komatsu E. et al., 2011, ApJS, 192, 18
Koopmans L. V., 2005, MNRAS, 363, 1136
Koopmans L. V. E., Treu T., Bolton A. S., Burles S., Moustakas L. A., 2006, ApJ, 649, 599
Koopmans L. V. et al., 2009, ApJ, 703, 51
Kormann R., Schneider P., Bartelmann M., 1994, A&A, 284, 285
Krist J., Hook R., Stoehr F., 2010, Astrophysics Source Code Library, record ascl:1010.057
Li R., Frenk C. S., Cole S., Gao L., Bose S., Hellwing W. A., 2016, MNRAS, 460, 363
MacKay D. J. C., 1992, PhD thesis, Caltech
MacKay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambridge
MacLeod C. L., Jones R., Agol E., Kochanek C. S., 2013, ApJ, 773, 35
McKee C. F., Ostriker E. C., 2007, Annu. Rev. Astron. Astrophys., 45, 565
Mertens F. G., Ghosh A., Koopmans L. V. E., 2017, MNRAS, 478, 3640
Moore B., 1994, Nature, 370, 629
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, 19
Morningstar W. R. et al., 2019, ApJ, 883, 14
Nightingale J. W., Dye S., 2015, MNRAS, 452, 2940
Nierenberg A. M., Treu T., Wright S. A., Fassnacht C. D., Auger M. W., 2014, MNRAS, 442, 2434
Nightingale J. W., Dye S., Massey R. J., 2018, MNRAS, 478, 4738
Oguri M., Rusu C. E., Falco E. E., 2014, MNRAS, 439, 2494
Oh S. H. et al., 2015, AJ, 149, 1
Oldham L. J., Auger M. W., 2018, MNRAS, 476, 133
Papastergis E., Giovanelli R., Haynes M. P., Shankar F., 2014, A&A, 574, 113
Planck Collaboration VI, 2020, A&A, 641, A6
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical Recipes in C, 2nd edn . Cambridge Univ. Press, Cambridge
Rasmussen C. E., Williams C. K., 2006, Gaussian Processes for Machine Learning. MIT Press, Cambridge
Schaye J. et al., 2015, MNRAS, 446, 521
Schneider P., Kochanek C. S., Wambsganss J., 2006, in Meylan G., Jetzer P., North P., eds, Saas-Fee Advanced Course, Vol. 33. Springer, Berlin
Skilling J., 2004, in Fischer R., Preuss R., von Toussaint U., eds, AIP Conf. Proc. Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. Am. Inst. Phys., New York, p. 395

[15] https://github.com/gvernard/verykool

Smith R. J., Lucey J. R., Conroy C., 2015, MNRAS, 449, 3441

Sonnenfeld A., Treu T., Suyu S. H., Marshall P. J., Auger M. W., Nipoti C., 2013, ApJ, 777, 98

Sonnenfeld A., Leauthaud A., Auger M. W., Gavazzi R., Treu T., More S., Komiyama Y., 2018, MNRAS, 481, 164

Spiniello C., Barnabè M., Koopmans L. V., Trager S. C., 2015, MNRAS, 452, L21

Suyu S. H. et al., 2014, ApJ, 788, L35

Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983

Tagore A. S., Keeton C. R., 2014, MNRAS, 445, 694

Tikhonov A., 1963, Sov. Math., 4, 1035

Vegetti S., Koopmans L. V., 2009, MNRAS, 392, 945

Vegetti S., Koopmans L. V. E., Bolton A., Treu T., Gavazzi R., 2010, MNRAS, 408, 1969

Vegetti S., Lagattuta D. J., McKean J. P., Auger M. W., Fassnacht C. D., Koopmans L. V., 2012, Nature, 481, 341

Vernardos G., 2022, MNRAS, 511, 4417

Vernardos G., Tsagkatakis G., Pantazis Y., 2020, MNRAS, 499, 5641

Vogelsberger M. et al., 2014, Nature, 509, 177

Warren S. J., Dye S., 2003, ApJ, 590, 673

Wayth R. B., Warren S. J., Lewis G. F., Hewett P., 2005, MNRAS, 360, 1333

White S. D., Rees M. J., 1978, MNRAS, 183, 341

Yildirim A., Suyu S. H., Halkola A., 2020, MNRAS, 493, 4783

APPENDIX A: CORNER PLOTS

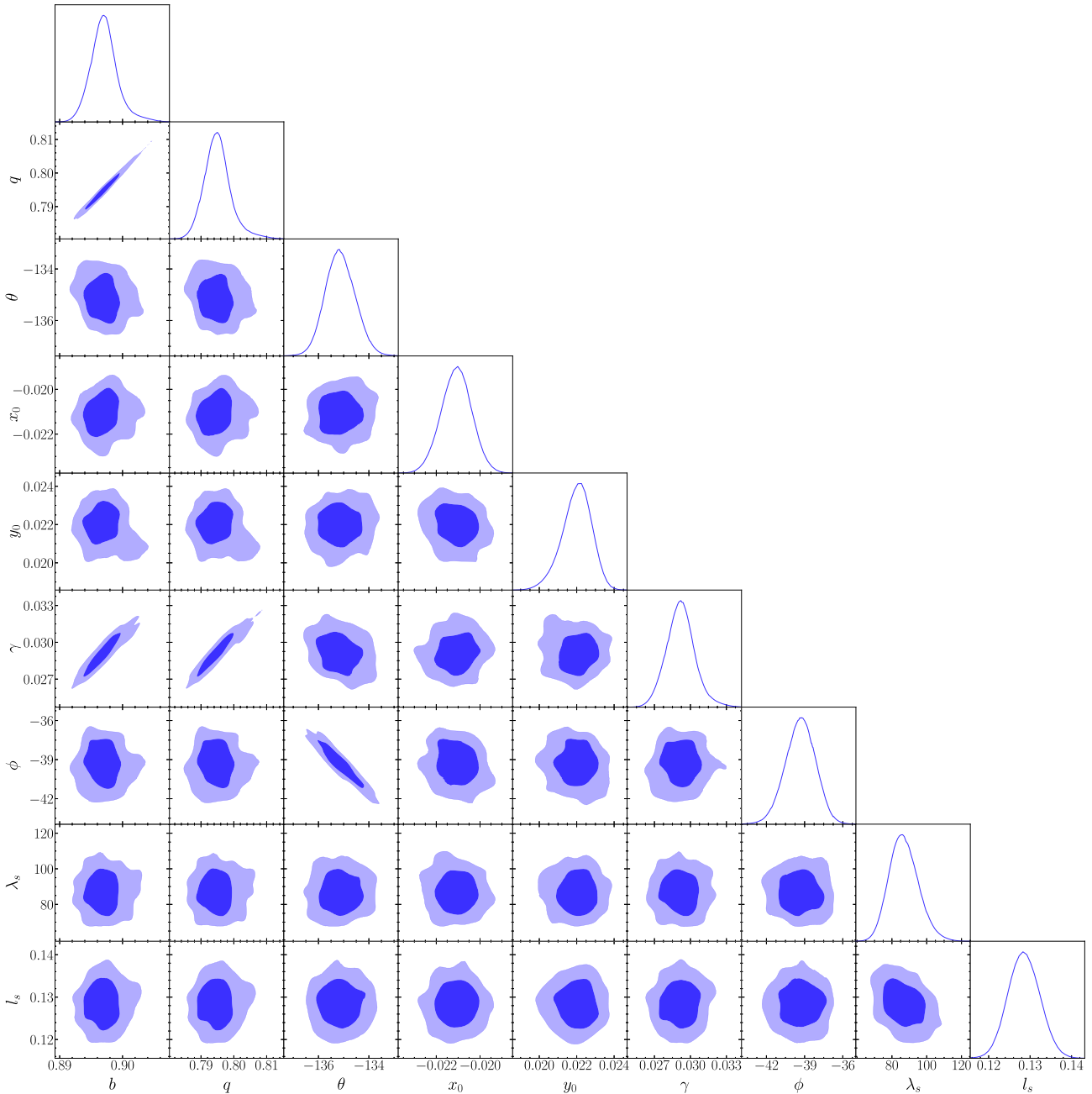Zoomed-in versions of the probability distributions shown in Figs 3 and 12.

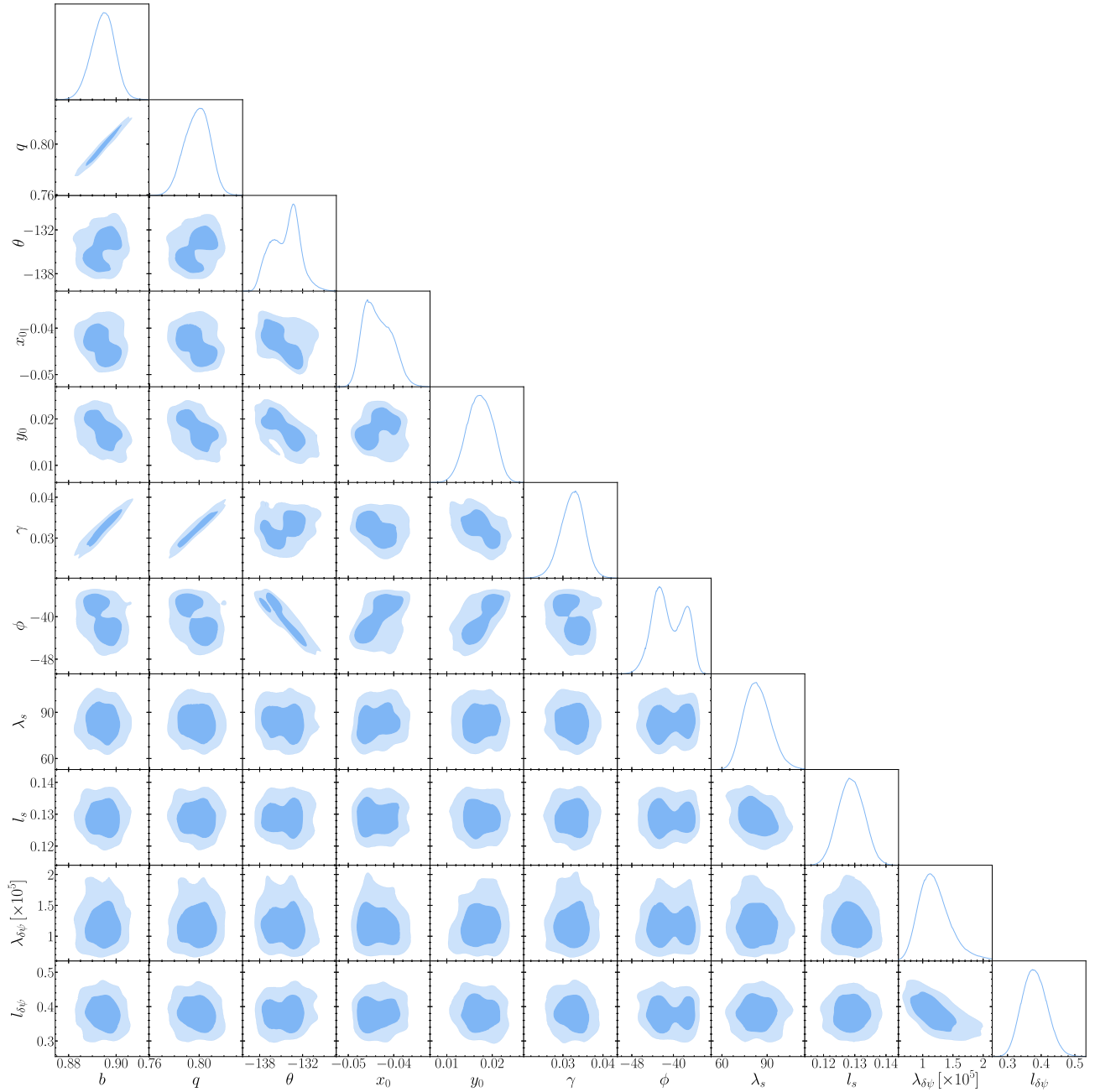**Figure A1.** Same as Fig. 3 with zoomed-in ranges to better show the shape of the 2D distributions.

Figure A2. Same as Fig. 12 for the FFF-smooth model with zoomed-in ranges to better show the shape of the 2D distributions.
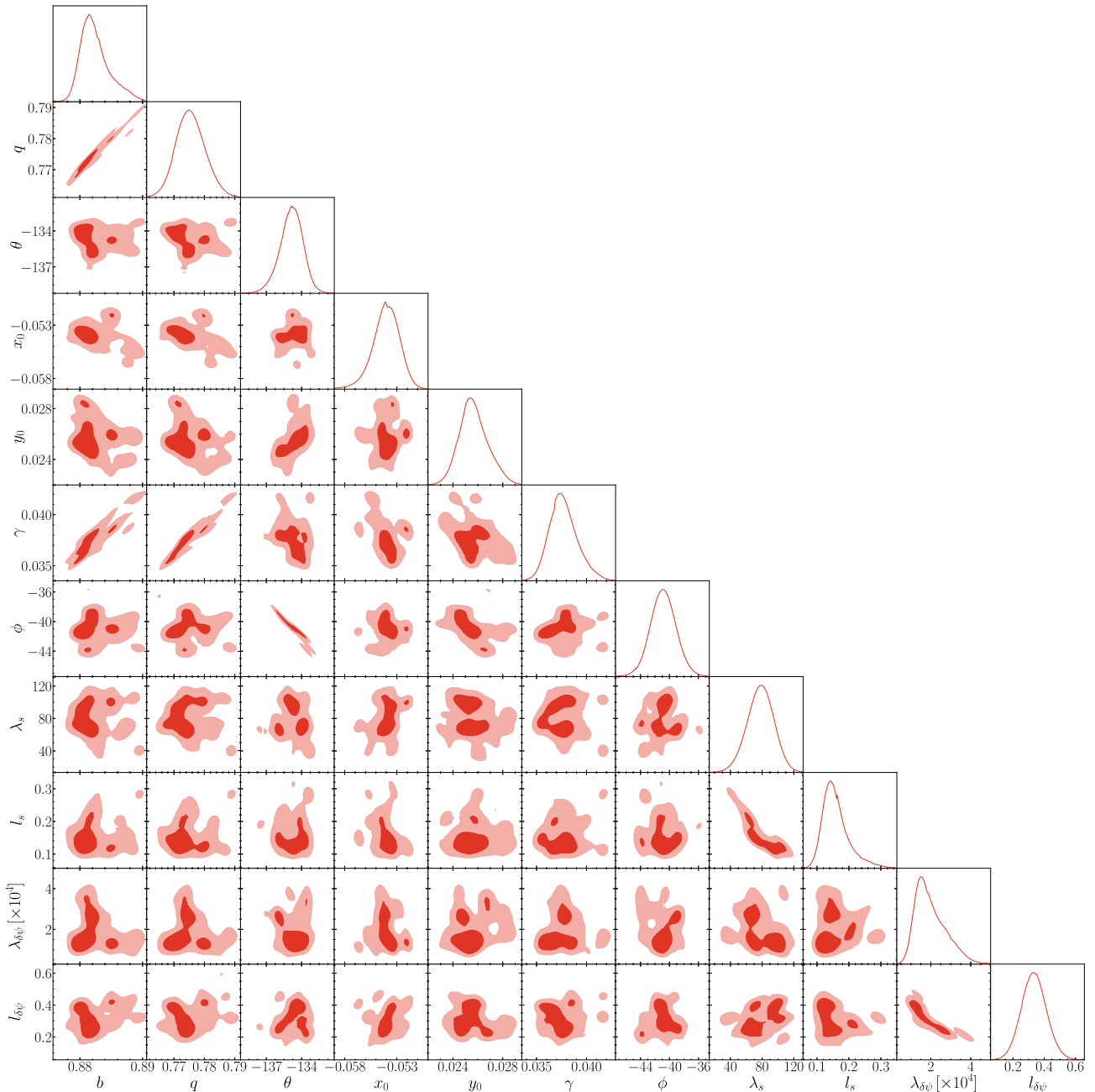
Figure A3. Same as Fig. 12 for the FFF-complex model with zoomed-in ranges to better show the shape of the 2D distributions.

APPENDIX B: THE SMOOTH POTENTIAL ABSORBS THE PERTURBATIONS TO A VERY SMALL EXTENT

In Fig. B1, we show the true smooth potential, which is an SIE with external shear described by the parameters η introduced in Section 3.1 (see also Table 1), as well as its MAP fits by the FFF-smooth and FFF-complex models described in Sections 3.3 and 3.4, respectively. These two models simultaneously fit the smooth potential and reconstruct its perturbations. Both models recover accurately the smooth potential parameters as listed in Tables 3 and 4 and shown in Fig. B1. Therefore, the observed differences in the reconstructions of $\delta\psi$ and s are purely due to their fundamental connection through equation (3) and the choice of regularization.

Finally, Fig. B2 shows the corresponding power spectra of the smooth potentials shown in Fig. B1. The power spectra are almost identical and drop smoothly with wavenumber k. Hence, the fitted smooth potentials can neither absorb nor introduce any spurious $\delta\psi$.
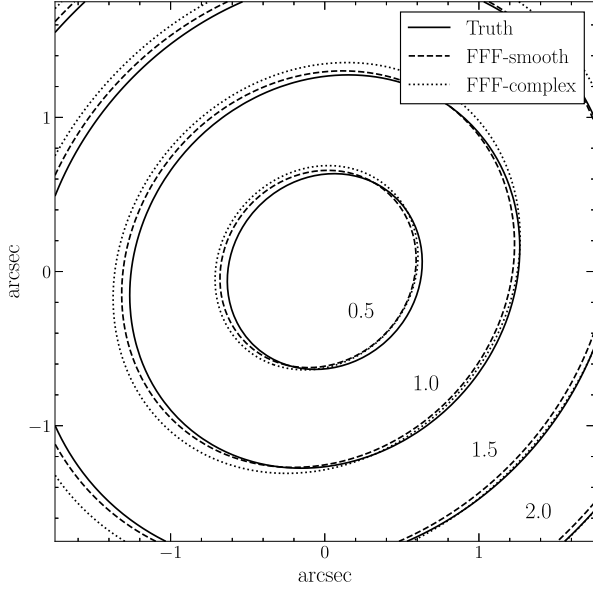
Figure B1. Contours of the true underlying smooth potential described in Section 3.1, as well as its MAP fits by the FFF-smooth and FFF-complex models (see Table 3). These models simultaneously fit the smooth potential and reconstruct its perturbations $\delta\psi$.
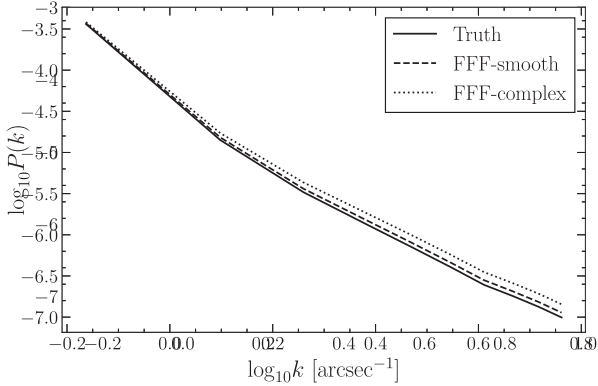

Figure B2. Fourier power spectrum of the smooth potentials shown in Fig. B1.

APPENDIX C: PARAMETER VALUES OF A SMOOTH POTENTIAL FITTED TO A PERTURBED SYSTEM

In Table C1, we list the mean and 68 per cent confidence intervals of the smooth SIE potential parameters, $\boldsymbol{\eta}$, obtained from the purely smooth models fitted to perturbed systems as presented in Sections 3.3 and 3.4.

Table C1. Mean values and 68 per cent confidence intervals for the lens potential ($\boldsymbol{\eta}$) and source regularization parameters (λ_s, $\boldsymbol{g}_s$) for the smooth models fitted to the perturbed mocks, as described in Sections 3.3 and 3.4, compared to the truth.

Name	Units	Truth	SMOOTH (Section 3.3)	SMOOTH $n = 1$ (Section 3.4)
b	arcsec	0.9	$0.926^{+0.002}_{-0.002}$	$0.897^{+0.004}_{-0.004}$
q	–	0.8	$0.856^{+0.003}_{-0.003}$	$0.803^{+0.008}_{-0.008}$
θ	°	-135	$-148.153^{+1.036}_{-0.754}$	$-136.765^{+0.012}_{-0.012}$
x_0	arcsec	0	$-0.021^{+0.000}_{-0.000}$	$-0.036^{+0.001}_{-0.001}$
y_0	arcsec	0	$0.019^{+0.003}_{-0.002}$	$0.031^{+0.001}_{-0.001}$
γ	–	0.03	$0.056^{+0.000}_{-0.000}$	$0.039^{+0.002}_{-0.002}$
ϕ	°	-40	$-36.599^{+0.851}_{-1.151}$	$-42.405^{+0.628}_{-0.628}$
λ_s	–	–	$11.181^{+1.020}_{-1.020}$	$71.368^{+15.930}_{-15.930}$
l_s	arcsec	–	$0.175^{+0.007}_{-0.007}$	$0.062^{+0.049}_{-0.049}$

This paper has been typeset from a TeX/LaTeX file prepared by the author.