FERMILAB-PUB-21-598-PPD

The challenge of blending in large sky surveys

Peter Melchior, Rémy Joseph, Javier Sanchez, Niall MacCrann and Daniel Gruen

Abstract | The increasing sensitivity of modern sky surveys allow ever fainter emissions of light to be detected, but it also increases the chances of noticeable overlap between multiple sources of light, a phenomenon called blending. The consequences of blending are expected to be among the leading systematic measurement uncertainties of future surveys, such as the Legacy Survey of Space and Time. This Perspective discusses two main approaches to addressing blending: attempting to separate individual sources and statistically correcting for the presence of blending at the population level. For both approaches, simultaneous access to data of multiple surveys will be critical to construct a joint data set that combines the strengths of each individual survey.

Blending occurs when multiple sources of light occupy the same region of the sky. If left unaccounted for, blending results in contaminated measurements of celestial sources that are traditionally considered isolated. Its consequences propagate into and alter the estimates of physical processes that are under investigation, and are expected to be among the leading systematic uncertainties of many scientific investigations for future surveys like the Legacy Survey of Space and Time (LSST) of the Vera C. Rubin Observatory.

The rate at which blending happens depends on two main parameters. First and foremost, the sensitivity (also called depth) of the observations determines how many emitters are detected per unit sky area. At optical and near-infrared wavelengths, shallow observations primarily show stars in the Milky Way, whereas source counts in deeper observations are dominated by galaxies at cosmological distances.

Second, the spatial resolution of the instrument determines the minimal unit of sky area in which multiple sources can be distinguished from each other. For modern telescopes, the resolution limit is typically determined by the 'seeing', the blurring caused by a turbulent atmosphere. In its absence, that is, with a ground-based adaptive optics system or a space telescope, the resolution is set by the diffraction limit of the telescope optics.

Modern ground-based surveys — like the Dark Energy Survey (DES)[1], the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP)[2] and the future LSST[3] — have been designed, among other criteria, to maximize the number of galaxies for which fluxes, sizes and ellipticities can be measured. Their observations are so deep that they reveal most stars in our Galaxy and an even larger number of galaxies. With the number density of sources on the sky, blending becomes inevitable, given the seeing conditions and sensitivity of these surveys (FIG. 1). On top of that, galaxies are spatially extended, that is, their apparent size is larger than the resolution limit of the instrument, which leads to even higher rates of blending. As a result, these surveys find 30–60% of all celestial sources to be blended[4–6]. Although each study adopts different definitions of blending, driven by their respective measurement methodology and scientific requirements, the blending rates consistently grow with increasing sensitivity.

The problems for follow-up analyses in astrophysics and cosmology that arise from blending are twofold. First, traditional measurement algorithms in astronomy assume independence of sources. In other words, they expect that a large sky area can be broken up into smaller chunks so that, in each chunk, there is only one source to be analyzed. If this assumption is violated, measurements, such as those of the flux or shape of a source, become contaminated[6,7]. Rejecting blended galaxies is possible, but not advisable. Depending on the adopted criterion, rejecting blended galaxies results in a loss of statistical power of a few percent to tens of percent for an LSST-like programme of measuring galaxy shapes for weak gravitational lensing[8], which defies the original intent of deeper observations. Moreover, such a scheme would preferentially reject galaxies in regions of high source density, with possible biases arising for cosmological applications[9]. We argue, in the next section, that simultaneous modelling multiple sources even inside one sky chunk can mitigate these measurement biases.

Second, a more pernicious scenario arises when multiple sources are so close to each other that a given observation and measurement method cannot recognize their presence individually. Unrecognized blending affects about 15% of all galaxies in an LSST-like survey[10] and creates strong outliers and systematic biases in ellipticity measurements[10–12]. It renders rejection and simultaneous modelling efforts formally impossible (but there is a remedy, as discussed in the next section) and calls for a subsequent, higher-level statistical correction of measurement biases (as discussed in the section on characterizing blending) or the integration of multiple data sets, especially the combination of ground-based and space-based surveys (as discussed in the Outlook).

BOX 1 presents a visual summary of the main effects and severity of blending for an LSST-like imaging survey.

Deblending efforts

Blending poses an inverse problem, known as 'deblending' in astronomy and 'source separation' in general. The goal of deblending is to reconstruct the properties of the individual sources from a combined, blended observation. Strictly speaking, the task is impossible because the problem is underconstrained. For any recorded photon, it is simply not known which source it came from. In general, the problem has k times

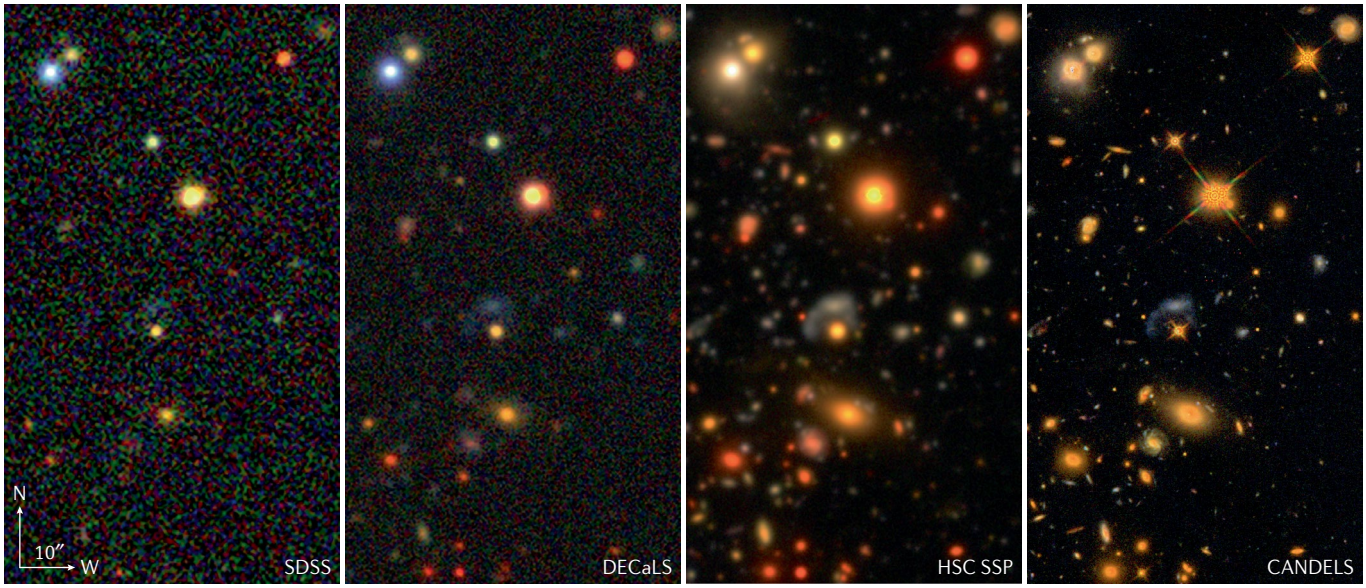


Fig. 1 | **The same sky region of 1.5 × 0.75 arcmin², observed by different surveys.** From left to right: the Sloan Digital Sky Survey (SDSS)[91] with a pixel scale of 0.396 arcsec, the DECam Legacy Survey (DECaLS)[92] with a pixel scale of 0.262 arcsec, the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP)[2] with a pixel scale of 0.168 arcsec and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) programme[93,94] with the Hubble Space Telescope and a pixel scale of 0.06 arcsec. Increasing the sensitivity yields a larger number of sources and reveals the fainter outskirts galaxies, both of which result in higher rates of blending. At fixed depth, increasing the spatial resolving power is the only means of reducing source confusion. SDSS and DECaLS images adapted with permission from Legacy Surveys/D. Lang (Perimeter Institute). HSC SSP image adapted with permission from NAOJ/HSC Collaboration.

as many degrees of freedom as constraints from data, where k is the number of blended sources. This scaling can be seen for imaging instruments, in which case one image per source is to be reconstructed, as well as for photon-counting instruments, in which case mixture modelling estimates a probability for every photon of being associated with any of the k sources[13]. To overcome the inherent degeneracies of this underconstrained inverse problem, one must make assumptions about the kind of blending and the properties of the sources being blended.

Virtually all deblending approaches adopt a linear mixing model, in which multiple sources simply add their respective emissions along the line of sight. This model is mathematically and practically the easiest one, and allows for a variety of linear methods to be applied. It assumes mutual independence between sources, which seems justified in astronomy because stars and galaxies that appear blended are very often not physically interacting but merely aligned from our perspective. In detail, the assumption is invalid because dust in a foreground galaxy absorbs and re-emits light from any background source. Fortunately, galaxies are only moderately opaque[14] — at least in the nearby universe — so that dust attenuation can be ignored to the first order.

One can further assume knowledge of the spatial light distribution of the sources. For instance, stars are point sources, the apparent shape of which is entirely determined by the point spread function of the telescope, so that only their luminosity and position have to be determined. This simplification allows for successful deblending even in highly crowded stellar fields[15–17]. Galaxies exhibit complex morphologies, and, thus, require additional assumptions. That such assumptions are at least approximately justified is supported by the long-standing realization that galaxy morphologies are not arbitrary but form groups, most prominently those of elliptical and spiral galaxies[18].

The main morphological features of many galaxies can, thus, be described by simple parametric models[19] or their approximations[20,21]. Stabilizing and speeding up the solutions for multi-source deblending usually requires iterative approaches, wherein the parametric model is fit to only one source at a time. One can then either mask pixels deemed to belong to other sources in the blend or subtract the assumed pixel values according to an earlier fit of these other sources[22–24]. Both approaches directly exploit the assumed independence between sources.

Non-parametric approaches can solve for multiple sources at once. A suitable approach that implements a linear mixing model and can formally handle an arbitrary number of sources was implemented in the Sloan Digital Sky Survey (SDSS) photo pipeline[25]. This approach can be extended with an approach based on matrix factorization[26,27]. To reduce the number of degrees of freedom, one can enforce morphological heuristics — for instance, that galaxies are symmetric under 180° rotation or are monotonically decreasing from their centres, through constrained optimization[28,29].

Recent advances in machine learning allow non-parametric methods to learn commonalities among observed galaxies, thereby, extending and generalizing heuristic constraints. Generative neural networks are capable of representing a wide range of galaxy morphologies. They are well-suited for deblending because they learn an encoding from which any observed galaxy, or at least its relevant aspects, can be reconstructed. A common architecture is the autoencoder[30,31], which can be trained to encode isolated galaxy images into a low-dimensional latent space and then to decode, that is, reconstruct, the image from the latent variables. Knowing how any galaxy can look (the task for the decoder, also called generator) strongly reduces the degeneracies that arise from the overlap of multiple galaxies: many deblending solutions may add up to a given blend, but the individual sources do not look like real galaxies and, therefore, can be dismissed.

Such generative models have been used in several deblending studies. For example,

the generator of an autoencoder trained on isolated galaxies can be retained while the encoder is retrained on blended images[32]. Doing so changes the task of the encoder to learn how to robustly identify the latent variables of the most prominent galaxy, while ignoring all others. The approach exploits an insight from blending statistics, namely, that configurations with one bright source and one or several much fainter sources are much more common than configurations with equally bright galaxies (BOX 1). The approach, thus, seeks to recover the most prominent source in the most common blend scenario. Going one step further, one can use an autoencoder extension, called a U-Net[33], to break up the image into multiple output channels of the same shape and obtain solutions for both sources in a two-source blend[34], which, for intermediate to faint galaxies, is the most common configuration (BOX 1).

Both these approaches are somewhat sensitive to the prevailing observational conditions (such as image resolution, the shape and width of the point spread function, sky brightness and its spatial variations, and detector artifacts). Modern surveys invest in careful characterization of these conditions, which suggests a strategy of explicitly modelling them as degradations of an ideal image. Doing so leaves for the network only the task of encoding galaxy images in idealized form — that is, noise-free, super-resolution and unconvolved[35]. Combining this strategy with the linear mixing model yields a hybrid Bayesian scheme that has an explicit form of the blending likelihood and a probabilistic generative network (of the PixelCNN++ architecture[36], for example) as a prior over all galaxy morphologies[37]. This direction appears promising thanks to deep learning methods gaining further capabilities and maturity, but care needs to be taken to train on realistic, high-quality images and to account for morphological changes of galaxies observed at different wavelengths.

Deblending approaches typically require knowledge of how many sources are present in a blend because they can only model the previously identified ones. This requirement renders deblending dependent on the task of source detection, which is performed prior to deblending. Whereas statistically optimal detection methods have been formulated for point sources[38–41], direct solutions for extended blended sources do not exist. Convolutional neural networks provide an implicit solution of object detection and classification, for instance, with the Mask R-CNN[42] architecture, and show good performance on images of moderate source density[43]. Increasing the density further makes detection and deblending exponentially more difficult[44], ultimately leading to the so-called confusion limit. At this sensitivity of a survey, further detections become statistically impossible due to the Poisson noise from the collective light emitted by all sources. The increasing depth of modern surveys requires techniques that are capable of operating close to the confusion limit. One promising approach is to break the interdependency of detection and deblending by exploiting the power-law shape of galaxy luminosity functions. Performing detections and deblending on the rarer bright sources first and then running a conventional or Mask R-CNN detection method on the residuals of the deblending model allows the detection of previously missed sources in heavily blended regions[45]. Such an iterative detection catalogue increases source detection completeness in blends and, thus, sharply reduces the rate of unrecognized blending.

Although most detection methods have been trained on simulated images resembling the observations of particular

Box 1 | **Blending characteristics for the LSST**

We created realistic simulated images that mimic the observing conditions of the Legacy Survey of Space and Time (LSST), using a cosmological simulation that determines the location, brightness and morphology of all sources. We then process these images with the LSST data management software stack and register when a known input source is contained within the area of a detected source in the image. Unlike studies on observations, we can unambiguously determine if blending has happened.

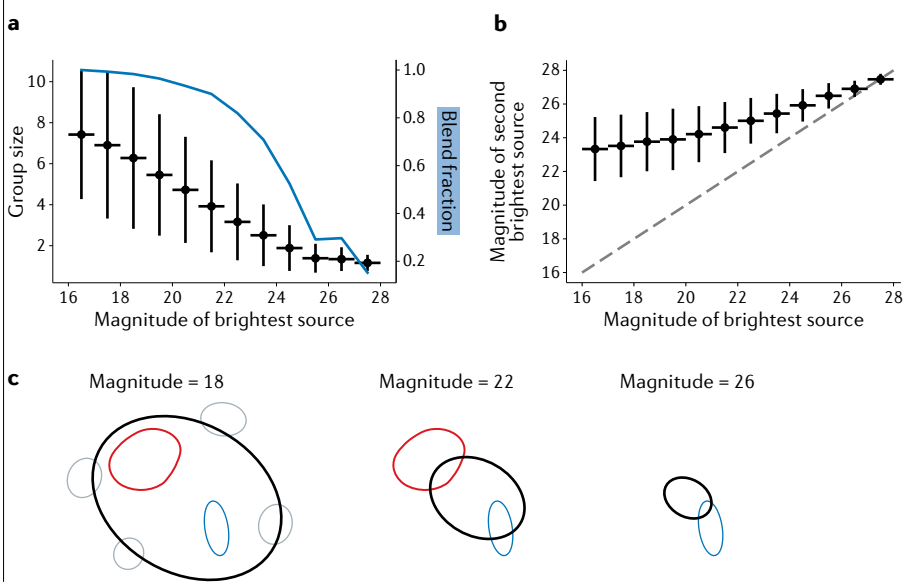
Owing to their apparent size on the sky, the brightest galaxies (*i*-band magnitude ~16) are effectively guaranteed to be blended (blue line in figure part **a**), typically forming groups of 8 ± 3 sources (black data points in figure part **a**), with the second brightest source in such a group usually being 8 magnitudes (that is, a factor of 1,500) fainter (figure part **b**). The measurement of the brightest sources is, thus, mostly unaffected by blending, but blending almost certainly hinders the detection and measurement of fainter sources in their vicinity.

Galaxies with intermediate brightness still have blending rates of more than 50% and group sizes of 2–4 (figure part **a**), but if they are blended, the secondary source is only 2–4 magnitudes (factors 6–40) fainter (figure part **b**). This is the regime in which blending is still very common and can be very noticeable in measurements.

At the faintest end, sources appear very small and blending becomes rare (figure part **a**), but if it happens, it involves another source of almost equal brightness (figure part **b**). This is the regime in which measurements, and even detection probabilities, are most strongly affected by blending.

A sketch to summarize these findings from blending statistics is shown in figure part **c**. Because brighter galaxies are intrinsically larger and their outskirts remain visible further away from their centres, they occupy more area on the sky (not shown to scale). With the same secondary source population in all three cases, blended groups get smaller with decreasing brightness of the primary source, but the secondaries become similarly bright.

In figure parts **a** and **b**, vertical error bars denote symmetrical 68% confidence intervals and horizontal error bars indicate the bin width of 1 magnitude. The dashed line in part **b** indicates equality between brightest and second brightest sources as a guide to the eye.



surveys, actual observations can be used instead if reliable detection catalogues are available. Doing so allows a U-Net to learn the separation between sources and the sky background. If the sources have also been assigned a classification label, one can further leverage the U-Net architecture and provide an image per label[46]. By running on multi-band images, this semantic segmentation approach classifies individual pixels into background, point source and several galaxy types, and is, thus, able to detect sources and possibly provide a physically motivated deblending solution based on morphological and colour features. The main challenge in such a scheme lies in the reliance on externally constructed detections and classifications. For both semantic segmentation and iterative detection and deblending, it is difficult to know when to stop, that is, to determine if a perceived variation in the data justifies reporting of a new source or different galaxy label.

Deblending can also be performed in the spectral domain. Given an assumed set of spectra that can be exhibited by sources, one can ask whether an observed spectrum is more likely to have come from a single source or a combination of sources. One can also determine the most likely spectra of each source using a statistical hypothesis test[47]. Furthermore, one can combine assumptions about spatial and spectral properties[28,48]. Deblending with spatial and spectral information — known as hyper-spectral modelling — exploits all separating features available for non-variable sources, and we are convinced of its effectiveness and feasibility. The current shortcoming of hyper-spectral modelling lies in the limitations in understanding of how various astrophysical processes shape the properties of galaxies. Comparisons between large hyper-spectral data sets[49–51] and galaxy-scale simulations[52–54] could lead to the definition of observationally recognizable galactic 'building blocks', such as bulges, bars, spiral arms, star-forming regions and tidal tails. Such building blocks are interesting in their own right as constraints on models for galaxy formation and evolution. Furthermore, a library or deep-learning model of them would make it possible to identify spectrally or spatially distinct structures and differentiate those that emerge within a single galaxy from those that are formed by the superposition of different galaxies. We believe that investigations in this direction are the next logical step in the development of hyper-spectral deblending.

Blending characterization and mitigation

An alternative to deblending is to characterize the effects of blending at the statistical level. Blending affects any pixel-level measurement, and subsequent scientific analyses, by altering one's ability to detect sources or to measure their properties cleanly. When deblending is not feasible at a level of systematic uncertainty subdominant to the statistical error budget of large surveys, such a characterization is essential.

Deep, space-based observations could provide reliable 'ground truth' information for blending studies[10] (FIG. 1), but, so far, have limited sky coverage. Therefore, simulations are critical for blending characterization[6,8]. Substantial simulation efforts are being made for the upcoming LSST[55,56], as well as ESA's Euclid mission[12,57] and NASA's Nancy Grace Roman Space Telescope[58]. In images generated from cosmological *N*-body simulations, the assumed galaxy morphologies either follow parametric Sérsic models[19] or are drawn from a limited number of high-fidelity images from the Hubble Space Telescope. The former lack the rich morphological diversity of galaxies, whereas the latter only provide a small sample of galaxy shapes.

The morphological complexity of simulated galaxies is likely to increase in the near future thanks to advances in neural network generators[35]. In addition, image simulations coming directly from cosmological hydrodynamics simulations[59] could soon provide fully realistic mock observations at the scale of wide-field surveys. Ideally, such simulated images would include complex dependencies, for instance, those between the clustering of galaxies and their spectral and morphological characteristics[52]. In the meantime, the lack of realism of ab initio simulations can be partially remedied by injecting mock sources into observations and measuring how well they are recovered[7,60,61]. Although the mock sources themselves are drawn from a parametric model or a small set of high-fidelity images, the properties of all other sources are, by construction, correct.

Regardless of the principle underlying the blending characterization, the assumption is made that measurements on observations exhibit the same kinds of difficulties, so that the corrections needed in practice can be directly read off from these prior characterization efforts. We detail blending-related approaches for the most important measurements below.

Detection. Detection completeness is a complex interplay of the concrete arrangement of sources in a sky region and the data processing pipeline. For isolated galaxies, the ability to detect is primarily limited by Poisson shot noise. For faint galaxies, this noise is dominated by the so-called sky brightness, which is a combination of several sources, such as sunlight scattered off the Moon or off interplanetary dust grains, airglow of the atmosphere. The increased shot noise from blending of detected and undetected sources does not itself noticeably impede detection, increasing the effective noise level for the LSST by only about 0.2%[4]. The dominating effect is the difficulty of detecting fainter sources in the vicinity of brighter ones, driven by algorithmic choices such as requiring a minimum measure of separability to consider the fainter source a legitimate detection. The detection completeness of potentially detectable sources, that is, sources that would have been detectable if they were isolated, is reduced by blending by approximately 10% in the DES[61] and by 20% in the LSST[4]. Source injection tests in deep observations from the Hubble Space Telescope similarly show that blending reduces stellar completeness from 1% at the brighter end to 20% close to the detection limit[62].

Source injection methods are also well suited to correct for the cumulative effects on the completeness of detection catalogues. A case study for angular clustering of galaxies in the DES[60] shows an excess signal by more than a factor of 2 on scales below 7 arcsec, attributed to blending. The clustering signal is expressed as the average probability of finding a source within an angular separation from another source, and is, thus, directly modulated by any change to the detection probability. This counter-intuitive excess possibly originates from the light envelope of bright sources aiding the detection of fainter ones in their vicinity. Injecting sources provides an estimate of the non-uniform detection probability of the survey and allows for such spurious contributions to be removed.

Photometry and photometric redshifts. Photometry — that is, the measurement of source flux, typically in multiple filter bands — is the basis for many statements about the physical state of a source, such as the temperature of a star or the stellar mass of a galaxy. In addition, the apparent colour of galaxies yields an estimate of their distance through the redshifting of photons caused by cosmic expansion. This approach,

called photometric redshift estimation, provides distance estimates for the vast majority of extragalactic sources.

Simulated images of Sérsic galaxy models with parameters resembling the expected galaxy population that will be observed by the LSST show that 62% of the simulated sources have at least 1% flux contribution from overlapping sources in the *i* band[4]. This fraction is reduced to 51% when only considering contaminating sources that are sufficiently bright to be individually detectable, highlighting the influence of a non-detectable source population that creates a faint but spatially structured photon background[63]. Even when blending is recognized, source injection tests show that galaxy photometry is affected at a level of 5%; stellar photometry even by 20%, largely independent of filter band[7]. In rare cases, blending causes catastrophic outliers in galaxy photometry, resulting in gross overestimation of source fluxes [61]. Even faint levels of flux contamination are problematic when they change the colour and, thus, the photometric redshift of a galaxy. In crowded fields, for instance, near the cores of clusters, such colour changes can happen at levels exceeding the requirements of next-generation cosmological experiments[64].

Morphology and weak gravitational lensing. One of the key aims of current and upcoming imaging surveys is to use gravitational lensing to map the distribution of matter in the Universe. Gravitational lensing is the bending of light from distant sources, primarily high-redshift galaxies, by the gravity of intervening mass. For most galaxies, this bending manifests itself as a slight change in size and ellipticity, a phenomenon called weak lensing. The primary effect, caused by the so-called shear $\boldsymbol{\gamma}$, changes the 2D ellipticity $\mathbf{e}$ of a source as

$$\mathbf{e} = \mathbf{e}\,|_{\gamma=0} + \frac{\partial \mathbf{e}}{\partial \boldsymbol{\gamma}}\,|_{\gamma=0}\boldsymbol{\gamma} + \mathcal{O}(\boldsymbol{\gamma}^2) \approx \mathbf{e}\,|_{\gamma=0} + \mathbf{R}\boldsymbol{\gamma}, \quad (1)$$

where $\mathbf{R} = \partial\mathbf{e}/\partial\boldsymbol{\gamma}$ is called the (linear) shear response.

Accurate measurements of the apparent ellipticity, thus, carry the imprints of cosmological lensing and give information about the properties of dark matter and dark energy. In addition, the lensing effect depends on the distance between the observer, the 'lens' that causes the light deflection and the source from which the light was initially emitted. To constrain the 3D structure of the cosmic matter field, one must, therefore, estimate the redshift z of the source galaxies, or, at least, the distribution of their redshifts $n(z)$. Weak lensing, therefore, requires accurate morphological measurements, as well as photometry for photometric redshift estimates.

Blending presents various challenges for weak lensing measurements. Ellipticity measurements are even more sensitive to blending than photometry because the effective influence of a pixel is proportional to the squared distance from the centre of the source. As a result, blending tends to align the semi-major axis of the best-fitting ellipse towards the neighbouring source. Unrecognized blending (identifiable in ground-based images by comparison with high-resolution space-based images of the same sky region) causes large outliers in ellipticity measurements and increases the overall uncertainties in ellipticities by 14%[10]. This number is confirmed by a statistical accounting of the influence of blends on ellipticity noise in simulated images that resemble those from the LSST[4]. The presence of non-detected objects also causes systematic biases in the amplitude of ellipticity measurements at the percent level[11,65,66]. Even if the neighbouring sources are completely identified, ellipticity and size measures remain uncertain and are usually biased towards large sizes and high ellipticities[67], because deblending algorithms are fundamentally imperfect (as discussed in the section on deblending efforts).

The high accuracy requirements of weak lensing suggest an approach where the measurements are made as robust to blending as possible, and residual biases are corrected statistically. Doing so by simulating every conceivable shape and brightness of a galaxy is time-consuming and sensitive to the underlying assumptions about the relevant galaxy parameters[11,68].

An alternative approach, called metacalibration, stems from recognizing that weak lensing is solely described by the change of ellipticity or sizes, with unlensed galaxy properties acting as a constant additive term (see Eq. (1)). Because galaxies have no preferred orientation, the constant term vanishes when averaging over many galaxies, so that $\langle\boldsymbol{\gamma}\rangle \approx \langle\mathbf{R}\rangle^{-1}\langle\mathbf{e}\rangle$. Calibration, thus, only needs to accurately determine the average of the response matrix $\mathbf{R}$, that is, how the measurements of ellipticities respond to a change in the shear, for instance, by artificial shearing (or stretching, for size-based lensing estimates) of sources in observed images and recording how their measurements change[69,70]. This approach can be performed on the actual galaxies in question, instead of simulations, and encompasses all stages of the data processing pipeline. As a result, metacalibration yields ellipticity measurements that are corrected to the very stringent requirements of upcoming weak lensing analyses[70].

This result remains nominally true for blended galaxies, even if deblending is not attempted[66]. Metacalibration can further be extended to include a separate detection step of the manipulated images, a technique dubbed metadetection, which can suppress biases arising from the interdependencies of detection and ellipticity measurements. For observations from the DES, the LSST and Euclid, these biases can be suppressed from a few percent to a few parts in 10,000 (REFS[66,71]).

The remaining problem for shape measurements arises because multiple blended galaxies are rarely physically associated, but are, instead, line-of-sight projections. As a result, the photons from individual galaxies experience different amounts of lensing on their way to the observer. In addition, weak lensing analyses require redshift estimates — typically from photometry, which is also affected by blending. Both of these effects can be summarized at the level of the redshift distribution of sources, $n(z)$, which ordinarily denotes the number of galaxies in a survey at a given redshift. However, any blended source with photons originating from different redshifts effectively contributes to $n(z)$ at multiple redshifts. The shear measurement response also modulates the effective distribution. The combined effect can be treated by constructing an effective redshift distribution for weak lensing $n_\gamma(z)$. The shape of this distribution can be determined from dedicated image simulations, which compute the measured ellipticity of a source when introducing an artificial shear at a redshift z. These simulations need to be populated with a realistic distribution of galaxies and must vary the lensing strength as a function of redshift[72]. Using $n_\gamma(z)$ instead of $n(z)$ provides a simple plug-in replacement for follow-up studies that accounts for the combined effects of blending on both the shape measurements and the redshift estimates. However, it requires many more calibration simulations compared with conventional constant-shear calibrations, which poses a substantial computing challenge to large surveys.

Outlook

With deblending methods continually improving and statistical mitigation approaches, in some cases, reaching levels

of calibration sufficient for upcoming cosmological studies, what is there left to do?

Many astronomical studies rely on accurate results for individual sources, often in configurations with a high probability of blending. Examples are galaxy clusters or extended low-surface-brightness galaxies[73–75]. The best way to aid the impossible task of deblending in these cases is to use superior data. Assessing detection completeness and photometry in ground-based imaging by comparing with space-based observations of the same region[10,76] is routinely done if overlapping data are available (FIG. 1). Such overlapping data, however, are still scarce.

For instance, the entire sky area covered by the Hubble Space Telescope amounts to only about 0.1% of the sky (Rick White, personal communication), with the largest contiguous area, the so-called Cosmic Evolution Survey (COSMOS) field[77,78], fitting comfortably in a single pointing of many modern survey telescopes. However, this situation will change in this decade. The launches of Euclid in 2022, of the Chinese Survey Space Telescope (CSST) in 2024 and of the Roman Space Telescope in 2025 complement the ground-based Rubin Observatory with three space-borne instruments that are optimized for wide-field surveys. The availability of overlapping imaging surveys will extend the wavelength coverage in the ultraviolet and near-infrared range. It also spurs the development of data processing approaches that either use high-resolution images as templates for more blended low-resolution data[79,80] or fully exploit the combined data set by jointly processing all available observations at the pixel level[81–84].

Statistical mitigation of blending can be remarkably accurate, but photometric measurements will remain challenging. One can read off blending-induced photometry biases from simulations, but the calculated biases are strongly dependent on the parameters of the image simulation or of the injected sources[7,72]. In addition to computational limitations, the critical issue is, thus, the realism of these simulations, specifically regarding the joint distributions of morphologies, spectra and redshifts.

The most convincing way forward again employs the use of superior data. Ideally, one would construct an overlapping data set from the most capable instruments in all relevant wavelengths, so that shortcomings of one data set can be compensated by another, as has routinely been done with follow-up spectroscopic campaigns of large imaging surveys[85–87]. But the instruments in question differ in their capabilities, observational restrictions and in the demands of other programmes they are intended to carry out. Full overlap over all of the sky is, thus, not attainable in the foreseeable future.

As a result, proposed agreements between multiple surveys are critical. Spending some fraction of their respective time budget on observations of common sky regions, the so-called deep fields[88–90], will make it possible to combine all available information. With common sky coverage from ground and space soon reaching tens of square degrees, these carefully chosen fields will be large enough to provide a representative sample of the complex joint distributions of galaxy properties. They are also premier target fields for deep spectroscopic follow-up programmes, either with moderately high resolution spectrographs from the ground or the upcoming slitless spectroscopy instruments of Euclid, CSST and Roman. Besides the immediate scientific opportunities, the analysis of these joint deep fields will be invaluable to test deblending methods, determine blending contaminations to various measurements without relying on simulations and provide training data for neural network generators.

Blending poses a profound challenge for the processing and analysis of data from deep sky surveys, the scope of which has only recently been fully recognized. We are optimistic that this challenge can be met with methodological advances and cross-survey cooperation.

Peter Melchior[1,2]✉, *Rémy Joseph*[1], *Javier Sanchez*[3,4], *Niall MacCrann*[5] *and Daniel Gruen*[6]

[1]*Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA.*

[2]*Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA.*

[3]*Fermi National Accelerator Laboratory, Batavia, IL, USA.*

[4]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL, USA.*

[5]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK.*

[6]*University Observatory Munich, Faculty of Physics, Ludwig-Maximilians-Universität München, Munich, Germany.*

✉*e-mail: peter.melchior@princeton.edu*

https://doi.org/10.1038/s42254-021-00353-y

Published online: 18 August 2021

1. Dark Energy Survey Collaboration et al. The Dark Energy Survey: more than dark energy – an overview. *Mon. Not. R. Astron. Soc.* **460**, 1270–1299 (2016).
2. Aihara, H. et al. The Hyper Suprime-Cam SSP survey: overview and survey design. *Publ. Astron. Soc. Jpn.* **70**, S4 (2018).
3. Ivezić, Ž. et al. LSST: from science drivers to reference design and anticipated data products. Preprint at *arXiv* http://arxiv.org/abs/0805.2366 (2008).
4. Sanchez, J., Mendoza, I., Kirkby, D. P. & Burchat, P. R. Effects of overlapping sources on cosmic shear estimation: Statistical sensitivity and pixel-noise bias. Preprint at *arXiv* http://arxiv.org/abs/2103.02078 (2021).
5. Bosch, J. et al. The Hyper Suprime-Cam software pipeline. *Publ. Astron. Soc. Jpn.* **70**, S5 (2018).
6. Samuroff, S. et al. Dark energy survey year 1 results: the impact of galaxy neighbours on weak lensing cosmology with IM3SHAPE. *Mon. Not. R. Astron. Soc.* **475**, 4524–4543 (2018).
7. Huang, S. et al. Characterization and photometric performance of the Hyper Suprime-Cam software pipeline. *Publ. Astron. Soc. Jpn.* **70**, S6 (2018).
8. Chang, C. et al. The effective number density of galaxies for weak lensing measurements in the LSST project. *Mon. Not. R. Astron. Soc.* **434**, 2121–2135 (2013).
9. Hartlap, J., Hilbert, S., Schneider, P. & Hildebrandt, H. A bias in cosmic shear from galaxy selection: results from ray-tracing simulations. *Astron. Astrophys*. **528**, A51 (2011).
10. Dawson, W. A., Schneider, M. D., Anthony Tyson, J. & James Jee, M. The ellipticity distribution of ambiguously blended objects. *Astrophys. J.* **816**, 11 (2015).
11. Hoekstra, H., Viola, M. & Herbonnet, R. A study of the sensitivity of shape measurements to the input parameters of weak-lensing image simulations. *Mon. Not. R. Astron. Soc.* **468**, 3295–3311 (2017).
12. Martinet, N. et al. Euclid preparation - IV. impact of undetected galaxies on weak-lensing shear measurements. *Astron. Astrophys*. **627**, A59 (2019).
13. Melchior, P. & Goulding, A. D. Filling the gaps: Gaussian mixture models from noisy, truncated or incomplete samples. *Astron. Comput.* **25**, 183–194 (2018).
14. Calzetti, D. The dust opacity of star-forming galaxies. *Publ. Astron. Soc. Pac.* **113**, 1449 (2001).
15. Stetson, P. B. DAOPHOT: a computer program for crowded-field stellar photometry. *Publ. Astron. Soc. Pac.* **99**, 191 (1987).
16. Linde, P. High precision crowded field photometry. *Highlights Astron.* **8**, 651–656 (1989).
17. Feder, R. M., Portillo, S. K. N., Daylan, T. & Finkbeiner, D. Multiband probabilistic cataloging: a joint fitting approach to point-source detection and deblending. *Astron. J.* **159**, 163 (2020).
18. Hubble, E. P. Extragalactic nebulae. *Astrophys. J.* **64**, 321–369 (1926).
19. Sérsic, J. L. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Bol. Asoc. Argent. Astron. Plata Argent.* **6**, 41–43 (1963).
20. Spergel, D. N. Analytical galaxy profiles for photometric and lensing analysis. *Astrophys. J. Suppl. Ser.* **191**, 58 (2010).
21. Hogg, D. W. & Lang, D. Replacing standard galaxy profiles with mixtures of Gaussians. *Publ. Astron. Soc. Pac.* **125**, 719 (2013).
22. Bertin, E. & Arnouts, S. SExtractor: software for source extraction. *Astron. Astrophys. Suppl. Ser.* **117**, 393–404 (1996).
23. Jarvis, M. et al. The DES science verification weak lensing shear catalogues. *Mon. Not. R. Astron. Soc.* **460**, 2245–2281 (2016).
24. Drlica-Wagner, A. et al. Dark energy survey year 1 results: The photometric data set for cosmology. *Astrophys. J. Suppl. Ser.* **235**, 33 (2018).
25. Stoughton, C., Lupton, R. H., Bernardi, M., Blanton, M. R. & Burles, S. Sloan digital sky survey: early data release. *Astron. J.* **123**, 485–548 (2002).
26. Paatero, P. & Tapper, U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* **5**, 111–126 (1994).
27. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
28. Melchior, P. et al. Scarlet: source separation in multi-band images by constrained matrix factorization. *Astron. Comput.* **24**, 129–142 (2018).
29. Melchior, P., Joseph, R. & Moolekamp, F. Proximal Adam: robust adaptive update scheme for constrained optimization. Preprint at *arXiv* https://ui.adsabs.harvard.edu/abs/2019arXiv191010094M (2019).
30. Hinton, G. E. & Zemel, R. in *Advances in Neural Information Processing Systems* Vol. 6 (eds Cowan, J. D., Tesauro, G. & Alspector, J.) 3–10

(Morgan-Kaufmann, 1994). https://proceedings.neurips.cc/paper/1993/file/9e3cfc48eccf81a0d57663e129aef3cb-Paper.pdf.
31. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. Preprint at *arxiv* http://arxiv.org/abs/1312.6114v10 (2013).
32. Arcelin, B., Doux, C., Aubourg, E. & Roucelle, C. Deblending galaxies with variational autoencoders: a joint multiband, multi-instrument approach. *Mon. Not. R. Astron. Soc.* **500**, 531–547 (2021).
33. Ronneberger, O., Fischer, P. & Brox, T. in *International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241 (Springer, 2015). https://doi.org/10.1007/978-3-319-24574-4_28.
34. Boucaud, A. et al. Photometry of high-redshift blended galaxies using deep learning. *Mon. Not. R. Astron. Soc.* **491**, 2481–2495 (2020).
35. Lanusse, F. et al. Deep generative models for galaxy image simulations. *Mon. Not. R. Astron. Soc.* **504**, 5543–5555 (2021).
36. Salimans, T., Karpathy, A., Chen, X. & Kingma, D. P. PixelCNN++: Improving the PixelCNN with discretized logistic mixture likelihood and other modifications. Preprint at *arXiv* http://arxiv.org/abs/1701.05517 (2017).
37. Lanusse, F., Melchior, P. & Moolekamp, F. Hybrid physical-deep learning model for astronomical inverse problems. Preprint at *arXiv* http://arxiv.org/abs/1912.03980 (2019).
38. Kaiser, N. Addition of images with varying seeing. Technical report http://pan-starrs.ifa.hawaii.edu/project/people/kaiser/imageprocessing/im++.pdf (2004).
39. Zackay, B. & Ofek, E. O. How to COAAD images. I. Optimal source detection and photometry of point sources using ensembles of images. *Astrophys. J.* **836**, 187 (2017).
40. Daylan, T., Portillo, S. K. N. & Finkbeiner, D. P. Inference of unresolved point sources at high galactic latitudes using probabilistic catalogs. *Astrophys. J.* **839**, 4 (2017).
41. Liu, R., McAuliffe, J. D. & Regier, J. Variational inference for deblending crowded starfields. Preprint at *arXiv* http://arxiv.org/abs/2102.02409 (2021).
42. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. Preprint at *arXiv* http://arxiv.org/abs/1703.06870 (2017).
43. Burke, C. J. et al. Deblending and classifying astronomical sources with Mask R-CNN deep learning. *Mon. Not. R. Astron. Soc.* **490**, 3952–3965 (2019).
44. Vaisanen, P., Tollestrup, E. V. & Fazio, G. G. Confusion limit resulting from galaxies: using the infrared array camera on board SIRTF. *Mon. Not. R. Astron. Soc.* **325**, 1241–1252 (2001).
45. Kamath, S. *Challenges for dark energy science: color gradients and blended objects*. PhD thesis, Stanford Univ (2020).
46. Hausen, R. & Robertson, B. E. Morpheus: a deep learning framework for the pixel-level analysis of astronomical image data. *Astrophys. J. Suppl. Ser.* **248**, 20 (2020).
47. Jones, D. M. & Heavens, A. F. Bayesian photometric redshifts of blended sources. *Mon. Not. R. Astron. Soc.* **483**, 2487–2505 (2019).
48. Joseph, R., Courbin, F. & L. Starck, J. Multi-band morpho-spectral component analysis deblending tool (MuSCADeT): deblending colourful objects. *Astron. Astrophys. Suppl. Ser.* **589**, A2 (2016).
49. Bryant, J. J. et al. The SAMI galaxy survey: instrument specification and target selection. *Mon. Not. R. Astron. Soc.* **447**, 2857–2879 (2015).
50. Bundy, K. et al. Overview of the SDSS-IV MaNGA survey: mapping nearby galaxies at Apache Point Observatory. *Astrophys. J.* **798**, 7 (2015).
51. Johnston, E. J. et al. SDSS-IV MaNGA: bulge–disc decomposition of IFU data cubes (BUDDI). *Mon. Not. R. Astron. Soc.* **465**, 2317–2341 (2017).
52. Hopkins, P. F. et al. FIRE-2 simulations: physics versus numerics in galaxy formation. *Mon. Not. R. Astron. Soc.* **480**, 800–863 (2018).
53. Kado-Fong, E., Kim, J.-G., Ostriker, E. C. & Kim, C.-G. Diffuse ionized gas in simulations of multiphase, star-forming galactic disks. *Astrophys. J.* **897**, 143 (2020).
54. Kim, W.-T., Kim, C.-G. & Ostriker, E. C. Local simulations of spiral galaxies with the TIGRESS framework. I. Star formation and arm spurs/feathers. *Astrophys. J.* **898**, 35 (2020).
55. Korytov, D. et al. CosmoDC2: A synthetic sky catalog for dark energy science with LSST. *Astrophys. J. Suppl. Ser.* **245**, 26 (2019).
56. The LSST Dark Energy Science Collaboration et al. The LSST DESC DC2 simulated sky survey. *Astrophys. J. Suppl. Ser.* **253**, 31 (2020).
57. Potter, D., Stadel, J. & Teyssier, R. PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys. *Comput. Astrophys. Cosmol.* **4**, 2 (2017).
58. Troxel, M. A. et al. A synthetic Roman Space Telescope High-Latitude Imaging Survey: simulation suite and the impact of wavefront errors on weak gravitational lensing. *Mon. Not. R. Astron. Soc.* **501**, 2044–2070 (2021).
59. Torrey, P. et al. Synthetic galaxy images and spectra from the Illustris simulation. *Mon. Not. R. Astron. Soc.* **447**, 2753–2771 (2015).
60. Suchyta, E. et al. No galaxy left behind: accurate measurements with the faintest objects in the dark energy survey. *Mon. Not. R. Astron. Soc.* **457**, 786–808 (2016).
61. Everett, S. et al. Dark energy survey year 3 results: measuring the survey transfer function with Balrog. Preprint at *arXiv* http://arxiv.org/abs/2012.12825 (2020).
62. Shipley, H. et al. HFF-DeepSpace Photometric Catalogs of the 12 Hubble Frontier Fields, Clusters, and Parallels: Photometry, Photometric Redshifts, and Stellar Masses. *Astrophys. J. Suppl. Ser.* **235**, 14 (2018).
63. Eckert, K. et al. Noise from undetected sources in dark energy survey images. *Mon. Not. R. Astron. Soc.* **497**, 2529–2539 (2020).
64. Gruen, D. et al. Dark energy survey year 1 results: the effect of intracluster light on photometric redshifts for weak gravitational lensing. *Mon. Not. R. Astron. Soc.* **488**, 4389–4399 (2019).
65. Kannawadi, A. et al. Towards emulating cosmic shear data: revisiting the calibration of the shear measurements for the Kilo-Degree Survey. *Astron. Astrophys.* **624**, A92 (2019).
66. Sheldon, E. S., Becker, M. R., MacCrann, N. & Jarvis, M. Mitigating shear-dependent object detection biases with metacalibration. *Astrophys. J.* **902**, 138 (2020).
67. Mandelbaum, R. et al. Weak lensing shear calibration with simulations of the HSC survey. *Mon. Not. R. Astron. Soc.* **481**, 3170–3195 (2018).
68. Kacprzak, T. et al. Measurement and calibration of noise bias in weak lensing galaxy shape estimation. *Mon. Not. R. Astron. Soc.* **427**, 2711–2722 (2012).
69. Huff, E. & Mandelbaum, R. Metacalibration: direct self-calibration of biases in shear measurement. Preprint at *arXiv* http://arxiv.org/abs/1702.02600 (2017).
70. Sheldon, E. S. & Huff, E. M. Practical weak-lensing shear measurement with metacalibration. *Astrophys. J.* **841**, 24 (2017).
71. Hoekstra, H., Kannawadi, A. & Kitching, T. D. Accounting for object detection bias in weak gravitational lensing studies. *Astron. Astrophys.* **646**, A124 (2021).
72. MacCrann, N. et al. DES Y3 results: blending shear and redshift biases in image simulations (2020). Preprint at *arXiv* http://arxiv.org/abs/2012.08567 (2020).
73. Connor, T. et al. Crowded field galaxy photometry: precision colors in the CLASH clusters. *Astrophys. J.* **848**, 37 (2017).
74. Greco, J. P. et al. Illuminating low surface brightness galaxies with the Hyper Suprime-Cam survey. *Astrophys. J.* **857**, 104 (2018).
75. Zhang, Y. et al. Dark energy survey year 1 results: detection of intracluster light at redshift ∼0.25. *Astrophys. J.* **874**, 165 (2019).
76. Palmese, A. et al. Comparing dark energy survey and *HST*–CLASH observations of the galaxy cluster RXC J2248.7–4431: implications for stellar mass versus dark matter. *Mon. Not. R. Astron. Soc.* **463**, 1486–1499 (2016).
77. Koekemoer, A. M. et al. The cosmos survey: Hubble space telescope advanced camera for surveys observations and data processing. *Astrophys. J. Suppl. Ser.* **172**, 196–202 (2007).
78. Scoville, N. et al. COSMOS: Hubble space telescope observations. *Astrophys. J. Suppl. Ser.* **172**, 38–45 (2007).
79. Merlin, E. et al. T-PHOT version 2.0: Improved algorithms for background subtraction, local convolution, kernel registration, and new options. *Astron. Astrophys.* **595**, A97 (2016).
80. Nyland, K. et al. An application of multi-band forced photometry to one square degree of SERVS: accurate photometric redshifts and implications for future science. *Astrophys. J. Suppl. Ser.* **230**, 9 (2017).
81. Rhodes, J. et al. Scientific synergy between LSST and Euclid. *Astrophys. J. Suppl. Ser.* **233**, 21 (2017).
82. Rhodes, J. et al. Cosmological synergies enabled by joint analysis of multi-probe data from WFIRST, Euclid, and LSST. *Bull. Am. Astron. Soc.* **51**, 201 (2019).
83. Chary, R. et al. Joint survey processing of LSST, Euclid and WFIRST: Enabling a broad array of astrophysics and cosmology through pixel level combinations of datasets. Preprint at *arXiv* http://arxiv.org/abs/1910.01259 (2019).
84. Joseph, R., Melchior, P. and Moolekamp, F. Joint survey processing: combined resampling and convolution for galaxy modelling and deblending. Preprint at *arXiv* https://arxiv.org/abs/2107.06984 (2021).
85. Strauss, M. A. et al. Spectroscopic target selection in the Sloan Digital Sky Survey: the main galaxy sample. *Astron. J.* **124**, 1810 (2002).
86. Dawson, K. S. et al. The Baryon oscillation spectroscopic survey of SDSS-III. *Astron. J.* **145**, 10 (2013).
87. Takada, M. et al. Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph. *Publ. Astron. Soc. Jpn.* **66**, R1 (2014).
88. Foley, R. J. et al. LSST observing strategy white paper: LSST observations of WFIRST deep fields. Preprint at *arXiv* http://arxiv.org/abs/1812.00514 (2018).
89. Koekemoer, A. et al. Ultra deep field science with WFIRST. *Bull. Am. Astron. Soc.* **51**, 550 (2019).
90. Hartley, W. G. et al. Dark energy survey year 3 results: deep field optical + near-infrared images and catalogue. Preprint at *arXiv* http://arxiv.org/abs/2012.12824 (2020).
91. York, D. G. et al. The Sloan Digital Sky Survey: Technical summary. *Astron. J.* **120**, 1579 (2000).
92. Dey, A. et al. Overview of the DESI legacy imaging surveys. *Astron. J.* **157**, 168 (2019).
93. Grogin, N. A. et al. CANDELS: the cosmic assembly near-infrared deep extragalactic legacy survey. *Astrophys. J. Suppl. Ser.* **197**, 35 (2011).
94. Koekemoer, A. M. et al. CANDELS: the cosmic assembly near-infrared deep extragalactic legacy survey — the Hubble Space Telescope observations, imaging data products, and mosaics. *Astrophys. J. Suppl. Ser.* **197**, 36 (2011).

Acknowledgements
J.S. acknowledges that this document was prepared using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under contract no. DE-AC02-07CH11359.

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Peer review information
Nature Reviews Physics thanks Dustin Lang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

RELATED LINKS
Euclid Deep Fields: https://www.cosmos.esa.int/web/euclid/euclid-survey
Hubble Legacy Archive: https://hla.stsci.edu/