The Characterisation, Subtraction, and Addition of Astronomical Images

Robert Lupton

Princeton University

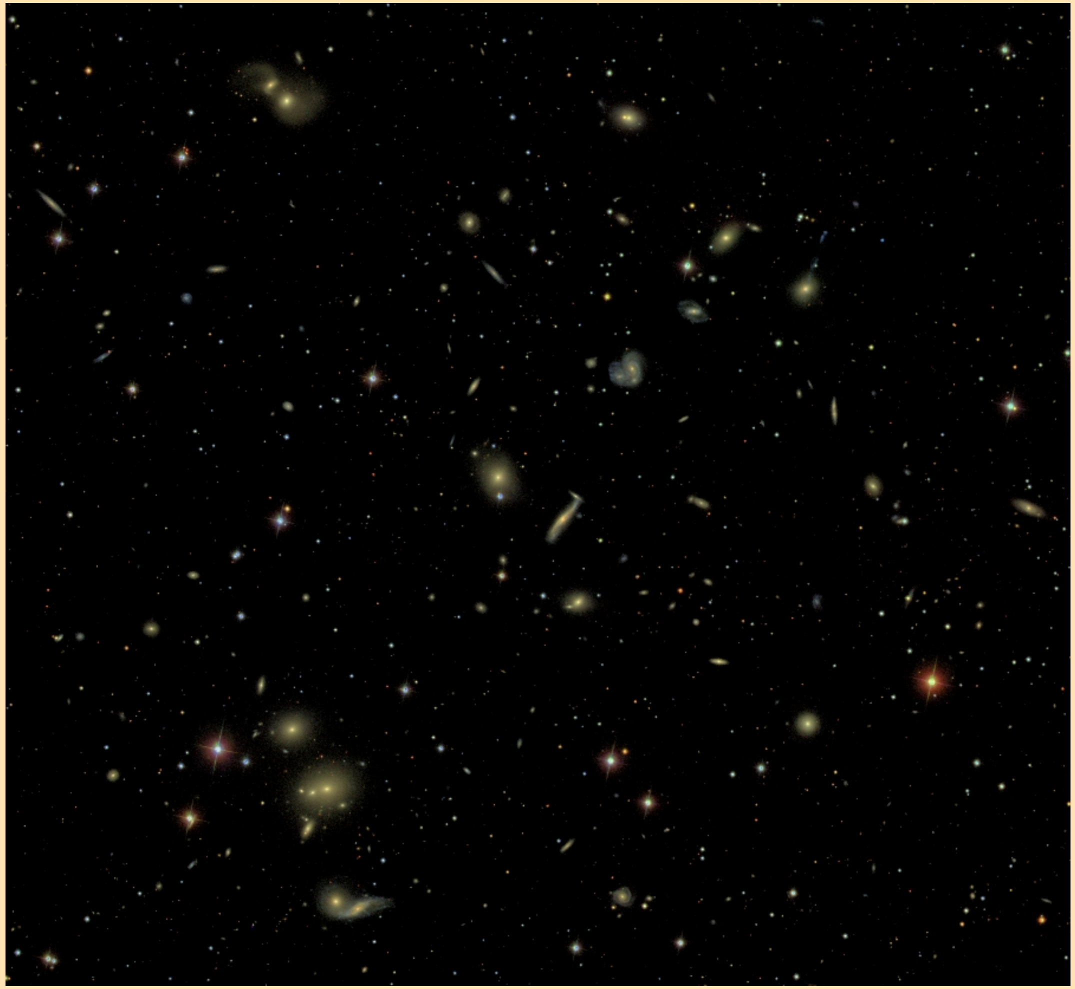
Introduction

Introduction

We can write down all of observational astronomy in terms of the model

$$I = \sum_{i \in \text{objects}} O_i \otimes \phi + \epsilon_i$$

where the O_i are our objects, ϕ is the PSF, and ϵ is the noise. If our objects are simple (e.g. if they're all stars, so $O_i = A_i \delta(x - x_i)$) then all that we have to do is estimate $3N$ numbers.



Astronomical PSFs

The PSF has at least three contributions:

Astronomical PSFs

The PSF has at least three contributions:

The detector

Astronomical PSFs

The PSF has at least three contributions:

The detector

- Charge diffusion within the photo-active material
- Variable sensitivity wihin the finite pixels.

Astronomical PSFs

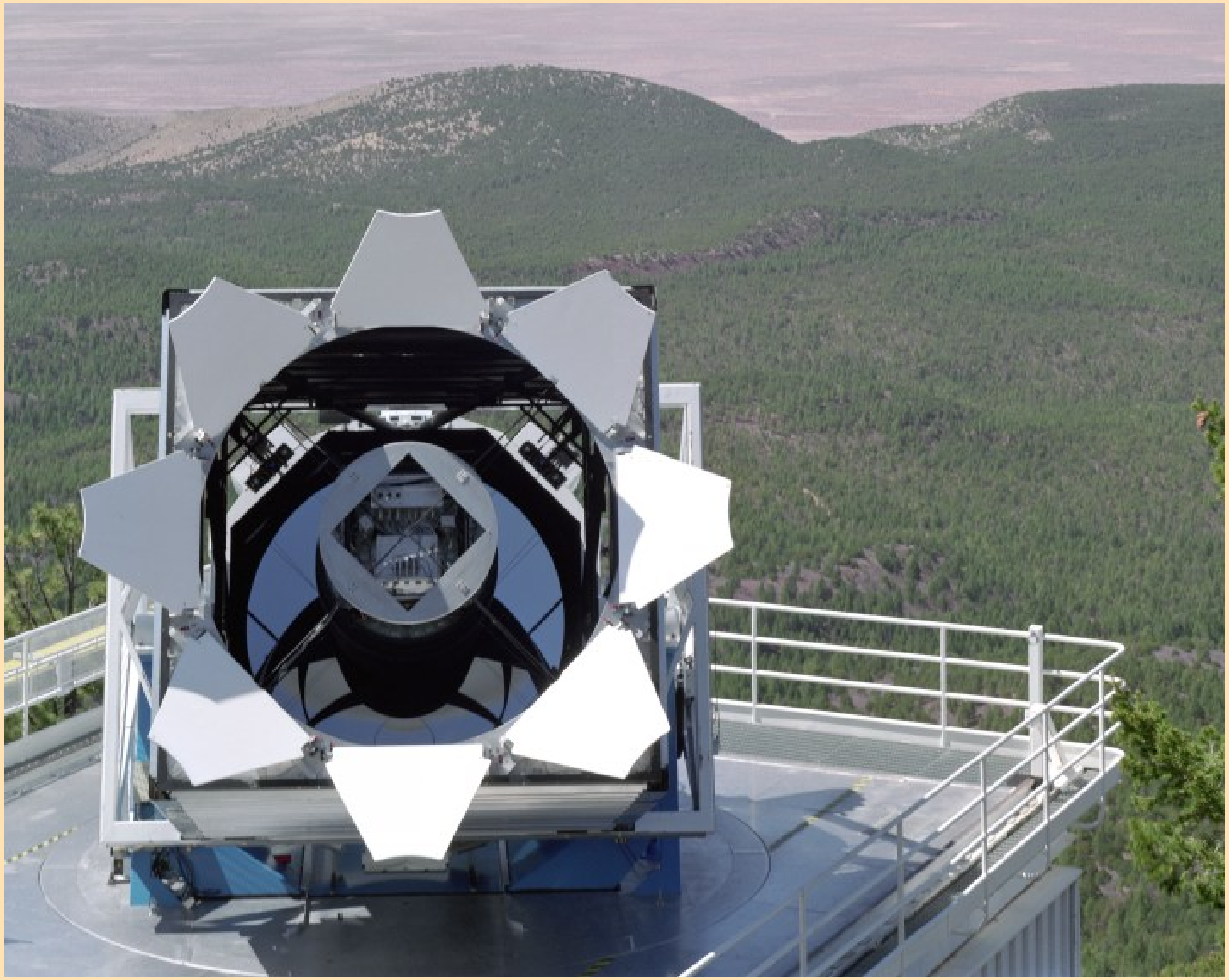
The PSF has at least three contributions:

The detector

- Charge diffusion within the photo-active material
- Variable sensitivity wihin the finite pixels.

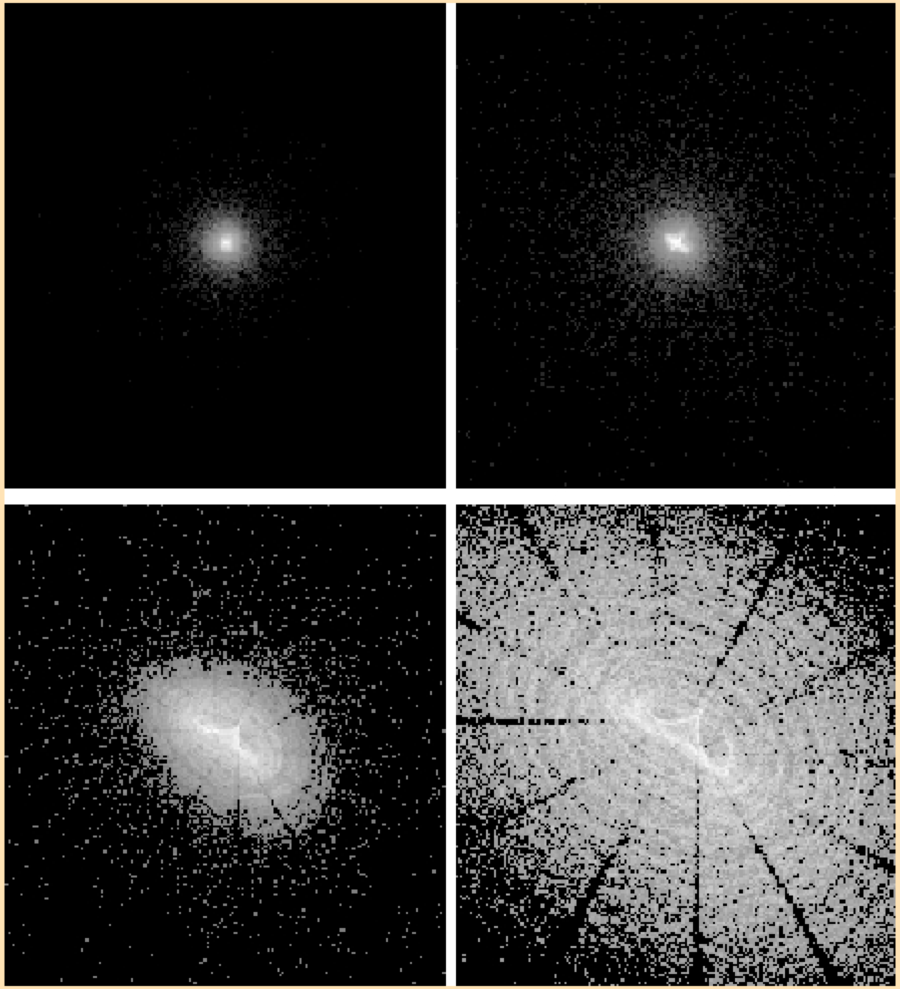
Critically sampling a PSF of rms size α requires a pixel of size $\sim \alpha$, so the pixel convolution increases the rms size of the PSF by c. $\sqrt{1 + 2/12} - 1 \sim 8\%$.

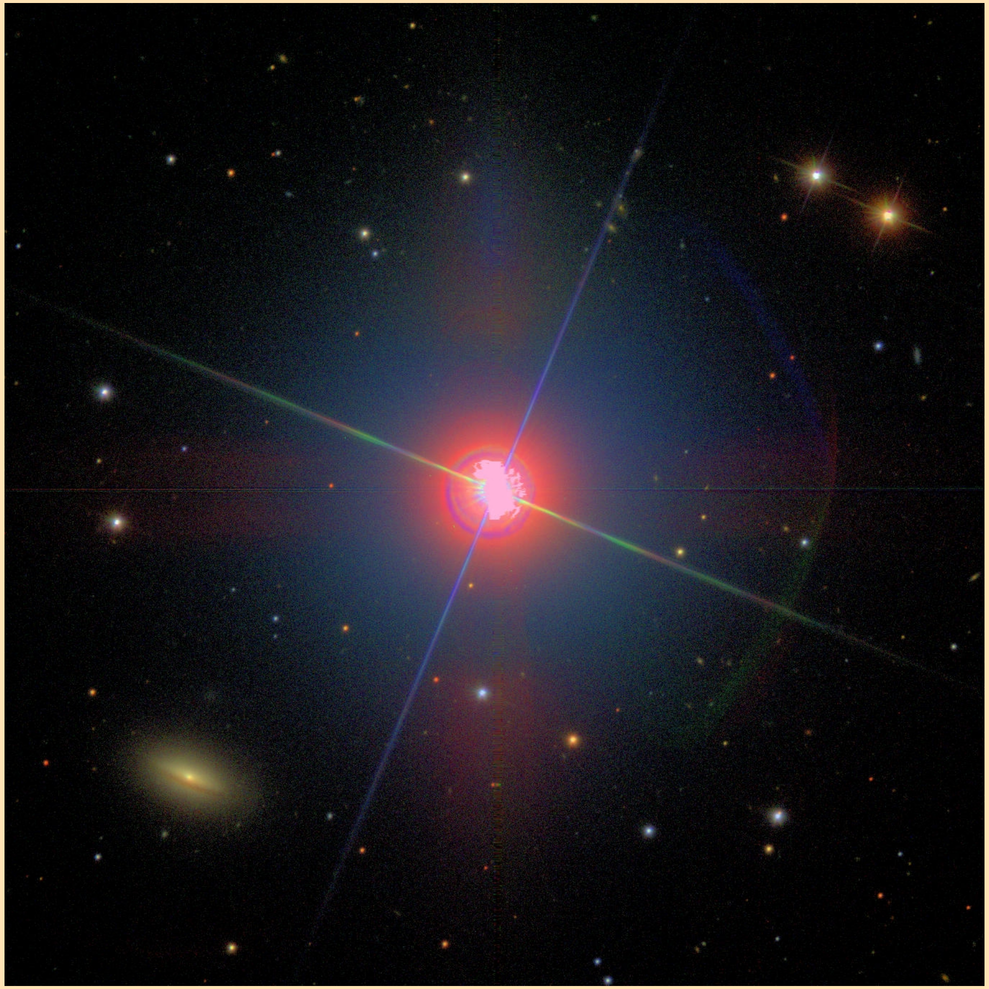
The telescope



The optics are usually designed to play at most a small rôle in the overall error budget.

The optics are usually designed to play at most a small rôle in the overall error budget.





NGC442

The Atmosphere

Turbulence in the atmosphere leads to telescopes providing less-than-perfect images.

The Atmosphere

Turbulence in the atmosphere leads to telescopes providing less-than-perfect images.

The high-atmosphere wavefront has a coherence length $r_0 \sim$ 10–50cm, corresponding to structure of the PSF over scales of arcminutes. Low-atmosphere turbulence has much larger spatial scales.

The Atmosphere

Turbulence in the atmosphere leads to telescopes providing less-than-perfect images.

The high-atmosphere wavefront has a coherence length $r_0 \sim$ 10–50cm, corresponding to structure of the PSF over scales of arcminutes. Low-atmosphere turbulence has much larger spatial scales.

The Kolmogorov-dominated PSF's Fourier transform is given by

$$\exp\left(-6.8839\,(2\pi k\lambda/r_0))^{5/3}\,/2\right)$$

where $\lambda/r_0 \equiv$ FWHM$/0.976$

The Atmosphere

Turbulence in the atmosphere leads to telescopes providing less-than-perfect images.

The high-atmosphere wavefront has a coherence length $r_0 \sim$ 10–50cm, corresponding to structure of the PSF over scales of arcminutes. Low-atmosphere turbulence has much larger spatial scales.

A sum of two Gaussians, $G(\alpha) + 0.1G(2\alpha)$, is a convenient and reasonably accurate representation of (the core of) the PSF.

Chromatic Effects

Chromatic Effects

- The Detector

Chromatic Effects

- The Detector
- The Telescope

Chromatic Effects

- The Detector
- The Telescope
- The Atmosphere

Sampling and Dithering

A telescope of diameter D band-limits the image of the sky at D/λ, and the atmosphere severly attenuates beyond r_0/λ; in general $r_0 \ll D$ and we design the instrumentation to sample a signal limited at the lower frequency r_0/λ but the signal has some power at higher (angular) frequencies.

Sampling and Dithering

A telescope of diameter D band-limits the image of the sky at D/λ, and the atmosphere severly attenuates beyond r_0/λ; in general $r_0 \ll D$ and we design the instrumentation to sample a signal limited at the lower frequency r_0/λ but the signal has some power at higher (angular) frequencies.

Unfortunately, detectors are expensive, and some data is taken which is not fully sampled.

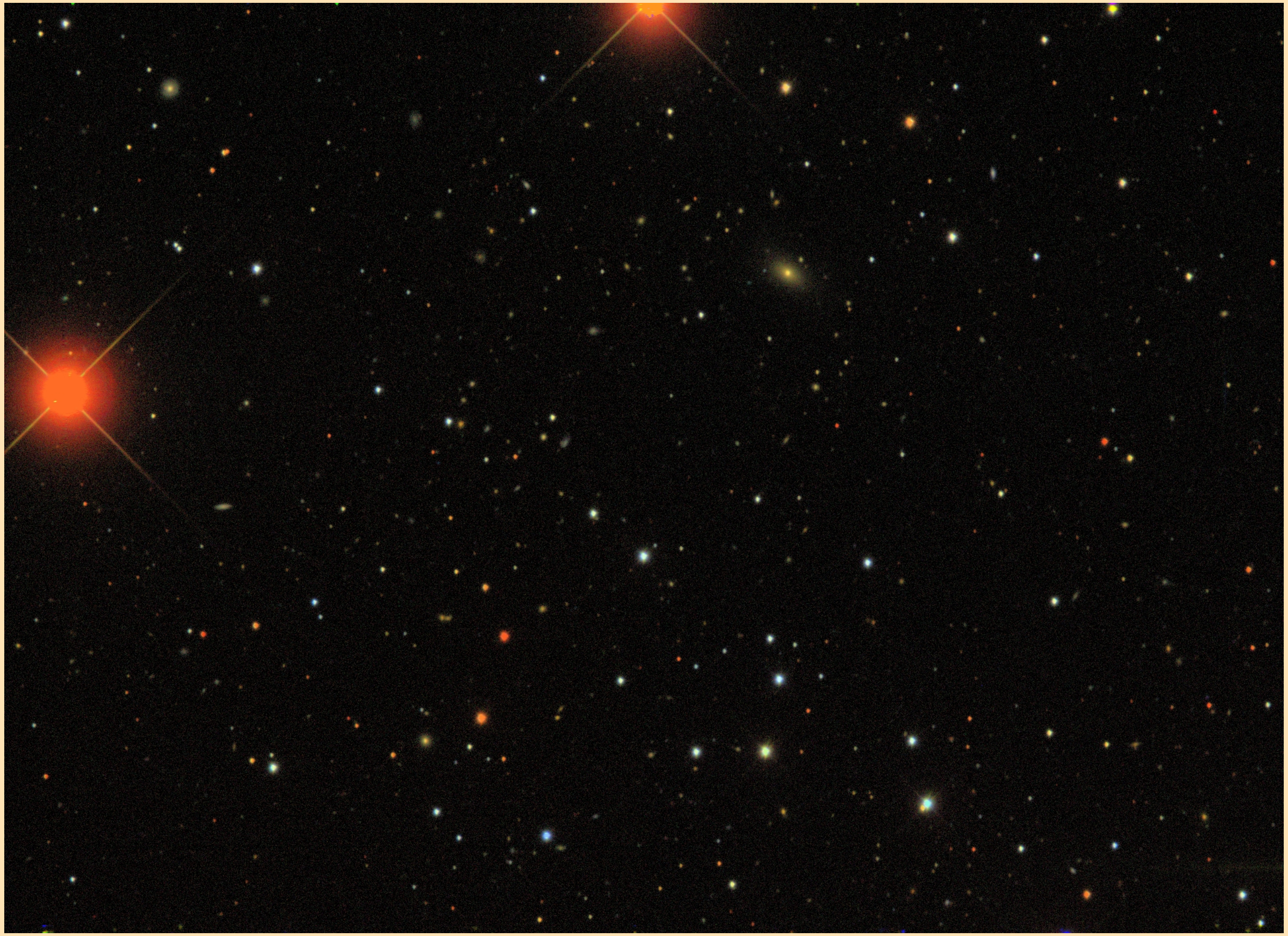
Sampling and Dithering

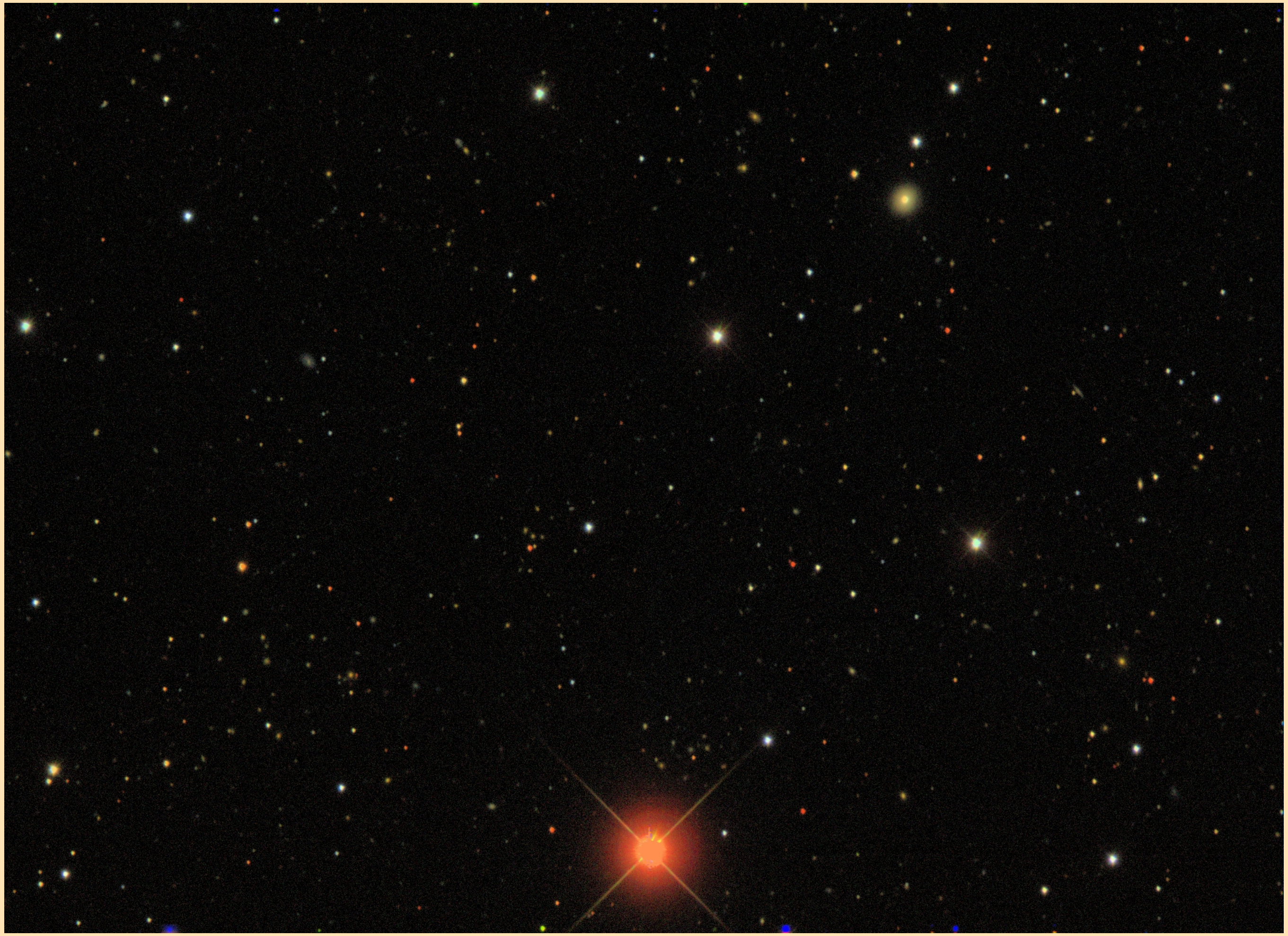
A telescope of diameter D band-limits the image of the sky at D/λ, and the atmosphere severly attenuates beyond r_0/λ; in general $r_0 \ll D$ and we design the instrumentation to sample a signal limited at the lower frequency r_0/λ but the signal has some power at higher (angular) frequencies.

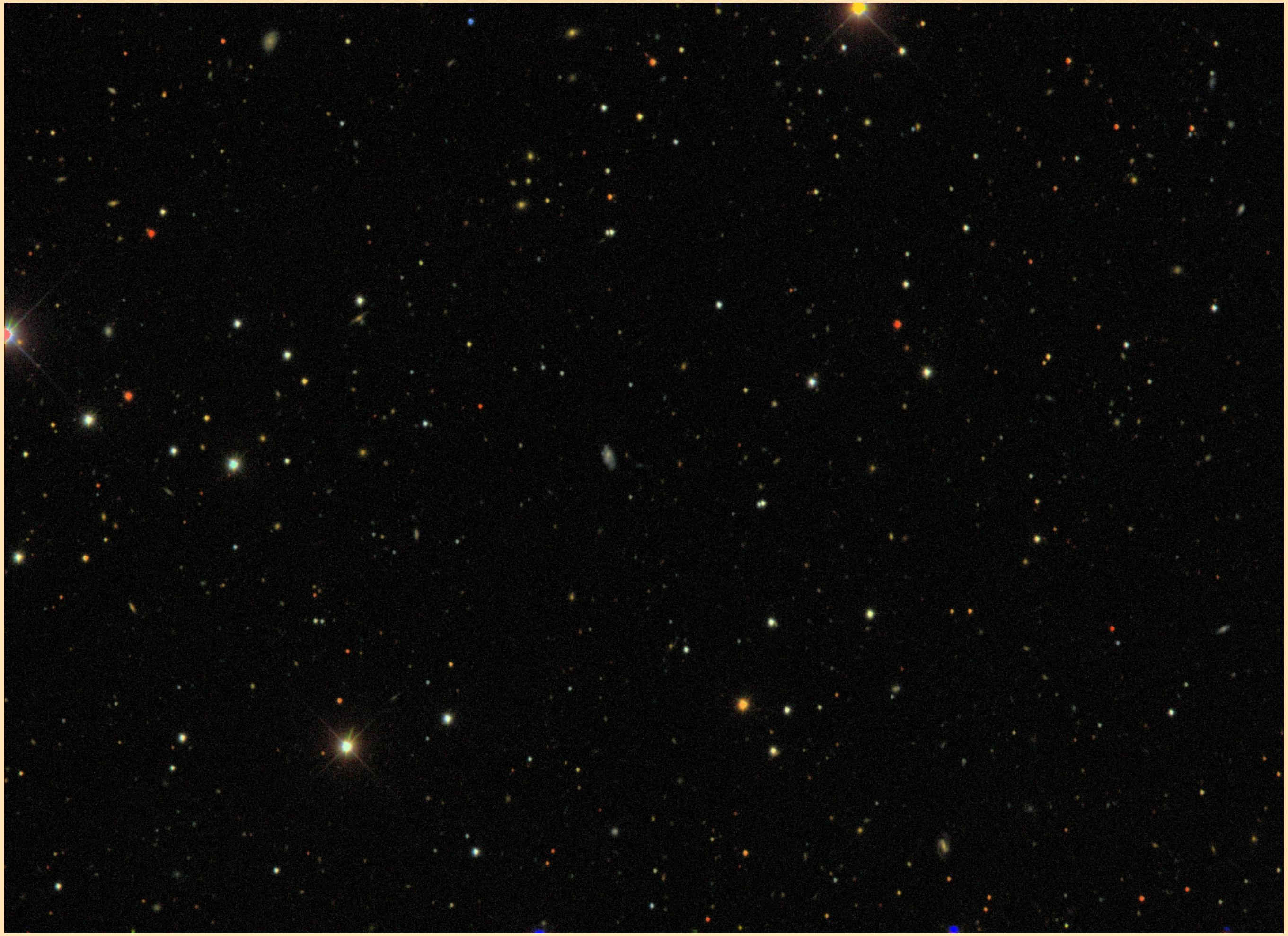
Unfortunately, detectors are expensive, and some data is taken which is not fully sampled.

In many cases, multiple exposures with the telescope's pointing varied a little are taken to recover the fully sampled image.

Estimating the PSF







What is it that I wish to estimate?

I'd be happy with a pixel-based represention — basically an image — but the outer parts of the PSF extend to several tens of arcseconds, and this would involve estimating the value of at least 20000 pixels.

What is it that I wish to estimate?

I'd be happy with a pixel-based represention — basically an image — but the outer parts of the PSF extend to several tens of arcseconds, and this would involve estimating the value of at least 20000 pixels.

Three approaches suggest themselves:

What is it that I wish to estimate?

I'd be happy with a pixel-based represention — basically an image — but the outer parts of the PSF extend to several tens of arcseconds, and this would involve estimating the value of at least 20000 pixels.

Three approaches suggest themselves:

- Employing a purely analytic approximation

What is it that I wish to estimate?

I'd be happy with a pixel-based represention — basically an image — but the outer parts of the PSF extend to several tens of arcseconds, and this would involve estimating the value of at least 20000 pixels.

Three approaches suggest themselves:

- Employing a purely analytic approximation
- Using an exact representation of the inner parts of the PSF and an analytic approximation to the outer parts

What is it that I wish to estimate?

I'd be happy with a pixel-based represention — basically an image — but the outer parts of the PSF extend to several tens of arcseconds, and this would involve estimating the value of at least 20000 pixels.

Three approaches suggest themselves:

- Employing a purely analytic approximation
- Using an exact representation of the inner parts of the PSF and an analytic approximation to the outer parts
- a hybrid scheme whereby we use an analytic form, and also carry a residual image

What is it that I wish to estimate?

I'd be happy with a pixel-based represention — basically an image — but the outer parts of the PSF extend to several tens of arcseconds, and this would involve estimating the value of at least 20000 pixels.

Three approaches suggest themselves:

- Employing a purely analytic approximation
- Using an exact representation of the inner parts of the PSF and an analytic approximation to the outer parts
- a hybrid scheme whereby we use an analytic form, and also carry a residual image

Because the PSF has spatial structure, we have to also ask how to represent this in any of these approaches.

From my discussion of the various contributions to the PSF you might think that it would make sense to model the PSF as a convolution of the constant-in-time component (the detector and telescope), and the constant-in-space component (the atmosphere).

From my discussion of the various contributions to the PSF you might think that it would make sense to model the PSF as a convolution of the constant-in-time component (the detector and telescope), and the constant-in-space component (the atmosphere).

To the best of my knowledge this has not been attempted, although Jarvis and Jain have modelled the PSF in an ensemble of images by studying its variation at a given point in the focal plane as a function of exposure number.

SDSS; KL Expansion of the PSF

We can use a set of objects identified as stars to form a KL basis, retaining the first n terms of the expansion:

$$P_{(i)}(u,v) = \sum_{r=1}^{r=n} a^r_{(i)} B_r(u,v)$$

where $P_{(i)}$ is the i^{th} PSF star, the B_r are the KL basis functions, and u, v are pixel coordinates relative to the origin of the basis functions.

SDSS; KL Expansion of the PSF

We can use a set of objects identified as stars to form a KL basis, retaining the first n terms of the expansion:

$$P_{(i)}(u, v) = \sum_{r=1}^{r=n} a^r_{(i)} B_r(u, v)$$

where $P_{(i)}$ is the i^{th} PSF star, the B_r are the KL basis functions, and u, v are pixel coordinates relative to the origin of the basis functions.

In determining the B_r, the $P_{(i)}$ are normalised to have equal peak value.

Once we know the B_r we can write

$$a^r_{(i)} \approx \sum_{l=m=0}^{l+m \leq N} b^r_{lm} x^l_{(i)} y^m_{(i)}$$

where x, y are the coordinates of the centre of the ith star.

Once we know the B_r we can write

$$a^r_{(i)} \approx \sum_{l=m=0}^{l+m \leq N} b^r_{lm} x^l_{(i)} y^m_{(i)}$$

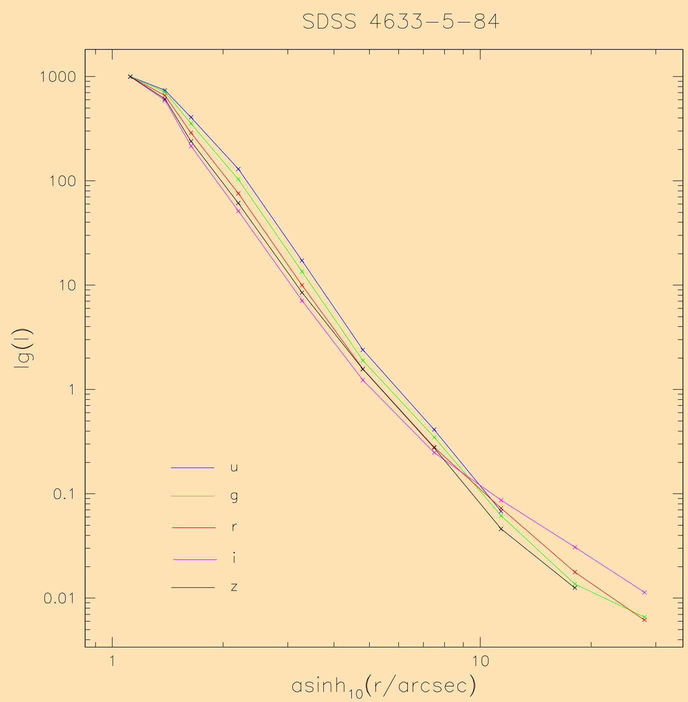
where x, y are the coordinates of the centre of the i^{th} star.

The b^r_{lm} are determined by minimising

$$\sum_i \left(\frac{P_{(i)}(u,v) - \sum_{r=1}^{r=n} a^r_{(i)} B_r(u,v)}{\sigma^2 + \Upsilon^2} \right)^2$$

The Outer Parts of the PSF

The KL model of the PSF extends to only about 7", but the outer parts contain a significant part of the total flux, and need to be modelled. We use partial radial profiles from stars of a range of magnitudes to construct a single profile in each filter.

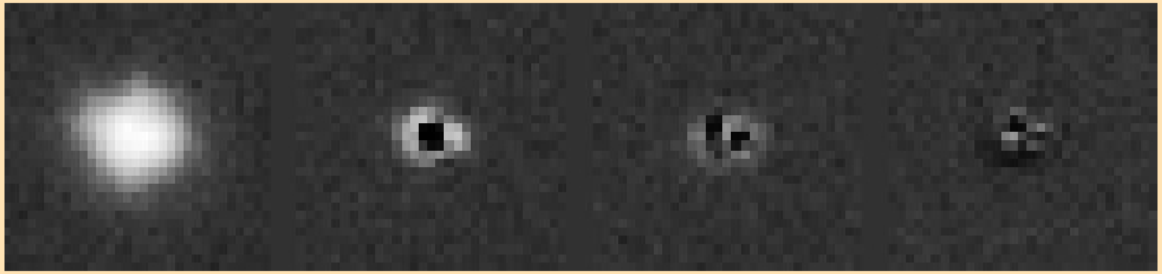

SDSS 4633-5-84
1000
100
10
1
0.1
0.01
lg(I)
1
10
asinh$_{10}$(r/arcsec)
u
g
r
i
z

Application to SDSS data

For each CCD, in each band, there are typically 15-25 stars in a frame that we can use to determine the PSF. We usually take $n = 3$ and $N = 2$ (i.e. the PSF spatial variation is quadratic).

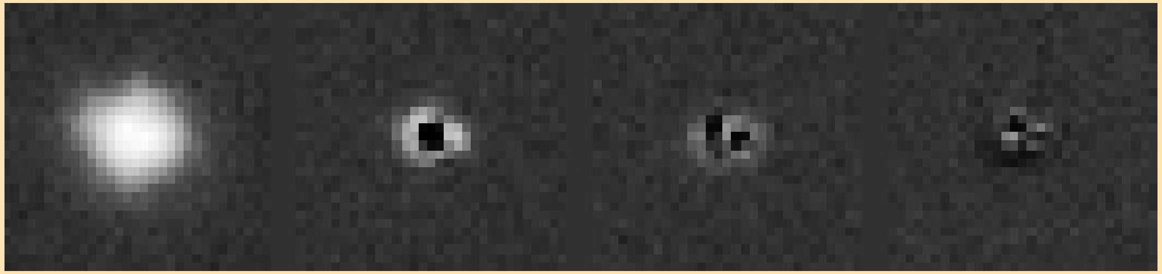
Application to SDSS data

For each CCD, in each band, there are typically 15-25 stars in a frame that we can use to determine the PSF. We usually take $n = 3$ and $N = 2$ (i.e. the PSF spatial variation is quadratic).

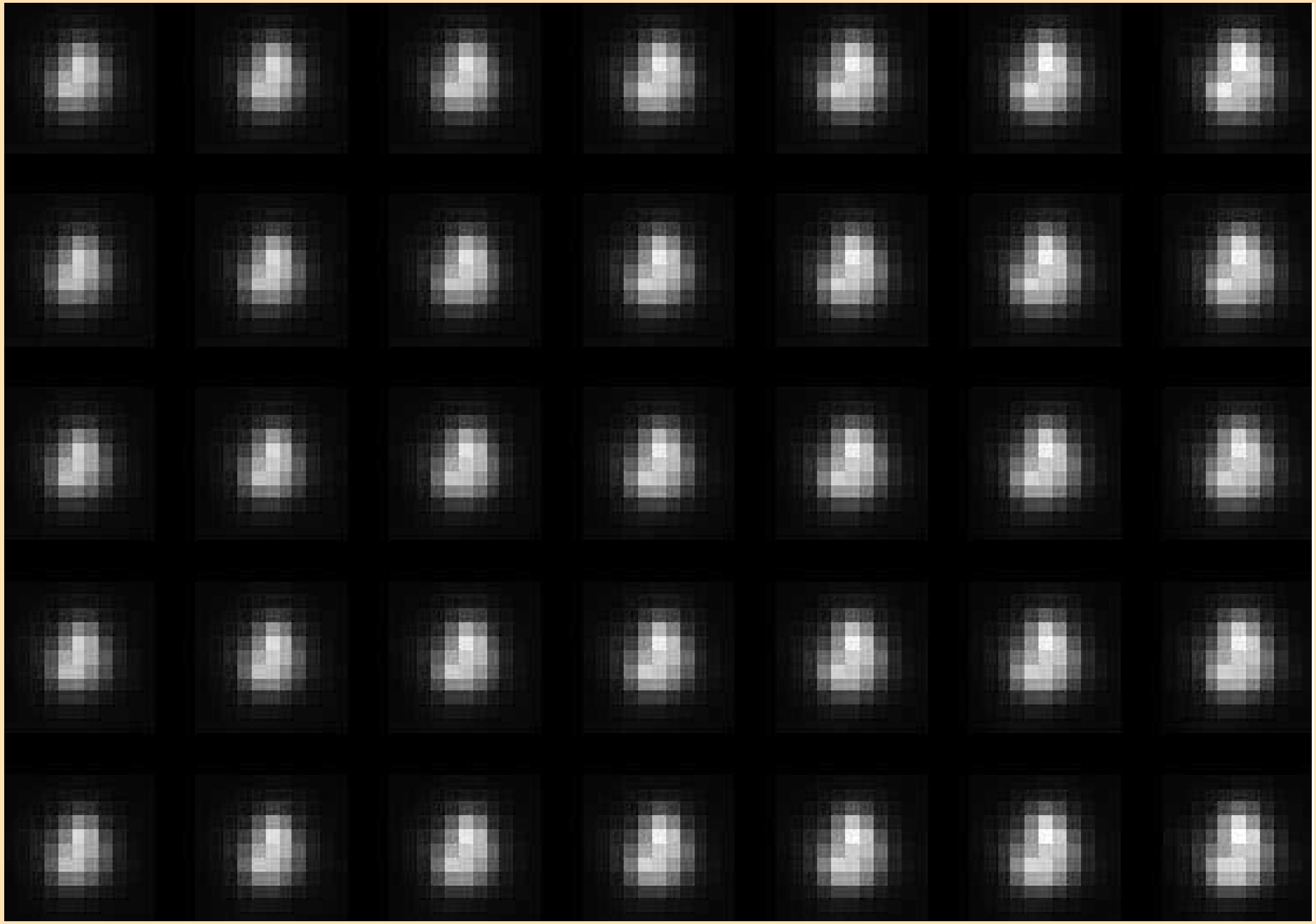


Application to SDSS data

For each CCD, in each band, there are typically 15-25 stars in a frame that we can use to determine the PSF. We usually take $n = 3$ and $N = 2$ (i.e. the PSF spatial variation is quadratic).



We force the KL basis images with $n > 0$ to have 0 mean in their outermost pixels, those more than approximately 7'' from the center of the star.



Reconstructed PSFs for 756-z6-700

Gauging the Success of the KL Expansion

We estimate the PSF at about 120 positions around the edge of the field, and also a few points in the interior. At each of these positions we calculate an aperture correction. The PSF model used in estimating the flux is based on a *local* Gaussian fit.

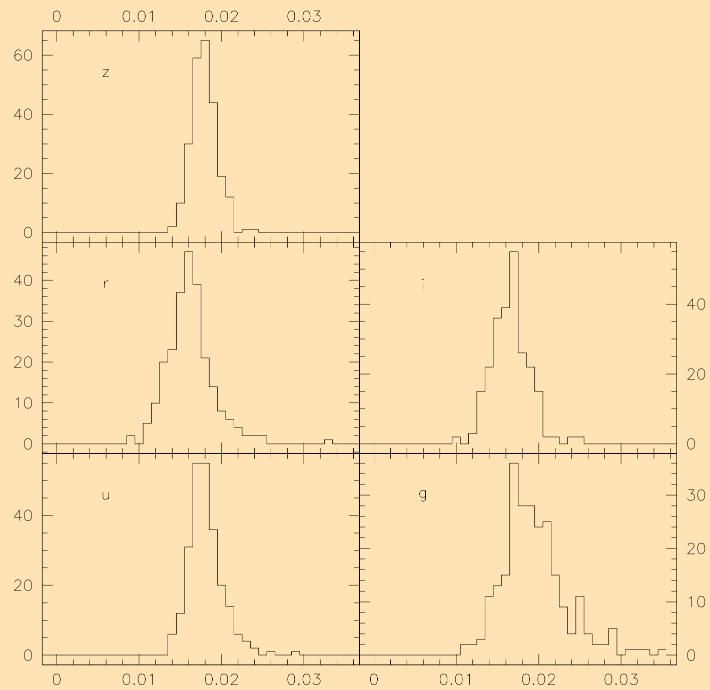
Gauging the Success of the KL Expansion

We estimate the PSF at about 120 positions around the edge of the field, and also a few points in the interior. At each of these positions we calculate an aperture correction. The PSF model used in estimating the flux is based on a *local* Gaussian fit.

We reject the PSF model if the maximum or minimum aperture correction seen differs from unity by more than 30%.

We then compare the PSF photometry based on the modeled KL PSFs to the aperture photometry for the same (bright) stars.

We then compare the PSF photometry based on the modeled KL PSFs to the aperture photometry for the same (bright) stars.



Problems with this approach

Problems with this approach

We need to identify isolated stars

Problems with this approach

We need to identify isolated stars

We assume that we are able to shift images without error.

Problems with this approach

We need to identify isolated stars

We assume that we are able to shift images without error.

This could be avoided by estimating the properties of n fully-sampled "KL" images from the data, rather than by constructing the basis directly.

Problems with this approach

We need to identify isolated stars

We assume that we are able to shift images without error.

Problems with this approach

We need to identify isolated stars

We assume that we are able to shift images without error.

You can fit a number of objects simultaneously using analytic PSF models, and this could be combined with the "KL" based-approach of the previous paragraph.

Problems with this approach

We need to identify isolated stars

We assume that we are able to shift images without error.

You can fit a number of objects simultaneously using analytic PSF models, and this could be combined with the "KL" based-approach of the previous paragraph.

The models used for spatial variability are generally polynomials. In the case of SDSS data this is probably not a very good basis (due to the coupling of time- and spatial-variability due to TDI).

For long exposures on large telescopes polynomials appear to work reasonably well, although it has been shown that, unsurprisingly, rational functions perform rather better.

Using an Ensemble of Exposures

In many applications there are simply not enough stars to map the PSF's spatial variations with sufficient accuracy.

An interesting approach [Jarvis and Jain] is to study a large number of exposures taken with a given telescope — including rich star fields — and ask about the statistical properties of this ensemble.

Using an Ensemble of Exposures

In many applications there are simply not enough stars to map the PSF's spatial variations with sufficient accuracy.

An interesting approach [Jarvis and Jain] is to study a large number of exposures taken with a given telescope — including rich star fields — and ask about the statistical properties of this ensemble.

They use a 4^{th} order polynomial fit to the spatial structure, and use 8–10 components to analyse 651 exposures with 10^5 stars per exposure.

Using an Ensemble of Exposures

In many applications there are simply not enough stars to map the PSF's spatial variations with sufficient accuracy.

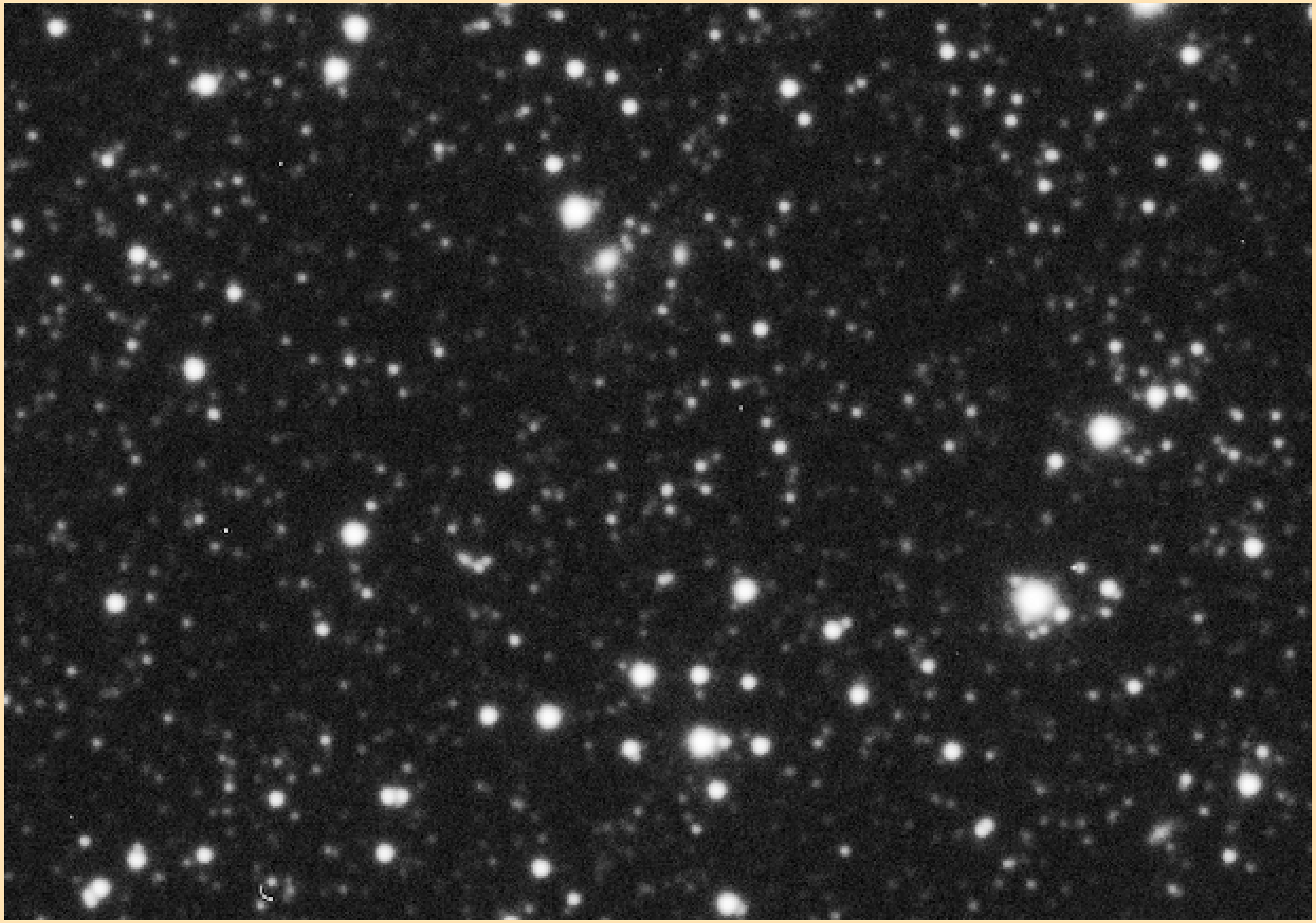
An interesting approach [Jarvis and Jain] is to study a large number of exposures taken with a given telescope — including rich star fields — and ask about the statistical properties of this ensemble.

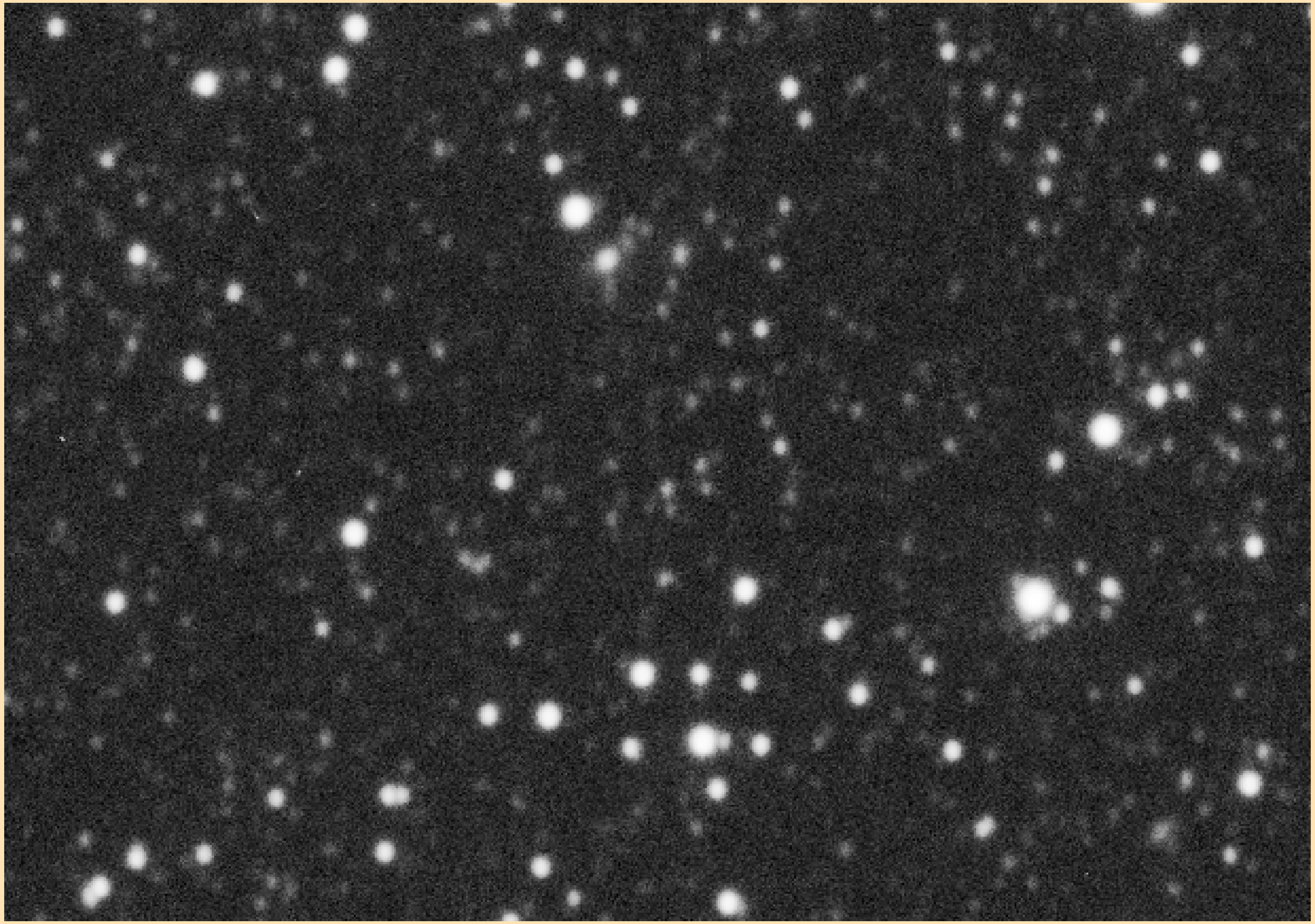
They use a 4^{th} order polynomial fit to the spatial structure, and use 8–10 components to analyse 651 exposures with 10^5 stars per exposure.

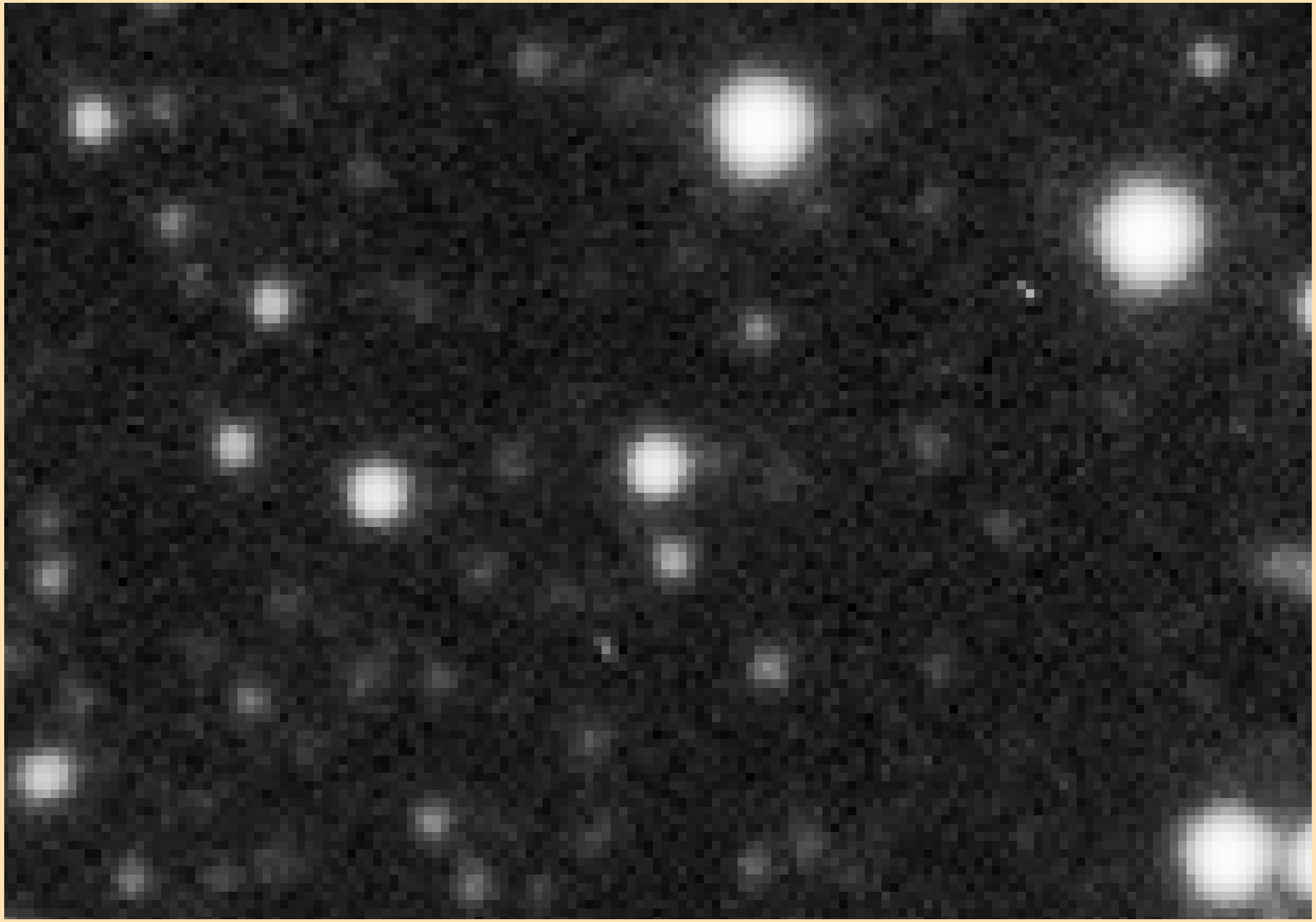
The parameters used to describe the PSF are not specified; for weak lensing studies it may be sufficient to measure a few weighted moments, but one could use a richer representation.

Image Subtraction

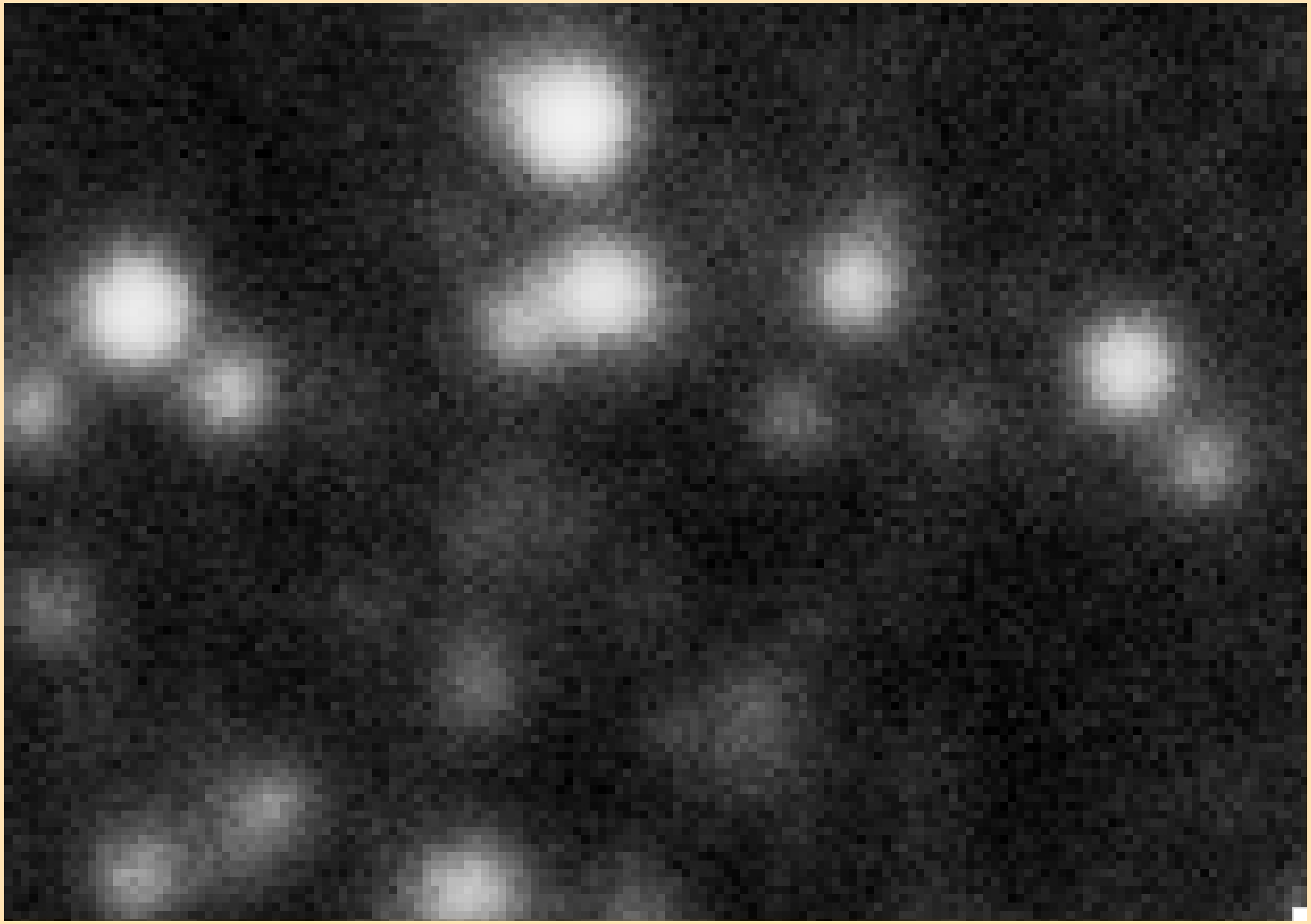
In many situations, only the time variable part of an image is of interest; the classic case is searching for gravitational micro-lensing in the direction of the Galactic bulge. In this case, we have two or more images of the same part of the sky, taken under different conditions; in particular the PSF will be different in the two exposures.

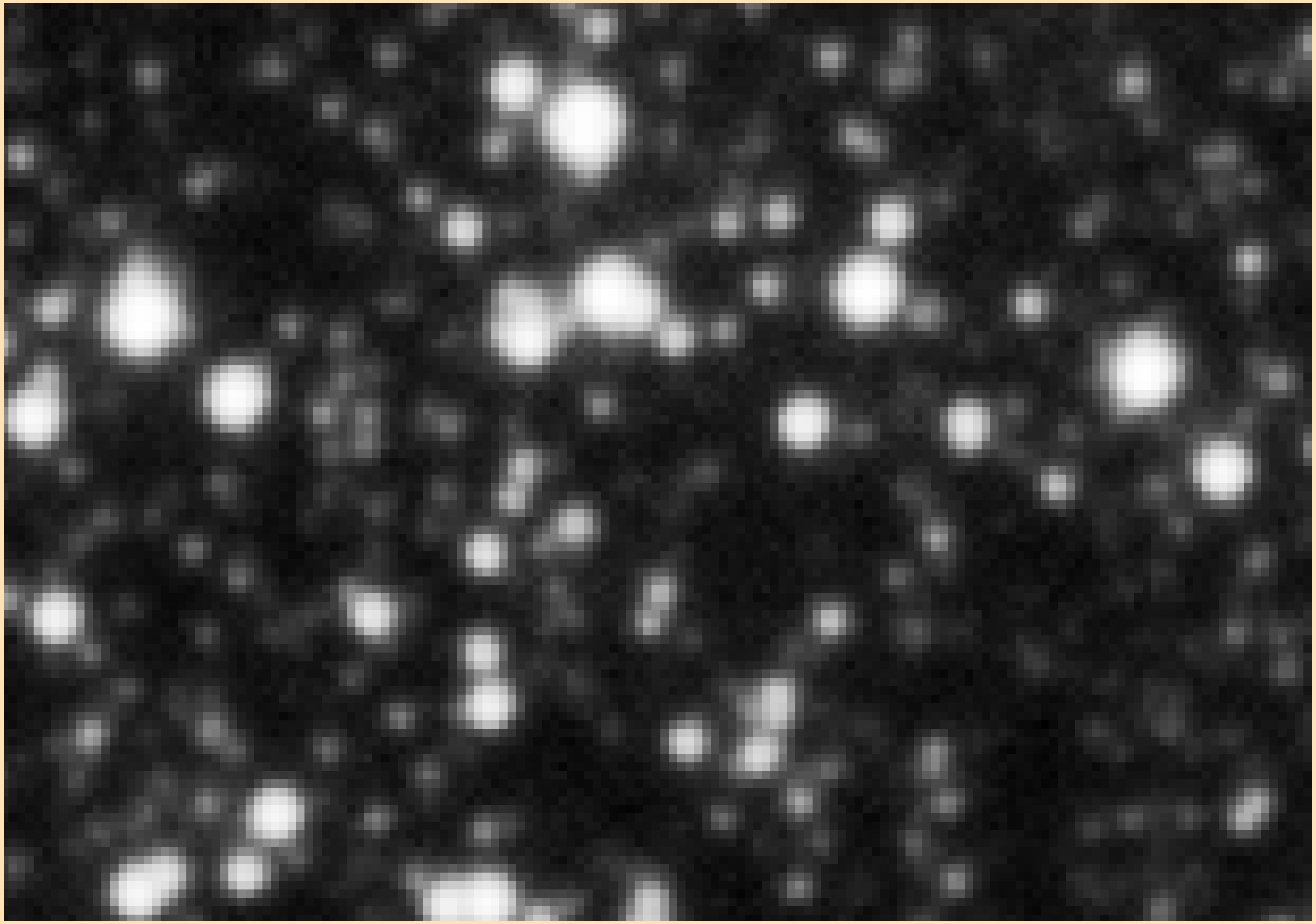






The classic solution to problem of searching for variability is to measure the brightness of each source in both images and compare the resulting catalogues.





An obvious alternative is to subtract the two images directly, allowing for the difference between the seeing in the two images.

An obvious alternative is to subtract the two images directly, allowing for the difference between the seeing in the two images.

Given two images $I_1 \equiv S \otimes \phi_1$ and $I_2 \equiv S \otimes \phi_2$, we can write the Fourier-transform of the (seeing-matched) difference as $I_1(k) - I_2(k) \times (\phi_1(k)/\phi_2(k))$.

An obvious alternative is to subtract the two images directly, allowing for the difference between the seeing in the two images.

Given two images $I_1 \equiv S \otimes \phi_1$ and $I_2 \equiv S \otimes \phi_2$, we can write the Fourier-transform of the (seeing-matched) difference as $I_1(k) - I_2(k) \times (\phi_1(k)/\phi_2(k))$.

Unfortunately, it's difficult to measure the outer parts of the PSFs well enough to carry out this Fourier division.

What really matters is how well the subtraction worked, and that the the residuals left by subtracting objects that *hadn't* varied should be as small as possible; that is, we should find the kernel K such that

$$R \equiv ||I_i - K \otimes I_2||$$

be minimised.

What really matters is how well the subtraction worked, and that the the residuals left by subtracting objects that *hadn't* varied should be as small as possible; that is, we should find the kernel K such that

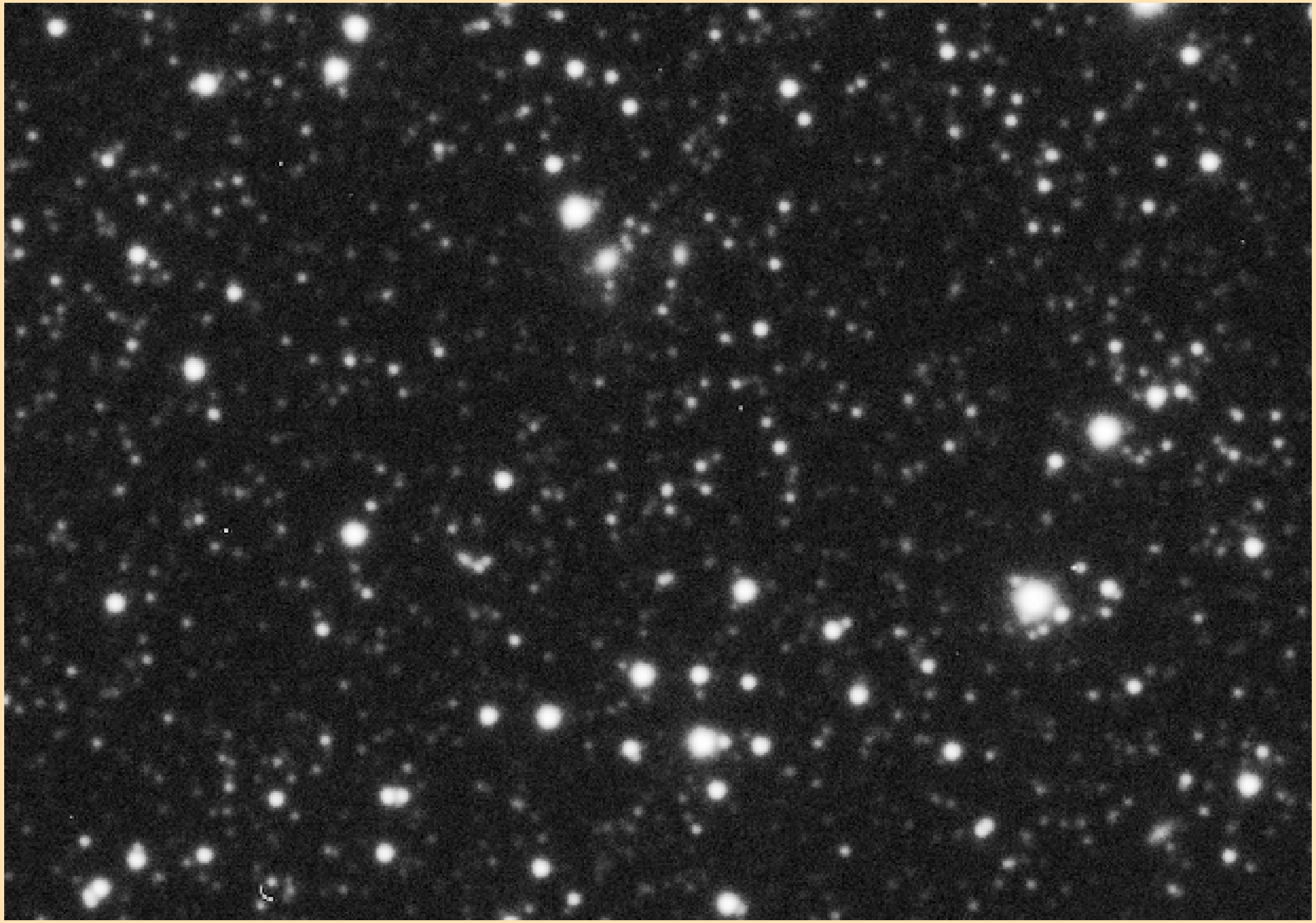
$$R \equiv ||I_i - K \otimes I_2||$$

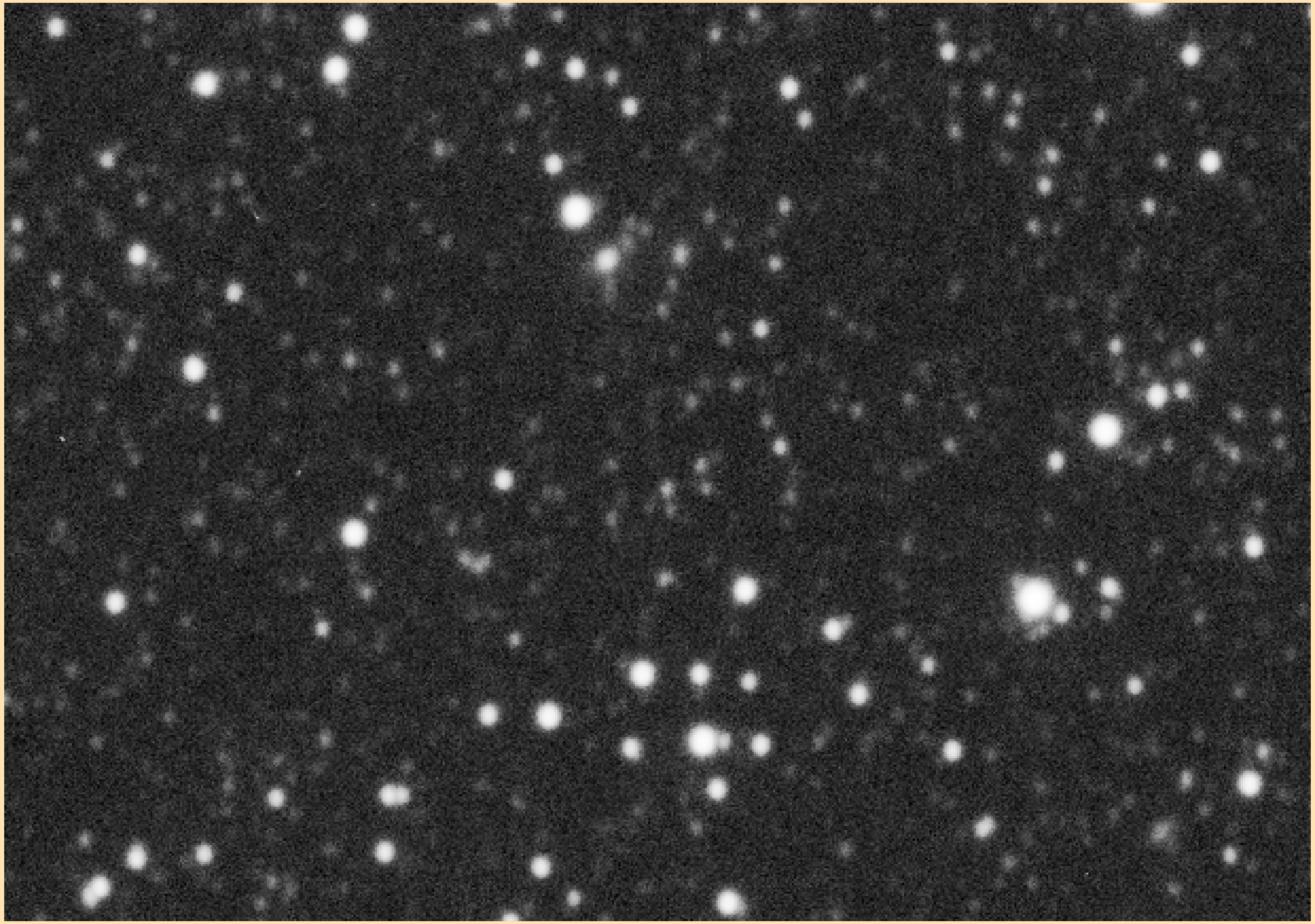
be minimised.

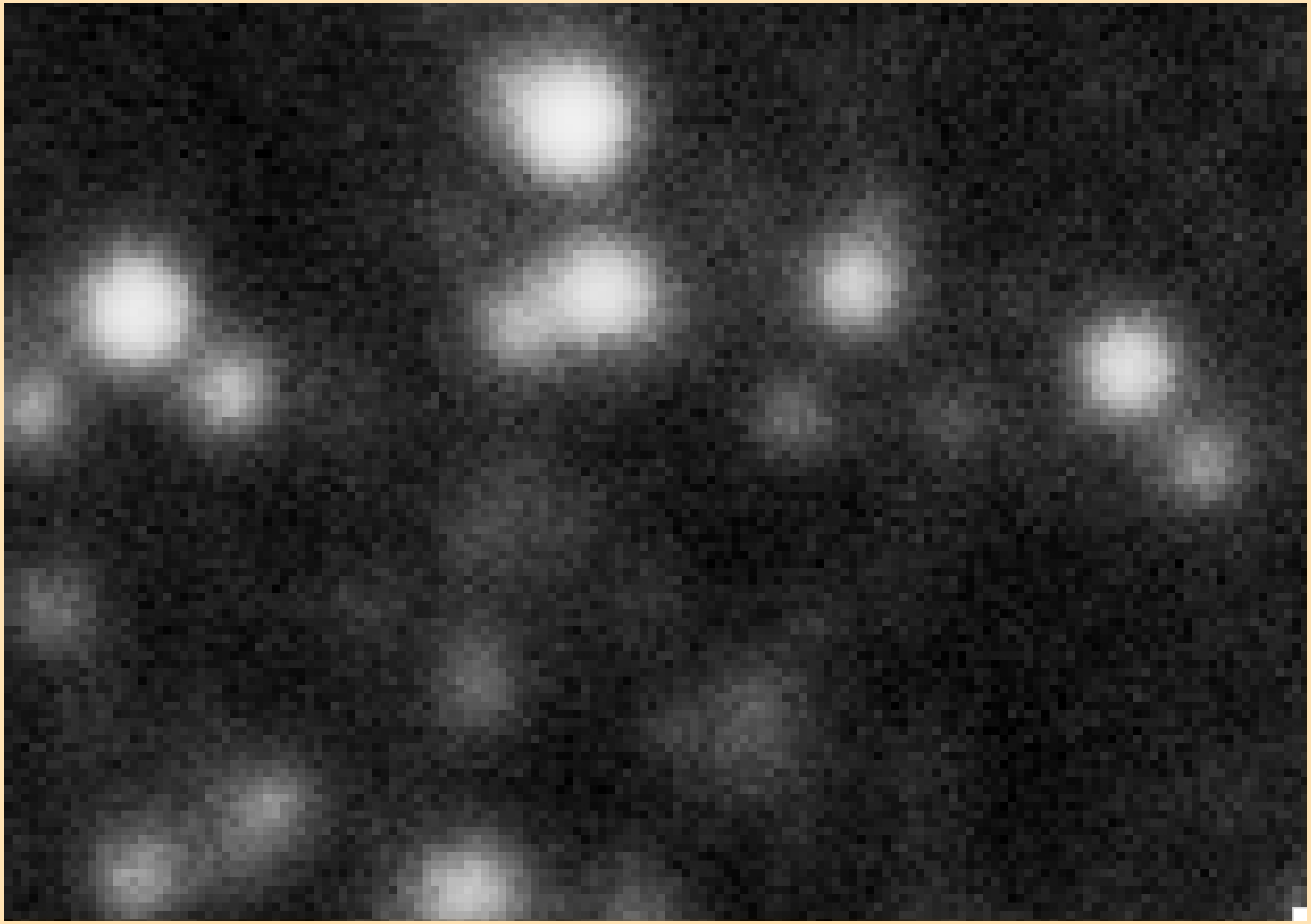
Let us write K as a sum of a set of basis functions:

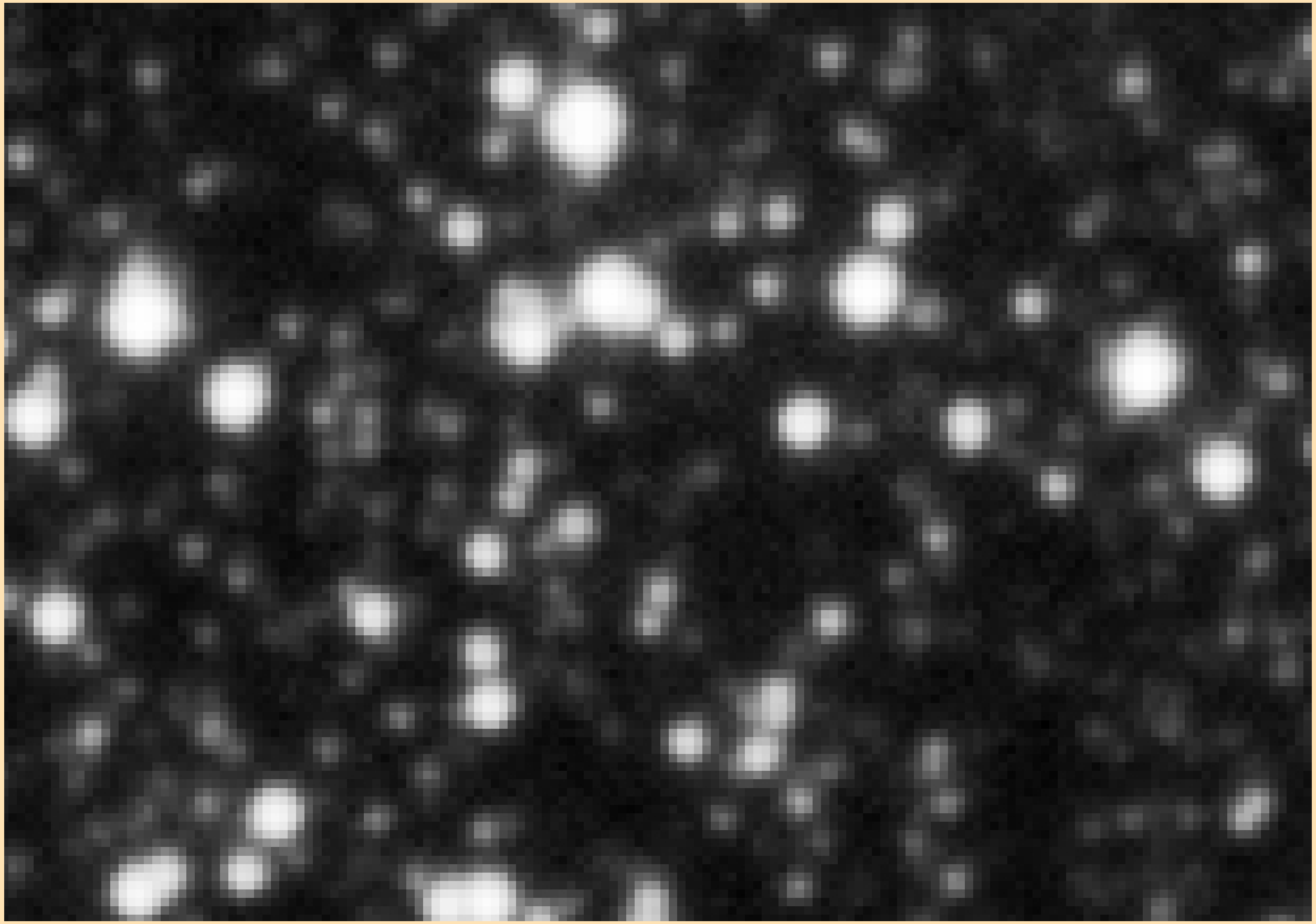
$$K(u, v) \equiv \sum_r a_r B_r(u, v).$$

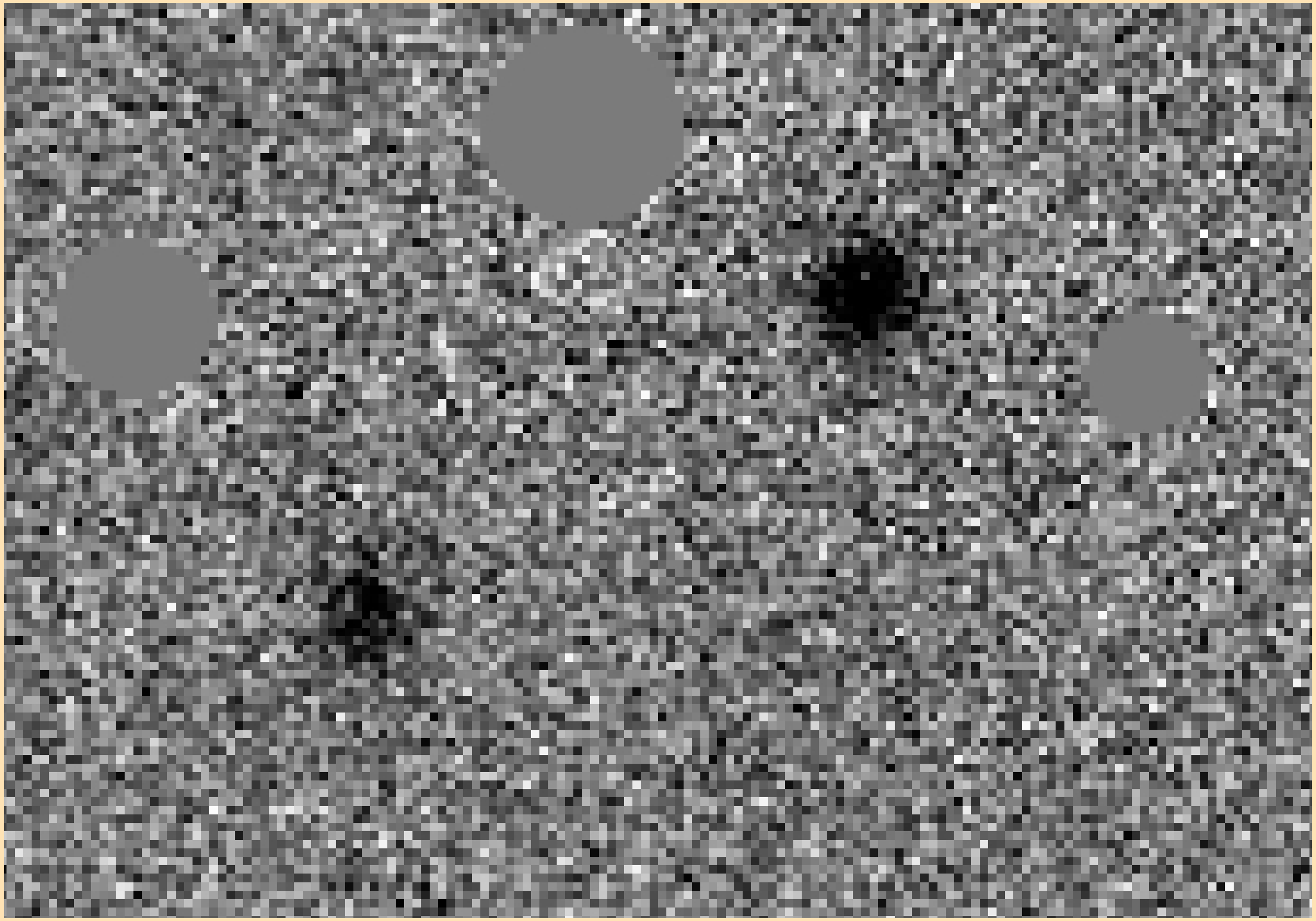
The task of subtracting two images is thus released to the problem of finding a set of a_r that minimise R; if we use an L_2 norm, this is a simple least-squares problem.











Spatially varying kernels can be handled by writing

$$K(u,v;x,y) = \sum_{r=1}^{r=n} \sum_{l=m=0}^{l+m\leq N} b^r_{lm} x^l_{(i)} y^m_{(i)} B_r(u,v)$$

Issues

Issues

- Optimal choice of basis functions

Issues

- Optimal choice of basis functions
- Representation of K's spatial variation

Issues

- Optimal choice of basis functions
- Representation of K's spatial variation
- Generation of the 'reference image' I_2

δ-function basis

An obvious idea is to write:

$$B_r(u, v) = \delta(u - u_r)\delta(v - v_r);$$

i.e. solve directly for K's pixel values and their spatial variation

δ-function basis

An obvious idea is to write:

$$B_r(u, v) = \delta(u - u_r)\delta(v - v_r);$$

i.e. solve directly for K's pixel values and their spatial variation

The resulting kernel is unacceptably noisy. We could add a cost function (e.g. the total variation?), but the problem would become non-linear.

A Sleazy alternative would be:

A Sleazy alternative would be:

- Estimate the kernel K and its spatial structure without any regularisation

A Sleazy alternative would be:

- Estimate the kernel K and its spatial structure without any regularisation
- Reconstruct K at a dense set of points within the field.

A Sleazy alternative would be:

- Estimate the kernel K and its spatial structure without any regularisation
- Reconstruct K at a dense set of points within the field.
- Perform a KL decomposition of these estimates of K, keep the significant terms, and re-expand their spatial structure.

A Sleazy alternative would be:

- Estimate the kernel K and its spatial structure without any regularisation
- Reconstruct K at a dense set of points within the field.
- Perform a KL decomposition of these estimates of K, keep the significant terms, and re-expand their spatial structure.

When the assembled Statistical Wisdom solves the spatially-varying PSF estimation problem, I'll be able to apply an identical algorithm to kernel estimation for image subtraction.

Choice of Reference Image

Choice of Reference Image

One approach would be to choose the image with the best seeing as I_2 (so as to avoid forcing K to do any deconvolution), but this has a number of problems:

Choice of Reference Image

One approach would be to choose the image with the best seeing as I_2 (so as to avoid forcing K to do any deconvolution), but this has a number of problems:

- It increases the noise in the difference image by $\sqrt{2}$

Choice of Reference Image

One approach would be to choose the image with the best seeing as I_2 (so as to avoid forcing K to do any deconvolution), but this has a number of problems:

- It increases the noise in the difference image by $\sqrt{2}$
- it leads to noise correlations

Choice of Reference Image

One approach would be to choose the image with the best seeing as I_2 (so as to avoid forcing K to do any deconvolution), but this has a number of problems:

- It increases the noise in the difference image by $\sqrt{2}$
- it leads to noise correlations
- it requires that we mask out a region the size of the convolution kernel K around the edges of the field and any defects

Choice of Reference Image

One approach would be to choose the image with the best seeing as I_2 (so as to avoid forcing K to do any deconvolution), but this has a number of problems:

- It increases the noise in the difference image by $\sqrt{2}$
- it leads to noise correlations
- it requires that we mask out a region the size of the convolution kernel K around the edges of the field and any defects
- If I_2 is not well sampled we have all the difficulties of resampling an aliased image.

Choice of Reference Image

One approach would be to choose the image with the best seeing as I_2 (so as to avoid forcing K to do any deconvolution), but this has a number of problems:

- It increases the noise in the difference image by $\sqrt{2}$
- it leads to noise correlations
- it requires that we mask out a region the size of the convolution kernel K around the edges of the field and any defects
- If I_2 is not well sampled we have all the difficulties of resampling an aliased image.

The obvious remedy is to generate a fully sampled, defect free image from the ensemble of N images.

Chromatic Effects in the PSF

How should we handle data taken at a variety of airmasses?

Chromatic Effects in the PSF

How should we handle data taken at a variety of airmasses?

A brute force approach would be to generate master templates at different airmasses. It seems probably that using the known colour distribution of the pixels in the image would be advantageous; but the obvious approaches (write the observed image as a linear combination of templates taken at different airmasses, or a template and a colour-template) do not appear to work, as we cannot write our data as a linear combination of the inputs.

Image Addition

Image Addition

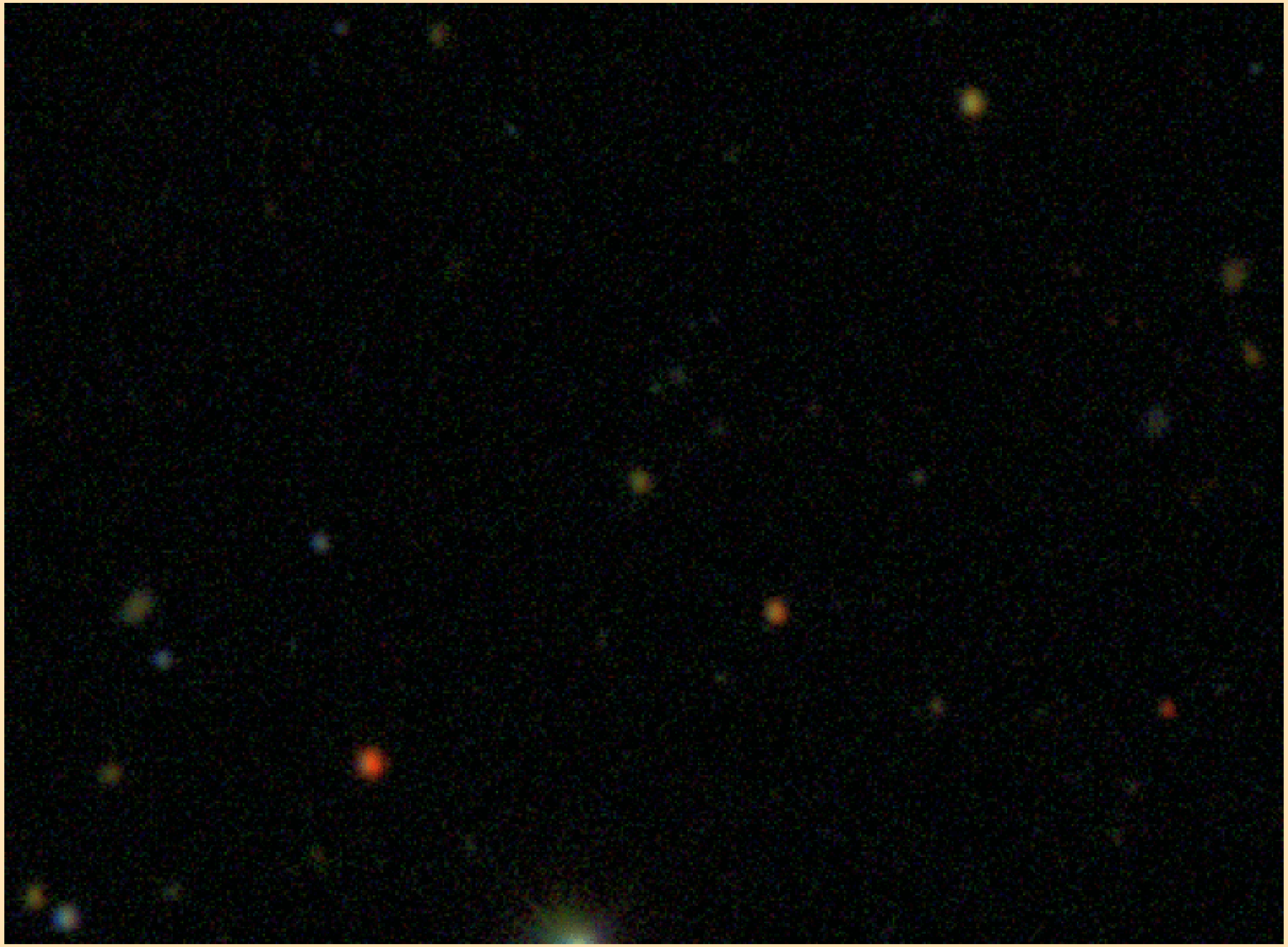


Image Addition

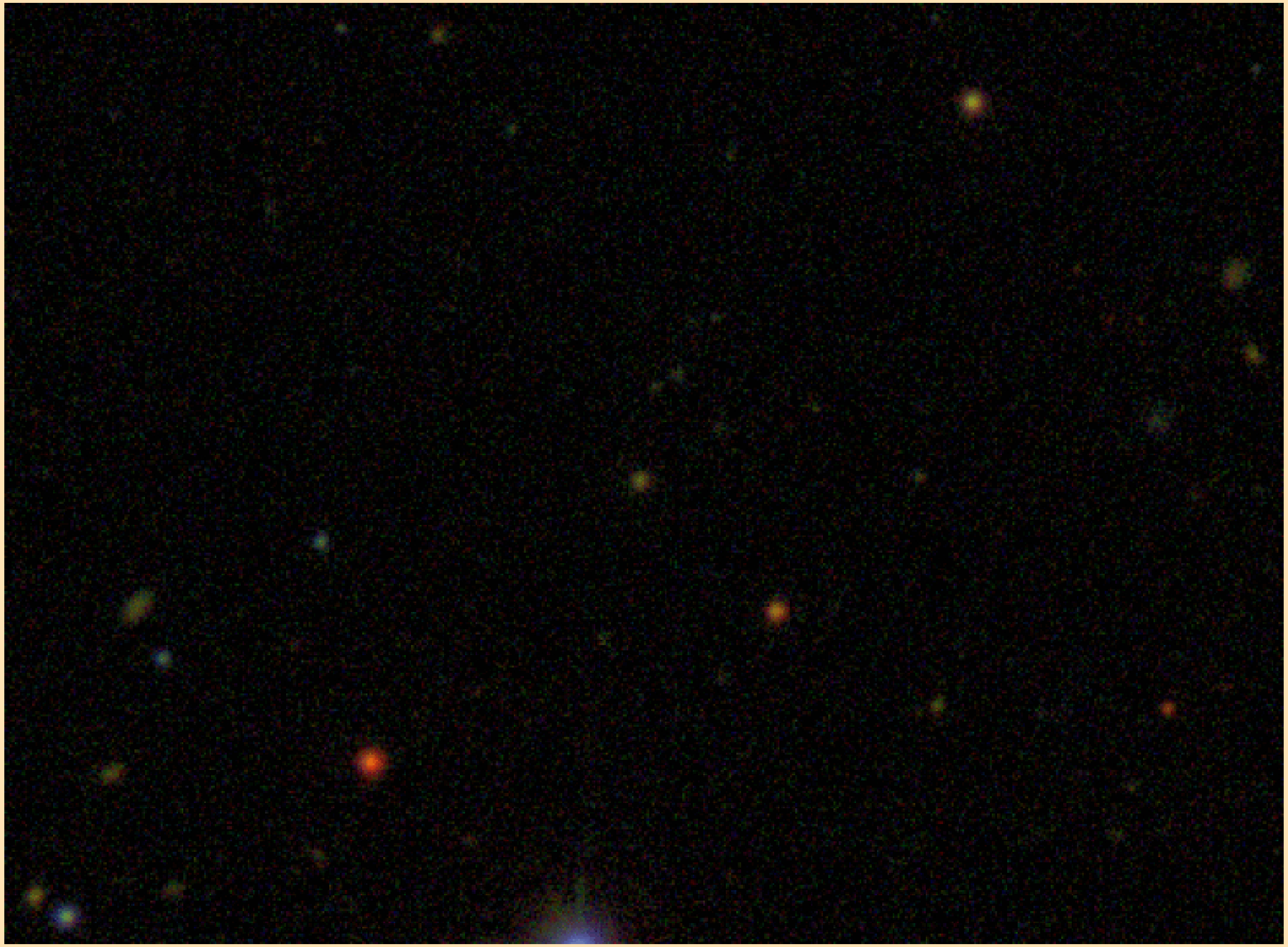


Image Addition

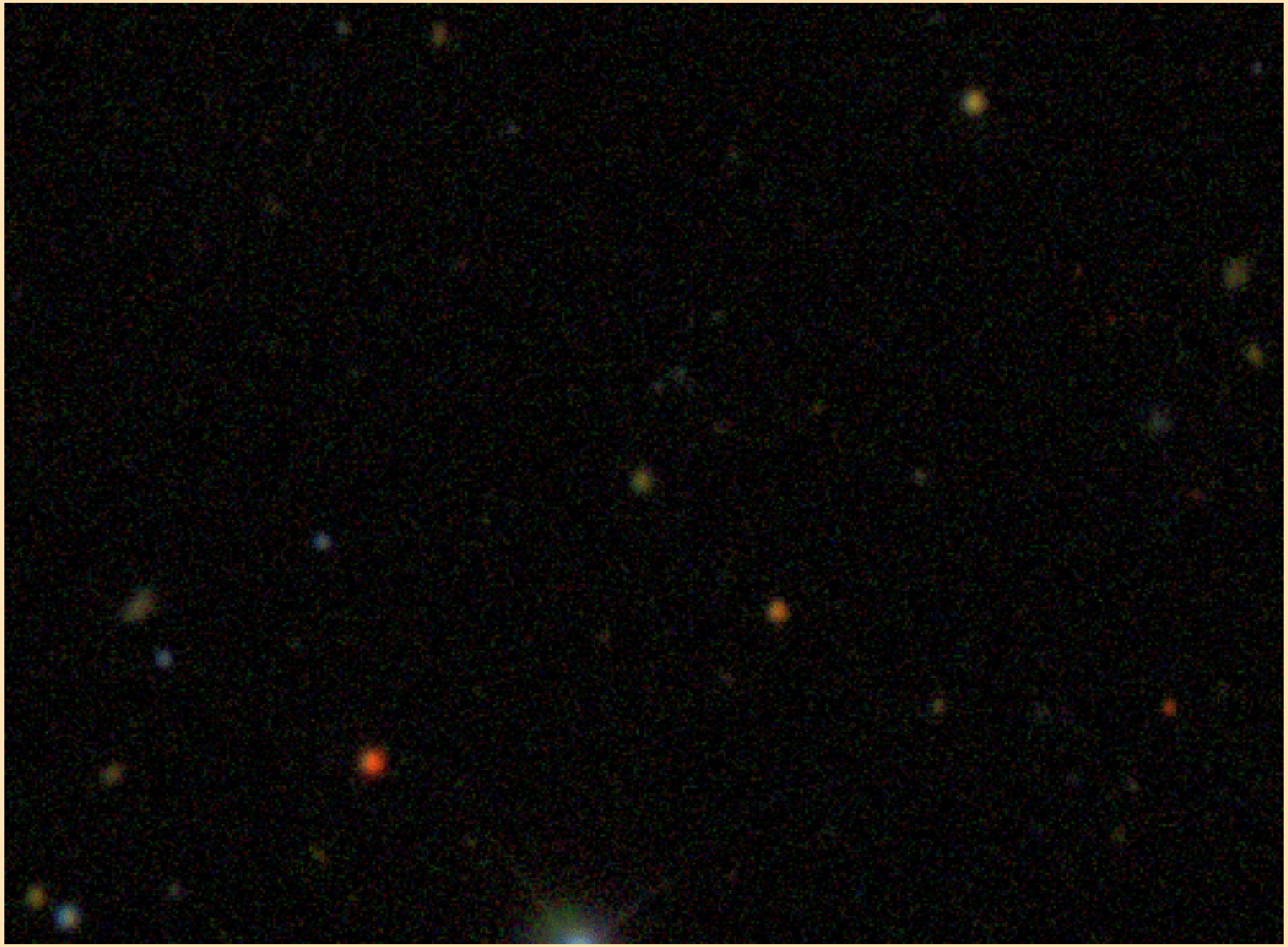


Image Addition

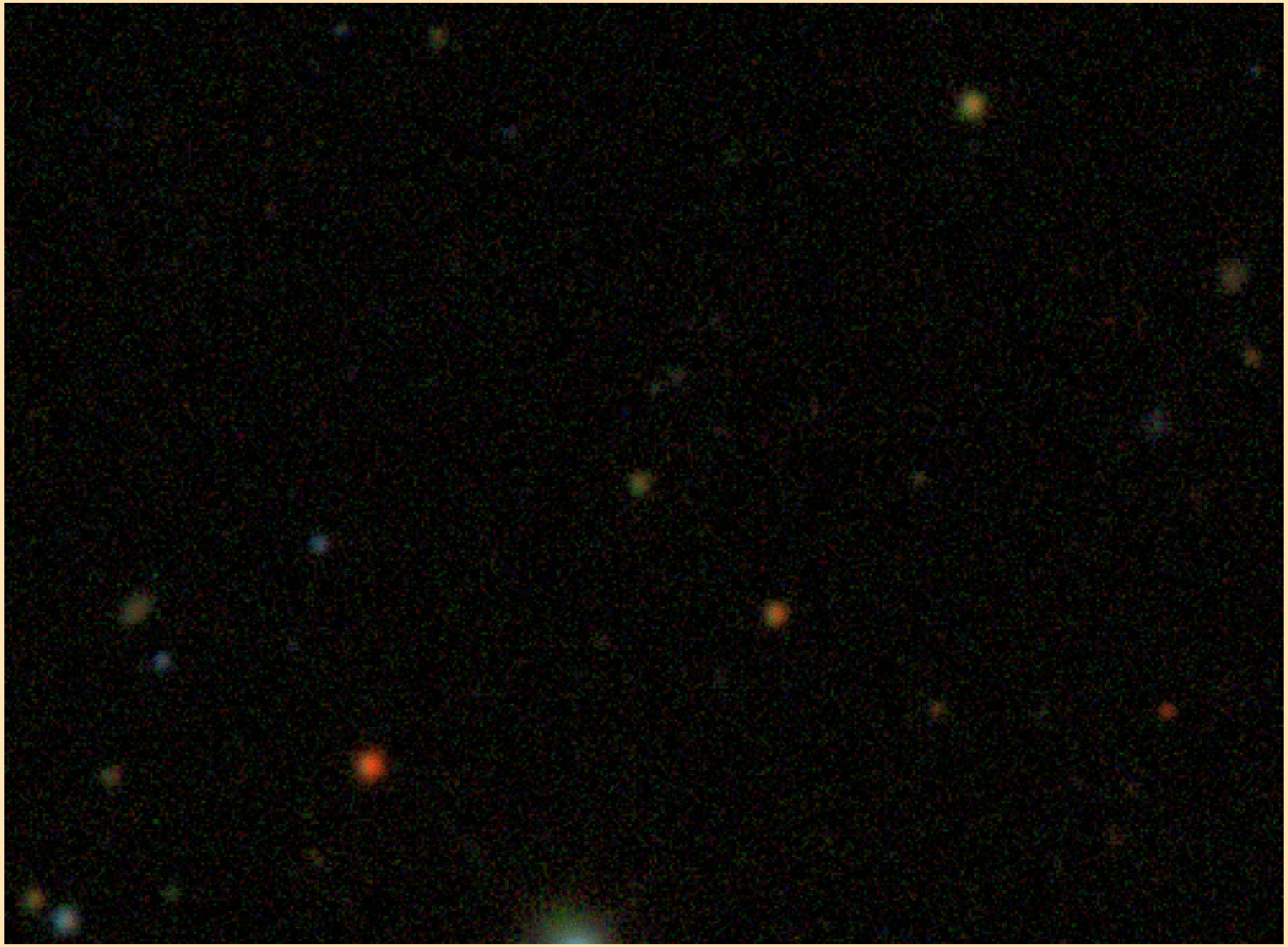


Image Addition

Image Addition

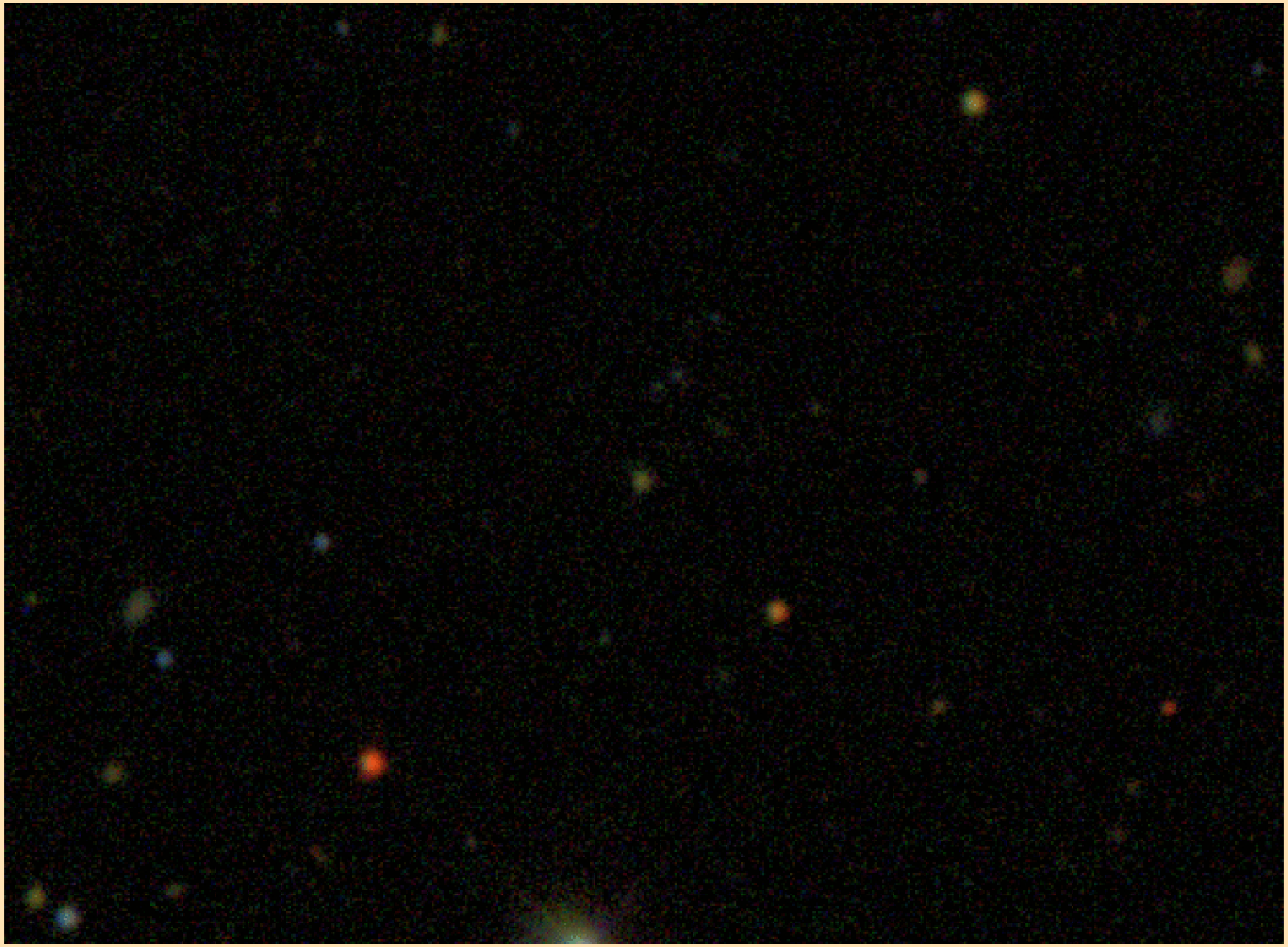


Image Addition

Image Addition

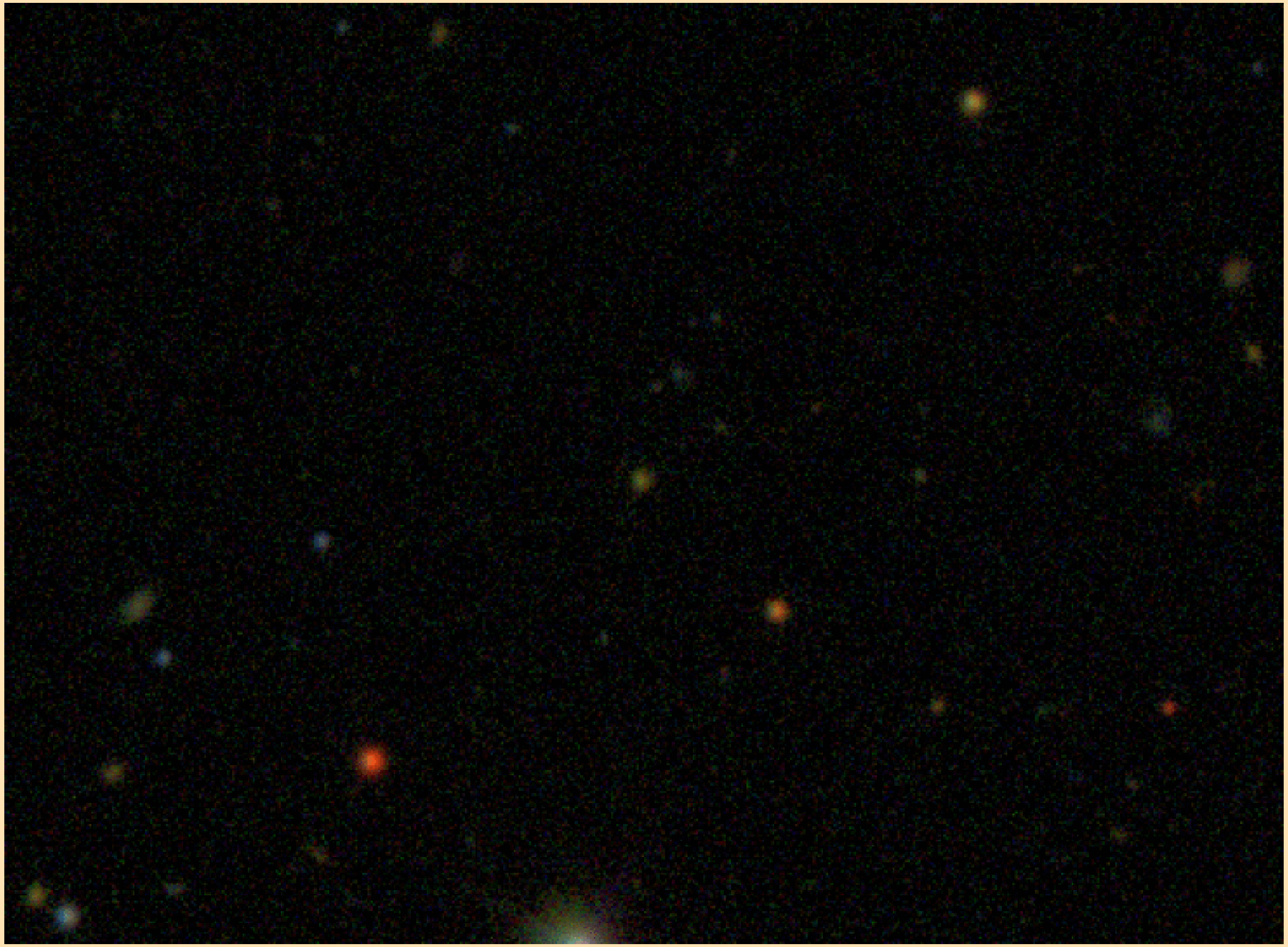


Image Addition

Image Addition

Image Addition

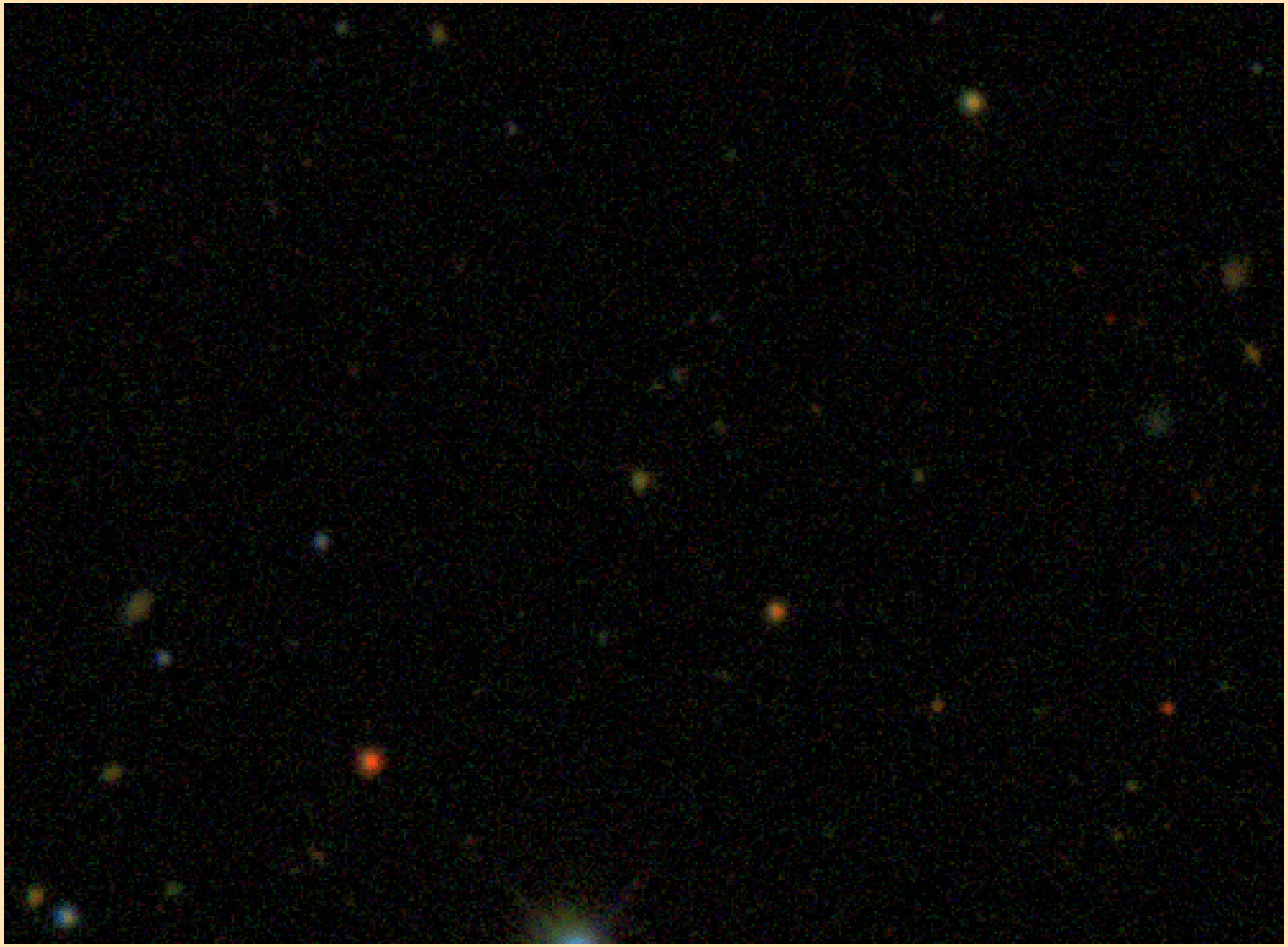


Image Addition

Image Addition

Image Addition

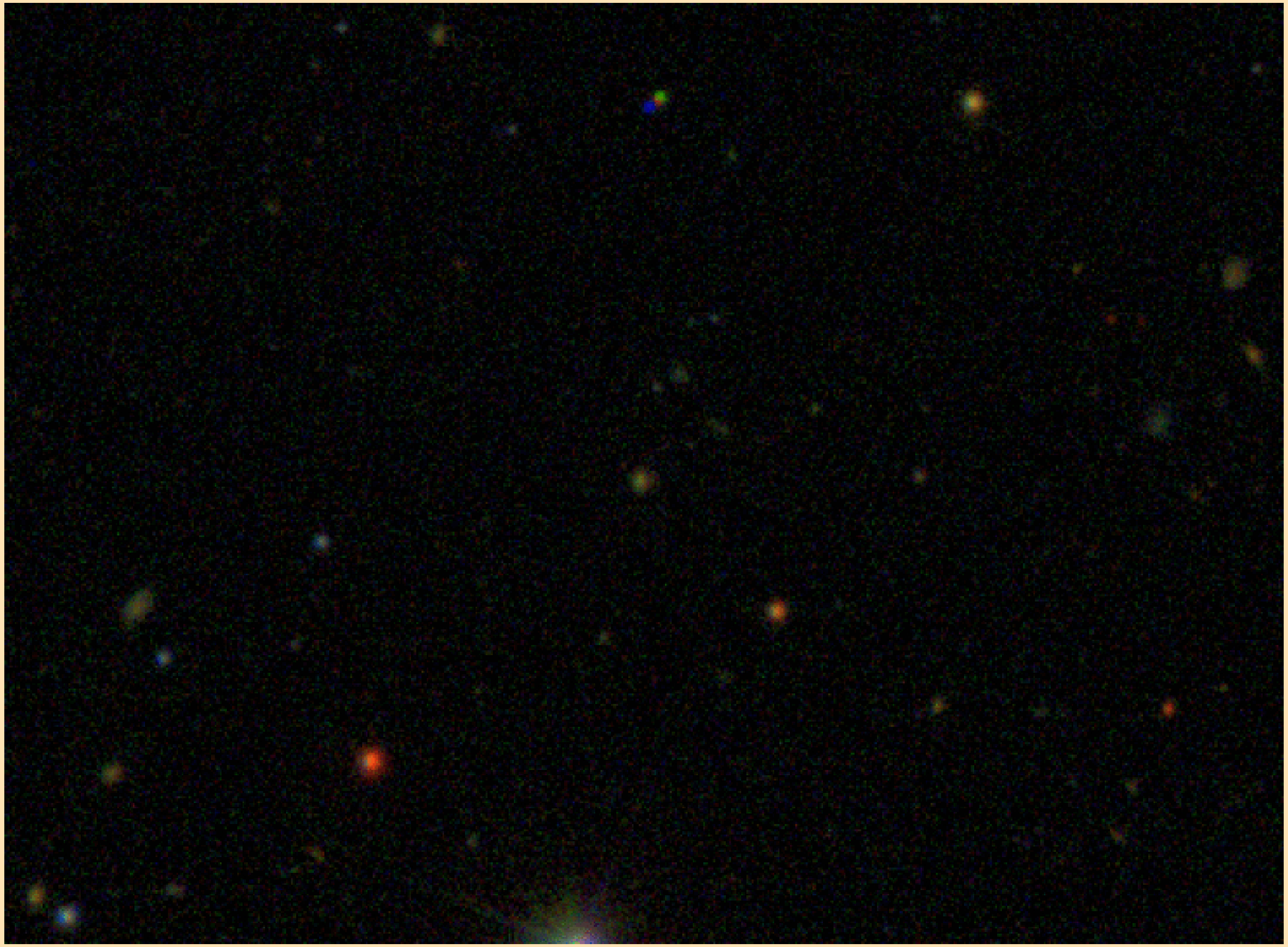


Image Addition

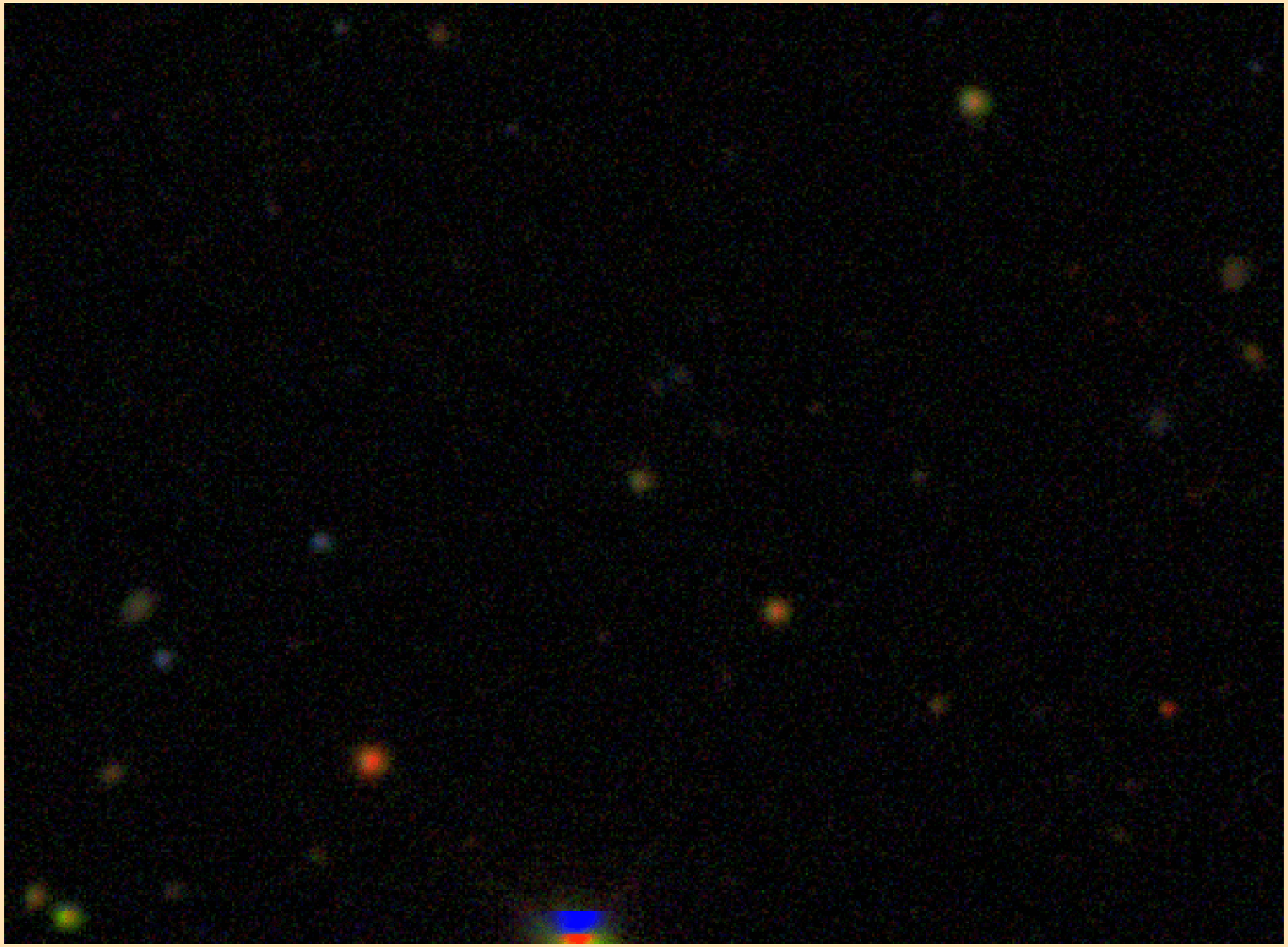


Image Addition

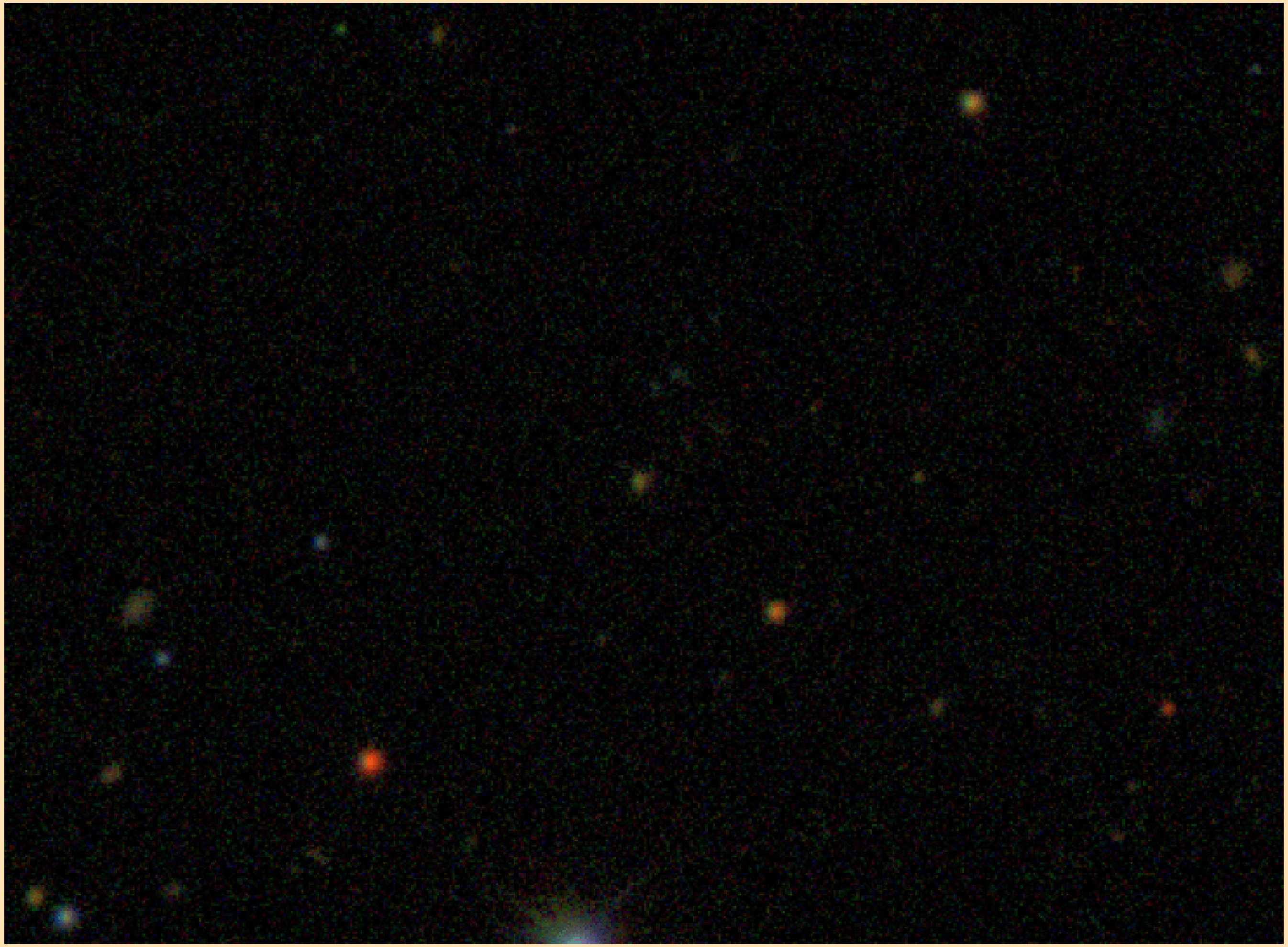


Image Addition

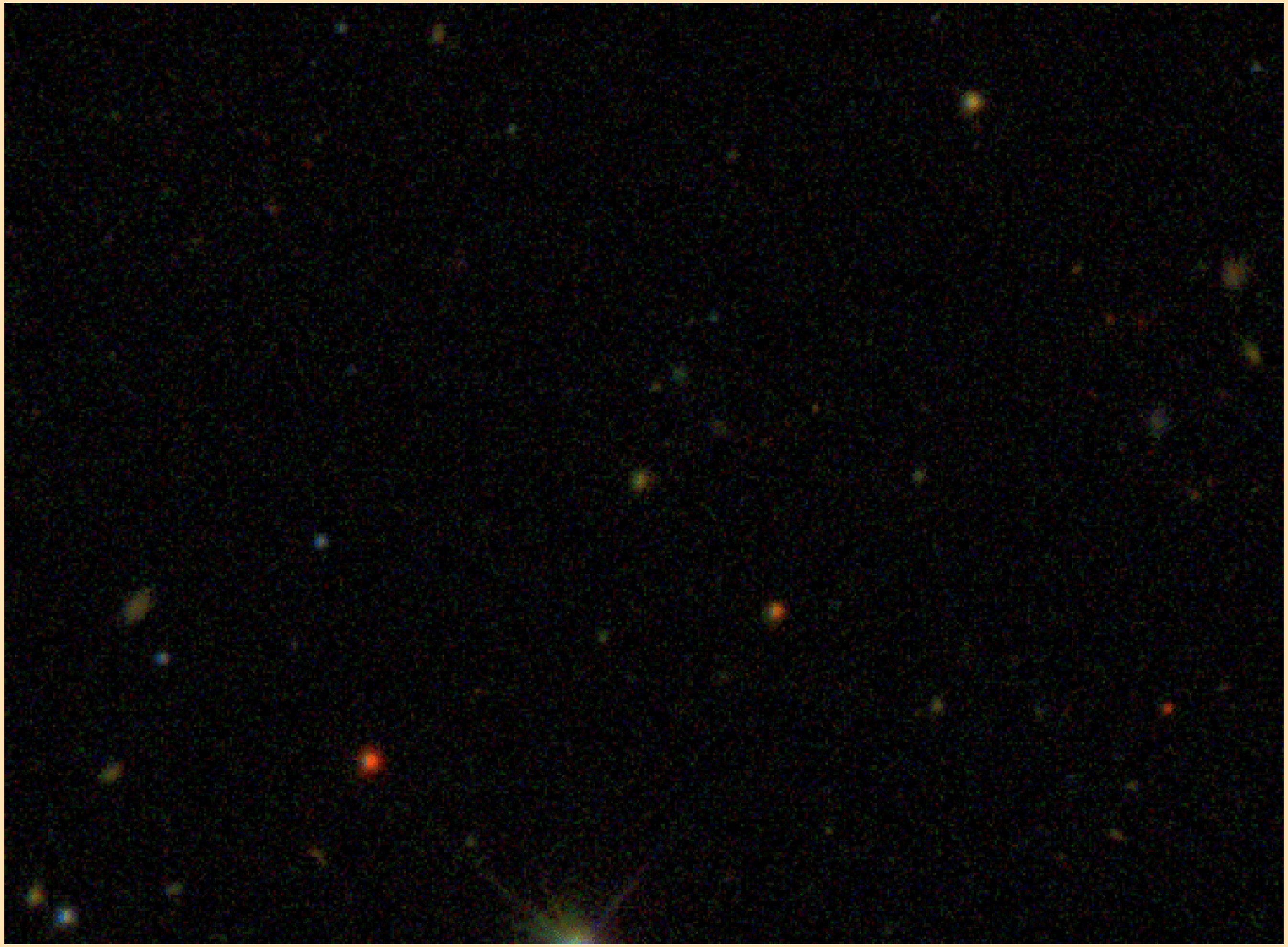


Image Addition

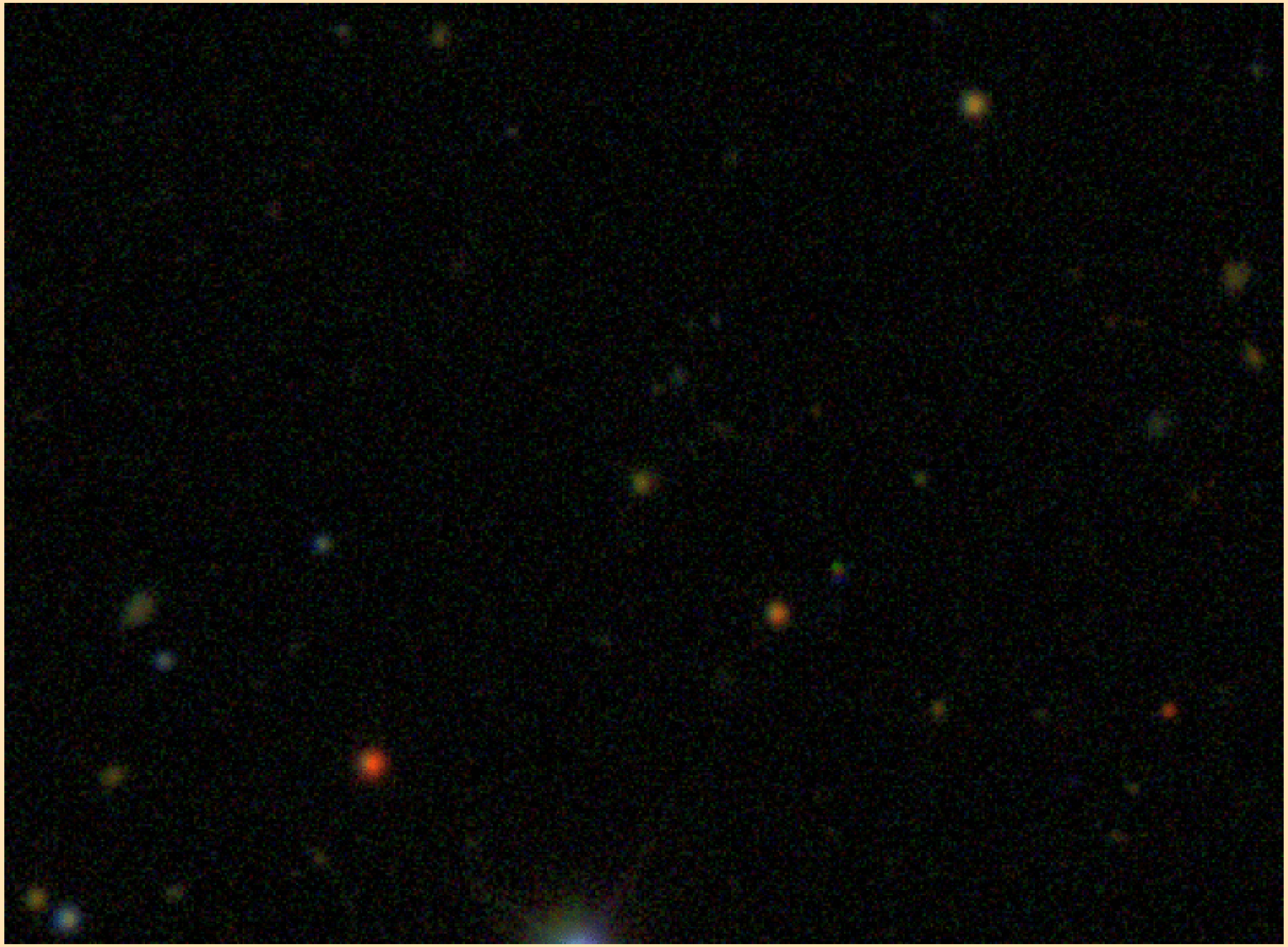
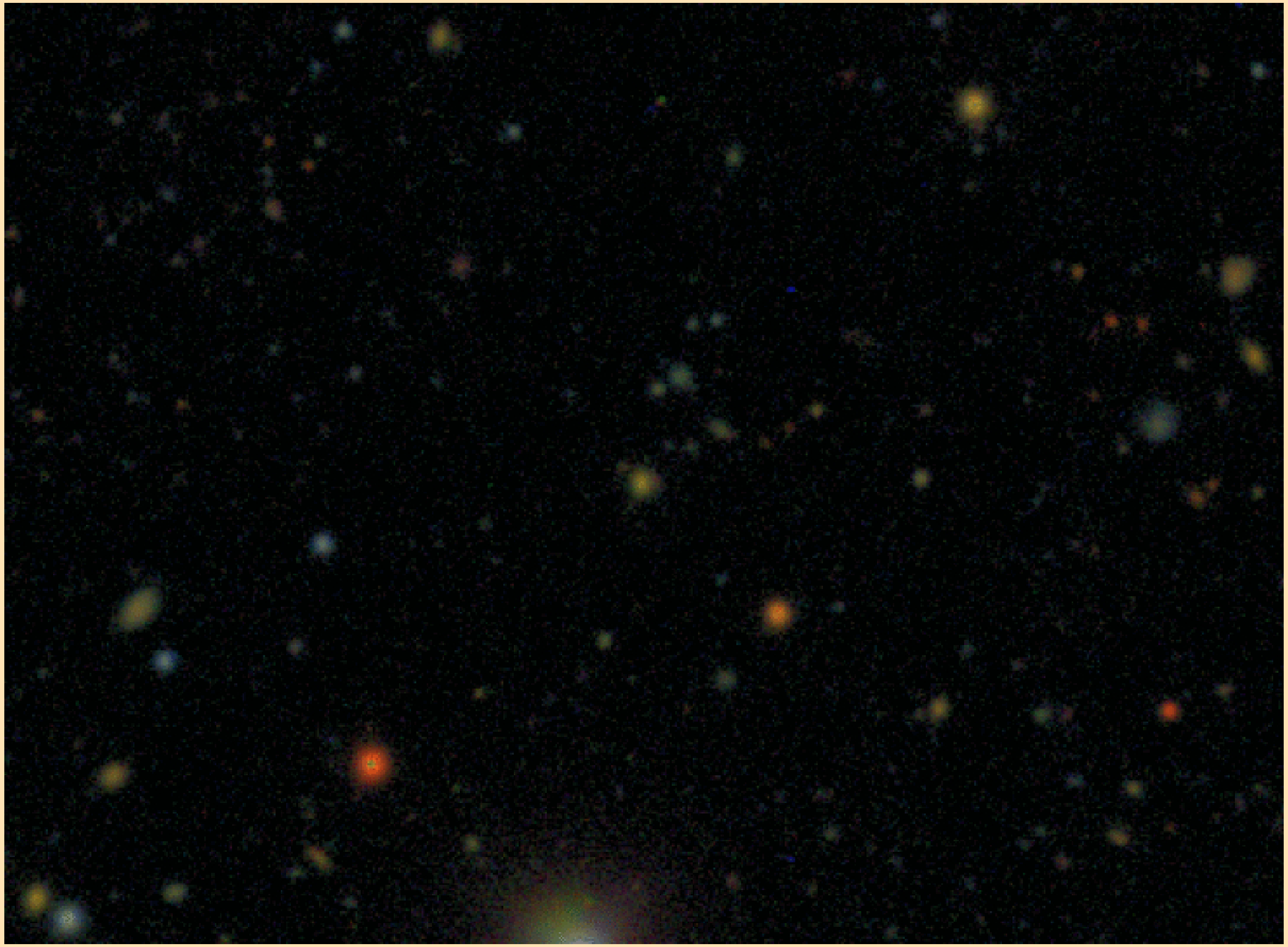


Image Addition



There are some situations when we probably really should want to add images; the generation of reference images for image subtraction is one such case.

There are some situations when we probably really should want to add images; the generation of reference images for image subtraction is one such case.

In other cases, it may well be better to handle an ensemble of N images as an estimation problem with N realisations of the data.

There are some situations when we probably really should want to add images; the generation of reference images for image subtraction is one such case.

In other cases, it may well be better to handle an ensemble of N images as an estimation problem with N realisations of the data.

In general each of these images has

1. a different (and possibly spatially varying) background level
2. its own footprint on the sky,
3. its own geometrical distortion
4. its own PSF structure.

Let us assume that the latter two are known.

Traditional Approaches; SWarp and Drizzle

Historically, most astronomers have used empirical approaches to adding images together (see also the poster by Jalobeanu et al.).

Traditional Approaches; SWarp and Drizzle

Historically, most astronomers have used empirical approaches to adding images together (see also the poster by Jalobeanu et al.).

For example, Bertin's SWarp :

1. Reads input images
2. Optionally reads corresponding variance maps
3. Constructs and removes a background image from each input
4. Resamples onto the desired output projection
5. Averages all of the resampled input images

Issues

Subtracting background levels is tricky

Issues

Subtracting background levels is tricky

The PSF in the final image will, in general, be discontinuous

Issues

Subtracting background levels is tricky

The PSF in the final image will, in general, be discontinuous

The coadded image's noise will be correlated

Issues

Subtracting background levels is tricky

The PSF in the final image will, in general, be discontinuous

The coadded image's noise will be correlated

The input images may not be well sampled.

One popular strategy, especially for HST data, is to use Drizzle ; the basic idea here is to map the observed pixels onto a fully-sampled grid, but to pretend that all the flux in concentrated into the central p% (typically $p \sim 50$) of the physical pixel. This preserves surface brightness and has good noise properties, but fails to preserve the PSF.

One popular strategy, especially for HST data, is to use Drizzle ; the basic idea here is to map the observed pixels onto a fully-sampled grid, but to pretend that all the flux in concentrated into the central p% (typically $p \sim 50$) of the physical pixel. This preserves surface brightness and has good noise properties, but fails to preserve the PSF.

For the case that the PSF is known and constant for all exposures, and in the absence of noise, we may use standard iterative techniques for non-uniformly sampled data.

One popular strategy, especially for HST data, is to use Drizzle ; the basic idea here is to map the observed pixels onto a fully-sampled grid, but to pretend that all the flux in concentrated into the central p% (typically $p \sim 50$) of the physical pixel. This preserves surface brightness and has good noise properties, but fails to preserve the PSF.

For the case that the PSF is known and constant for all exposures, and in the absence of noise, we may use standard iterative techniques for non-uniformly sampled data.

It is unclear how well these will scale for very large, noisy, datasets.

Estimation of the True Sky

An alternative approach that has not, so far as I know, been employed in anger was proposed by Kaiser.

Estimation of the True Sky

An alternative approach that has not, so far as I know, been employed in anger was proposed by Kaiser.

If an point in an image is given by f_p where the pixel index p runs over all the pixels in a set of images and the p^{th} pixel is associated with a PSF g_p, in the limit of uncorrelated Gaussian noise n_p, we may write down the likelihood of the (Fourier transform) true sky $\boldsymbol{I}$, thought of as a vector, as

$$\mathcal{L}(\boldsymbol{I}) = \boldsymbol{I}\boldsymbol{\phi} - \frac{1}{2}\boldsymbol{I}A^{-1}\boldsymbol{I} + \text{constant}$$

where $\boldsymbol{\phi} \equiv \sum_p f_p g_p / \sigma_p^2$ and $A \equiv \sum_p g_p g_p / \sigma_p^2$.

Estimation of the True Sky

An alternative approach that has not, so far as I know, been employed in anger was proposed by Kaiser.

If an point in an image is given by f_p where the pixel index p runs over all the pixels in a set of images and the p^{th} pixel is associated with a PSF g_p, in the limit of uncorrelated Gaussian noise n_p, we may write down the likelihood of the (Fourier transform) true sky $\boldsymbol{I}$, thought of as a vector, as

$$\mathcal{L}(\boldsymbol{I}) = \boldsymbol{I}\boldsymbol{\phi} - \frac{1}{2}\boldsymbol{I}A^{-1}\boldsymbol{I} + \text{constant}$$

where $\boldsymbol{\phi} \equiv \sum_p f_p g_p/\sigma_p^2$ and $A \equiv \sum_p g_p g_p/\sigma_p^2$.

This is minimised at the point

$$\bar{\boldsymbol{I}} = A^{-1}\boldsymbol{\phi}$$

As Kaiser points out, multiplying by A^{-1} is tantamount to deconvolving the orginal data, but nevertheless the data f_p enter into the estimator $\boldsymbol{I}$ only through the sum $\boldsymbol{\phi}$; in otherwords, $\boldsymbol{\phi}$ is a sufficient statistic for $\boldsymbol{I}$.

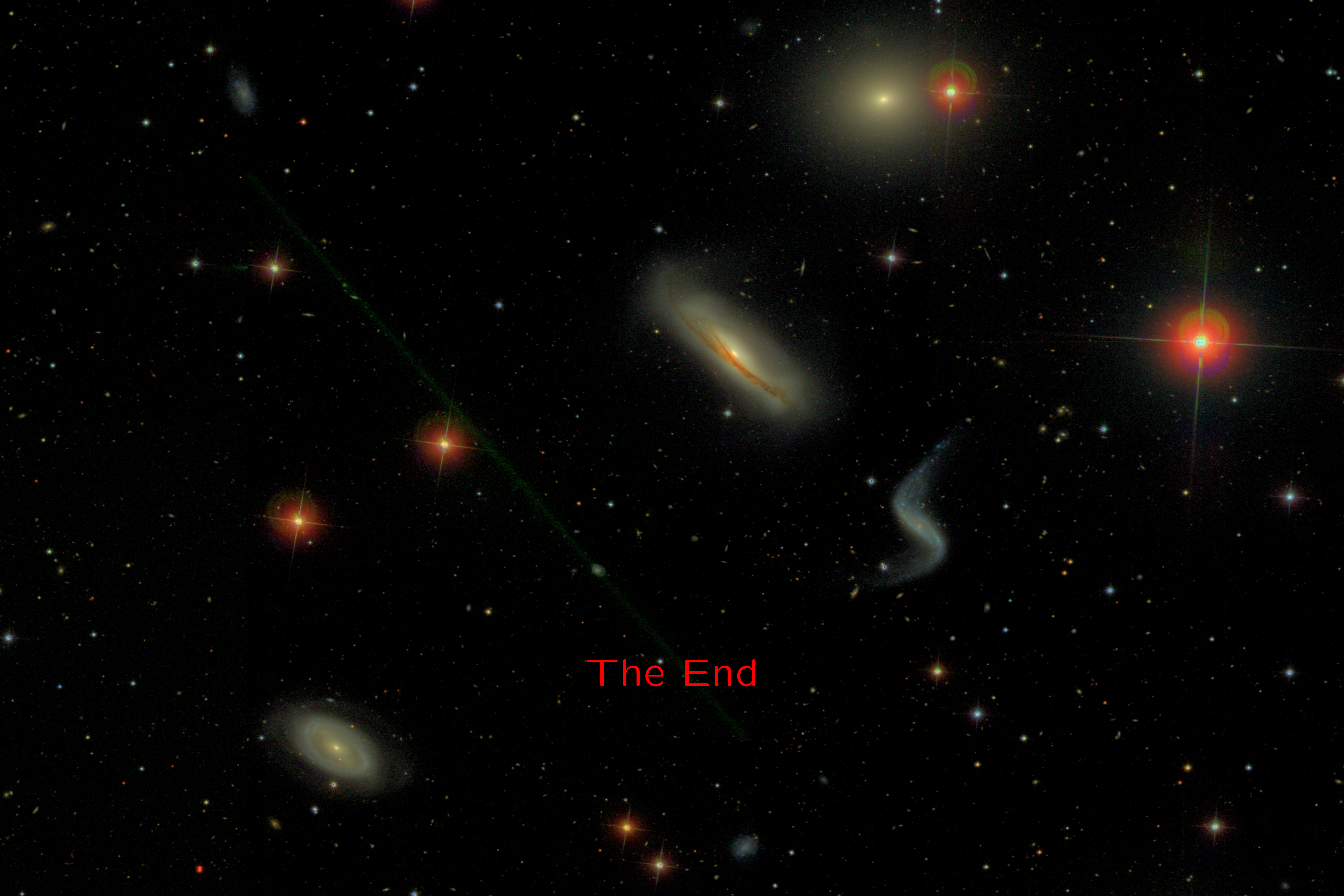
As Kaiser points out, multiplying by A^{-1} is tantamount to deconvolving the orginal data, but nevertheless the data f_p enter into the estimator $\boldsymbol{I}$ only through the sum $\boldsymbol{\phi}$; in otherwords, $\boldsymbol{\phi}$ is a sufficient statistic for $\boldsymbol{I}$.

In real space,

$$\boldsymbol{\phi}(\boldsymbol{r}) = \sum_i \left(g_i^\dagger \otimes f_i\right)_{\boldsymbol{r}} / \sigma_i^2$$

which has effective PSF

$$\sum_i \left(g_i^\dagger \otimes g_i\right)_{\boldsymbol{r}} / \sigma_i^2$$


The End