

The Devil’s in the Details: Photometric Biases in Modern Surveys

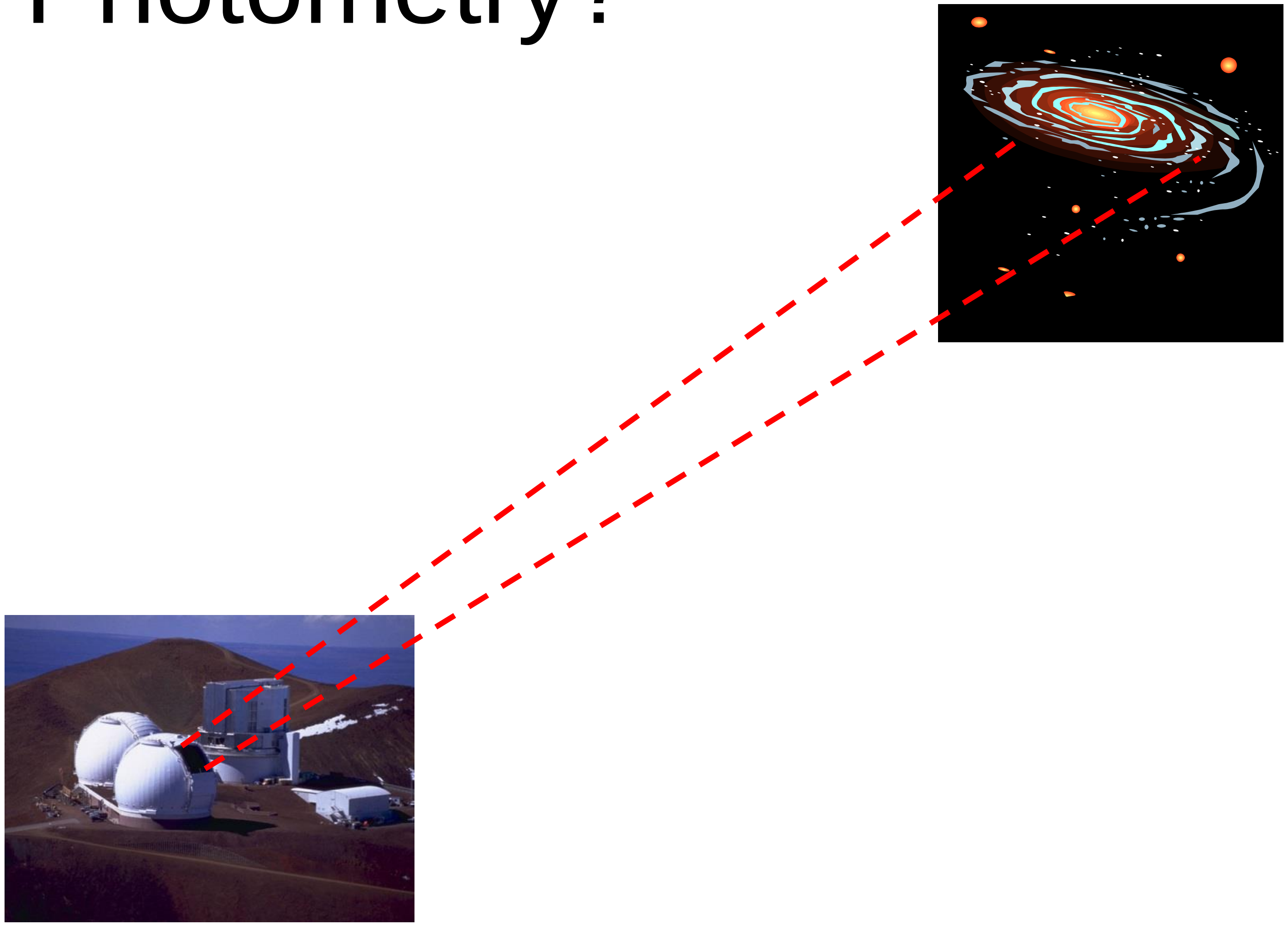
Stephen Portillo[1,2,*] and **Josh Speagle**[1,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington)

[*]Equal contribution

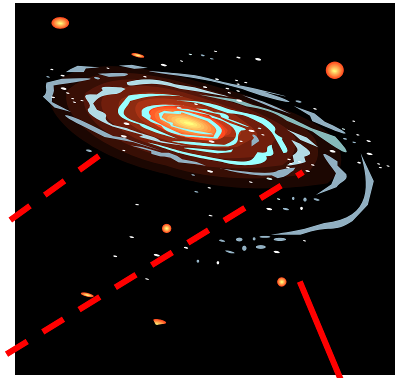
Portillo, Speagle, & Finkbeiner (**PSF**) subm., arxiv:1902.02374

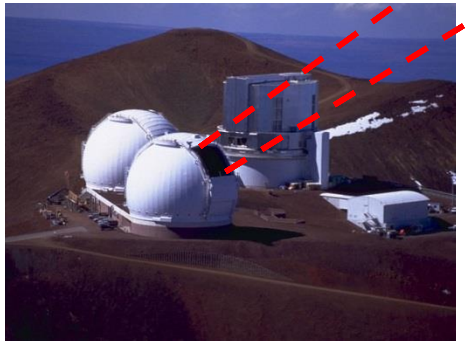
What is Photometry?



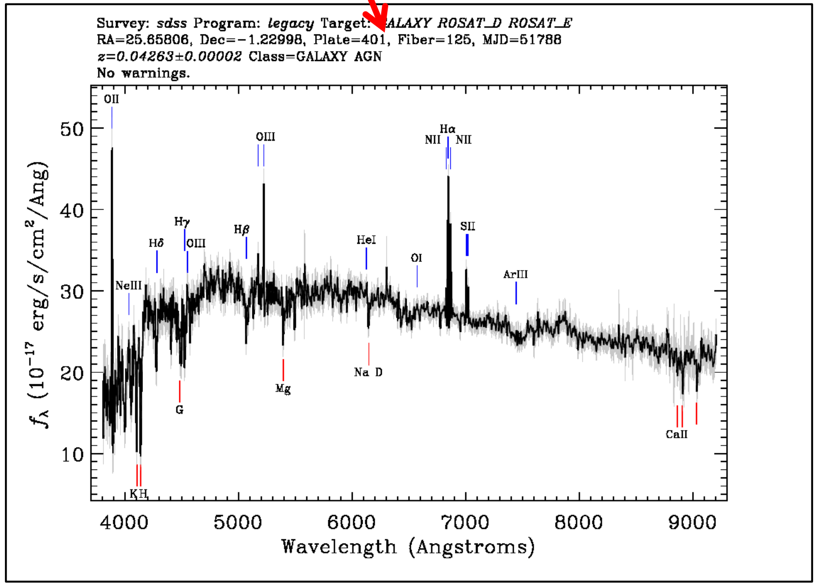
Keck Telecope (Hawaii)

What is Photometry?

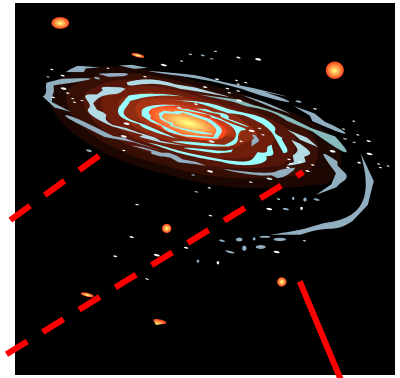


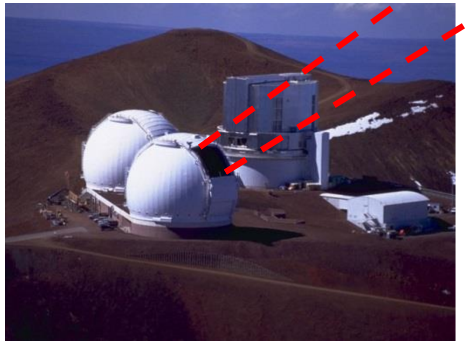


Keck Telecope (Hawaii)

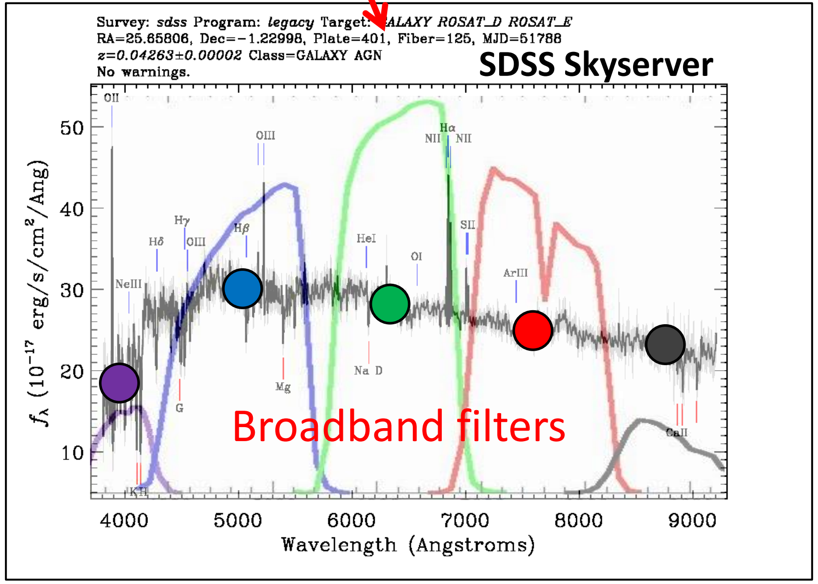


What is Photometry?



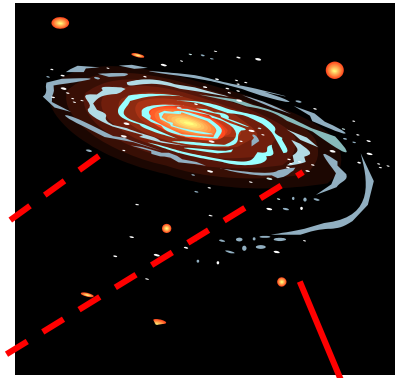


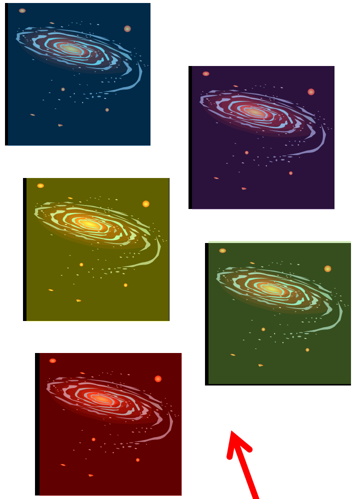
Keck Telecope (Hawaii)

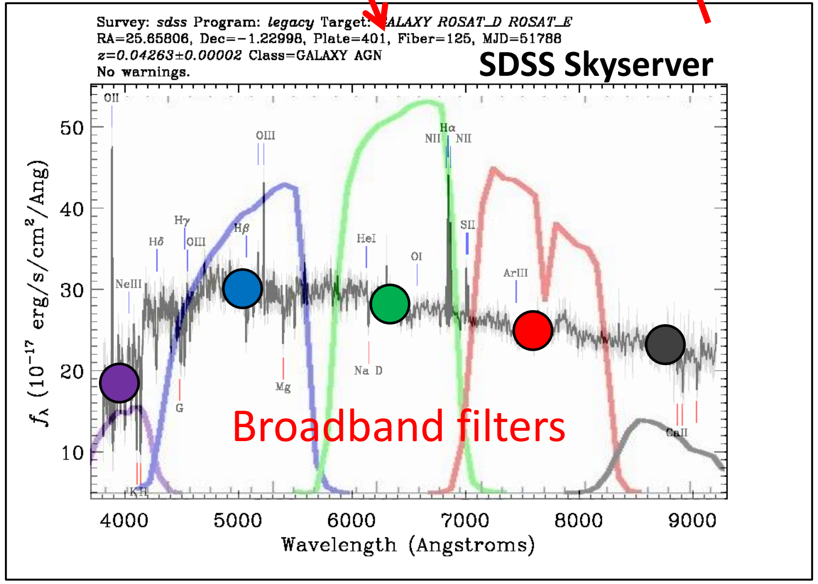


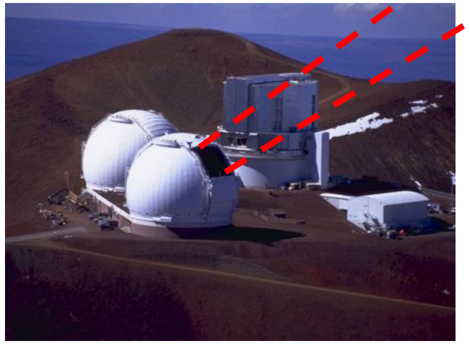
What is Photometry?

Spectral Energy Distribution (SED)



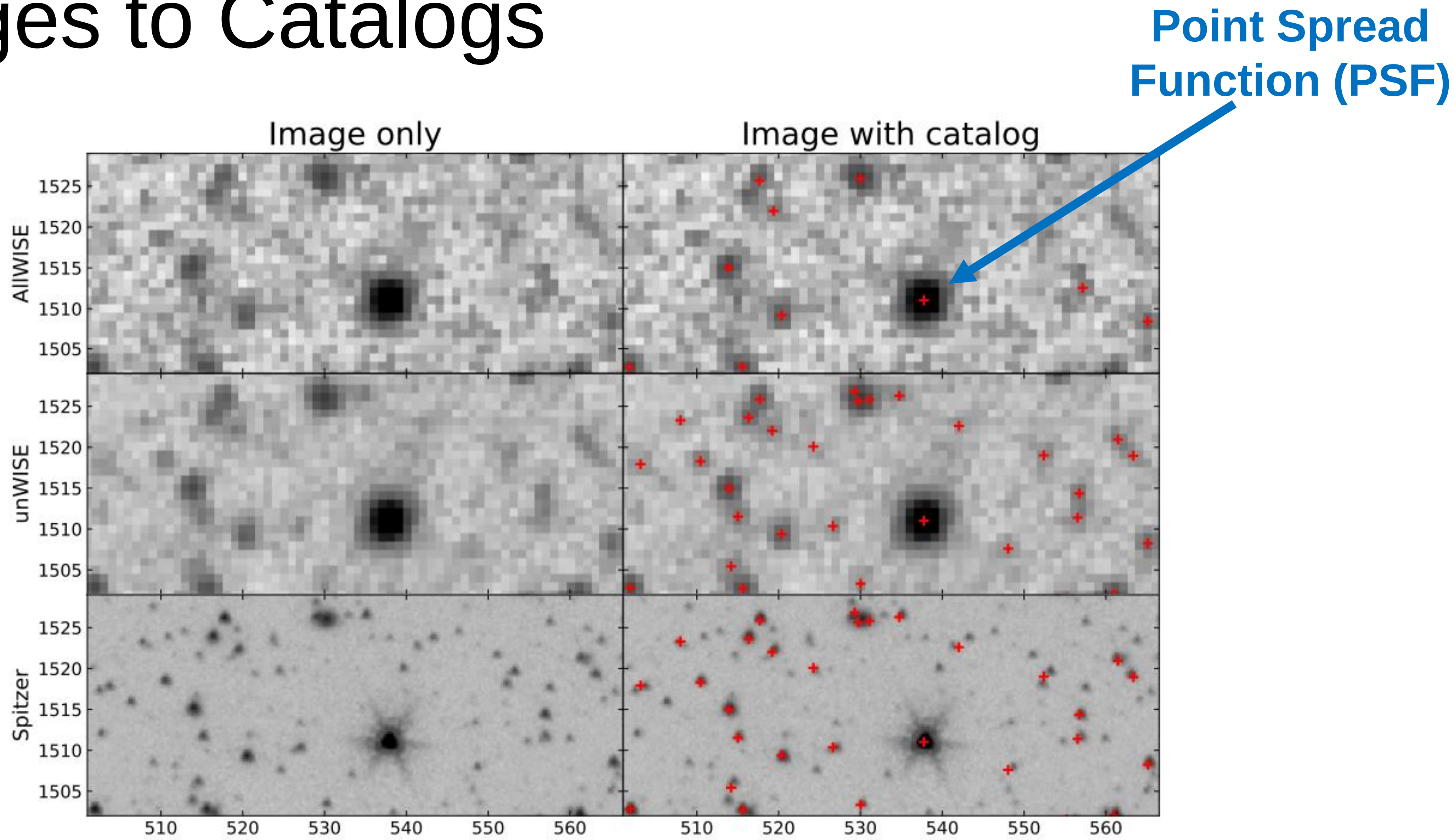






Keck Telecope (Hawaii)

Images to Catalogs



unWISE: Schlafly et al. (2019)

So…photometry?

- Most of my work focuses on using photometry from large surveys.

So…photometry?

"Big Data"-oriented work

- Most ~~of my work~~ focuses on using photometry from large surveys.

So…photometry?

“Big Data”-oriented work

- Most ~~of my work~~ focuses on using photometry from large surveys.
- Understanding the data is important.

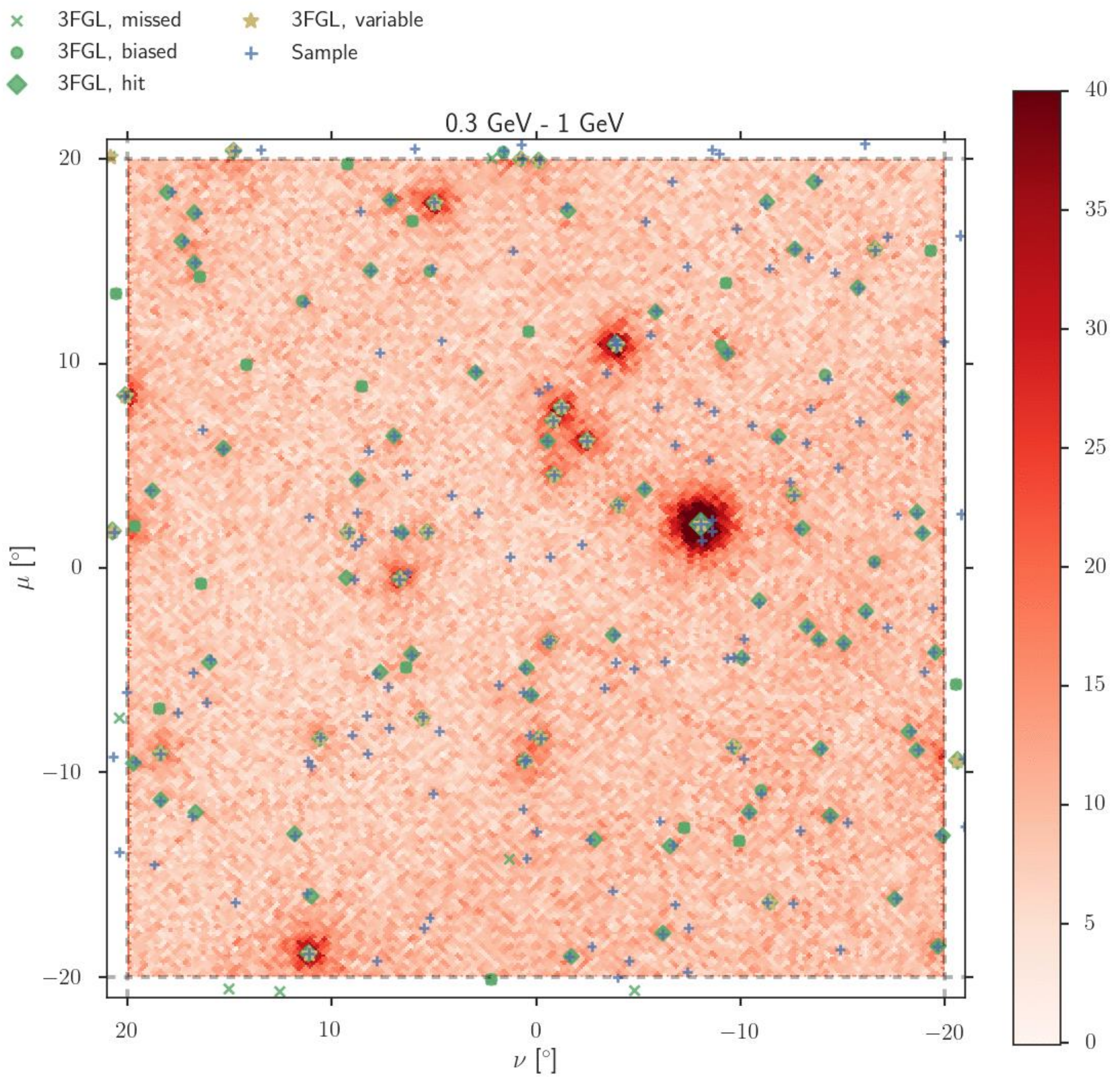
So…photometry?

“Big Data”-oriented work

- Most ~~of my work~~ focuses on using photometry from large surveys.
- Understanding the data is important.
- Small effects can add up over large populations.

Starting Point

- The Finkbeiner group built up **PCAT** (Probabilistic Cataloging):
  - sampling from the transdimensional space of all possible catalogs

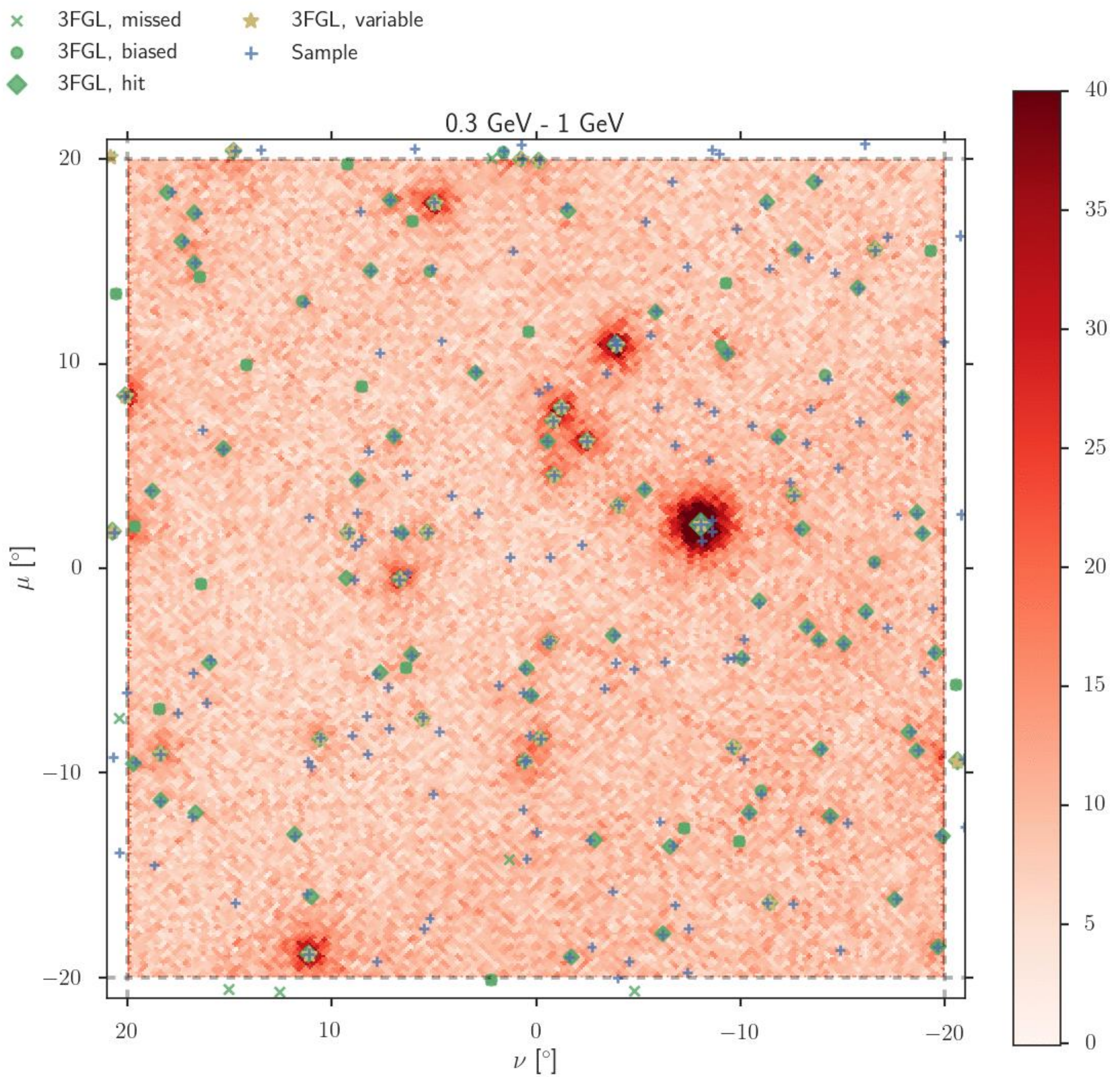


Daylan, Portillo, & Finkbeiner (2016)

Starting Point

- The Finkbeiner group built up **PCAT** (Probabilistic Cataloging):
  - sampling from the transdimensional space of all possible catalogs

PCAT



How well can you model a single source?

Daylan, Portillo, & Finkbeiner (2016)

Results

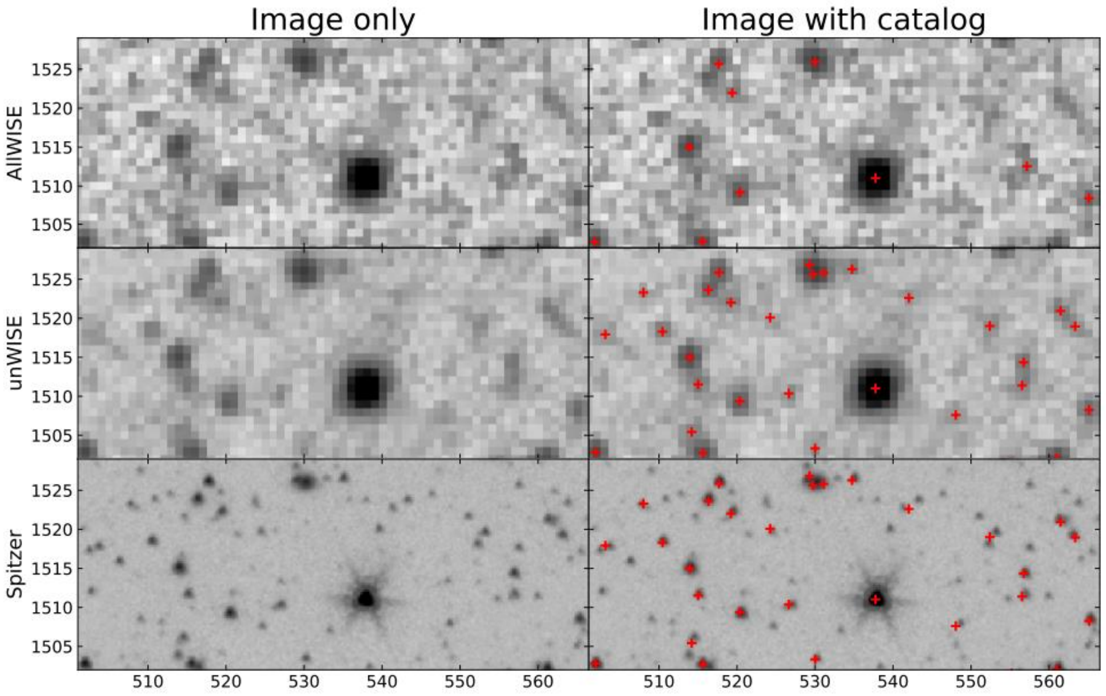
- Estimated fluxes are biased.
- Uncertainties are underestimated.

Results

- Estimated fluxes are biased.
- Uncertainties are underestimated.

First reaction:

- No surprise: photometry is hard.
- Model mismatch (PSF, source)
- Blending issues
- Background estimation
- Unresolved sources
- Detection limits/selection effects
- Etc.



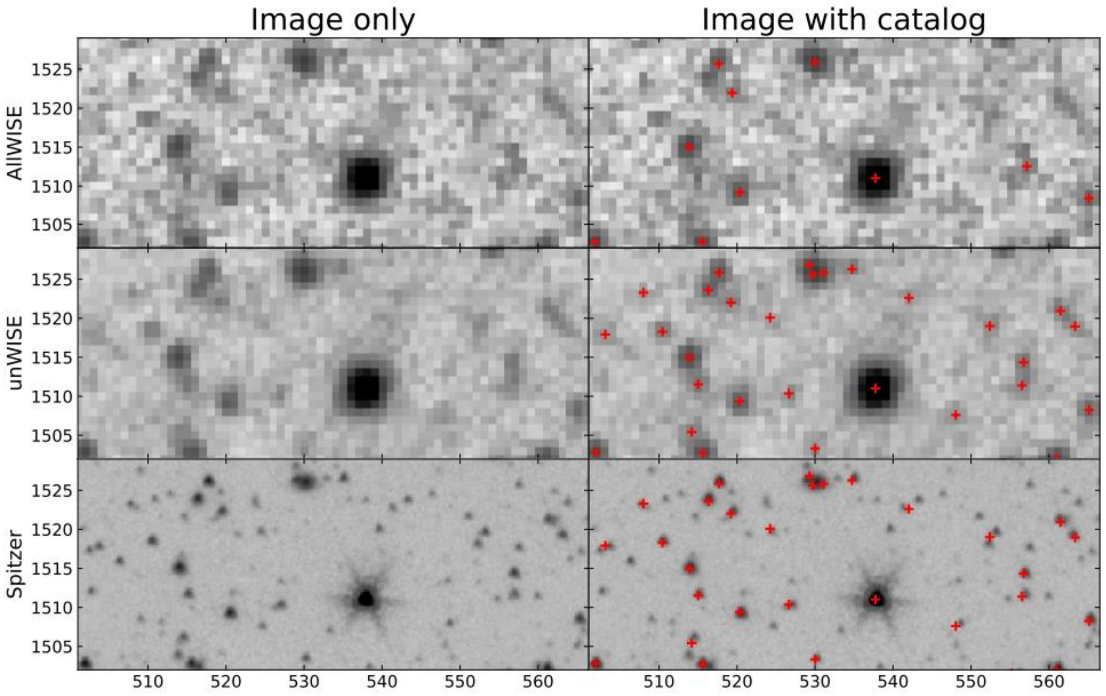
unWISE: Schlafly et al. (2019)

Results

- Estimated fluxes are biased.
- Uncertainties are underestimated.

First r[illegible]action:

- N[illegible]ise: photo[illegible] hard.
- M[illegible]t[illegible]rce)
- Blend[illegible]
- Bac[illegible]
- [illegible] sou[illegible]
- De[illegible]ion limits/selec[illegible] effects
- Etc.



unWISE: Schlafly et al. (2019)

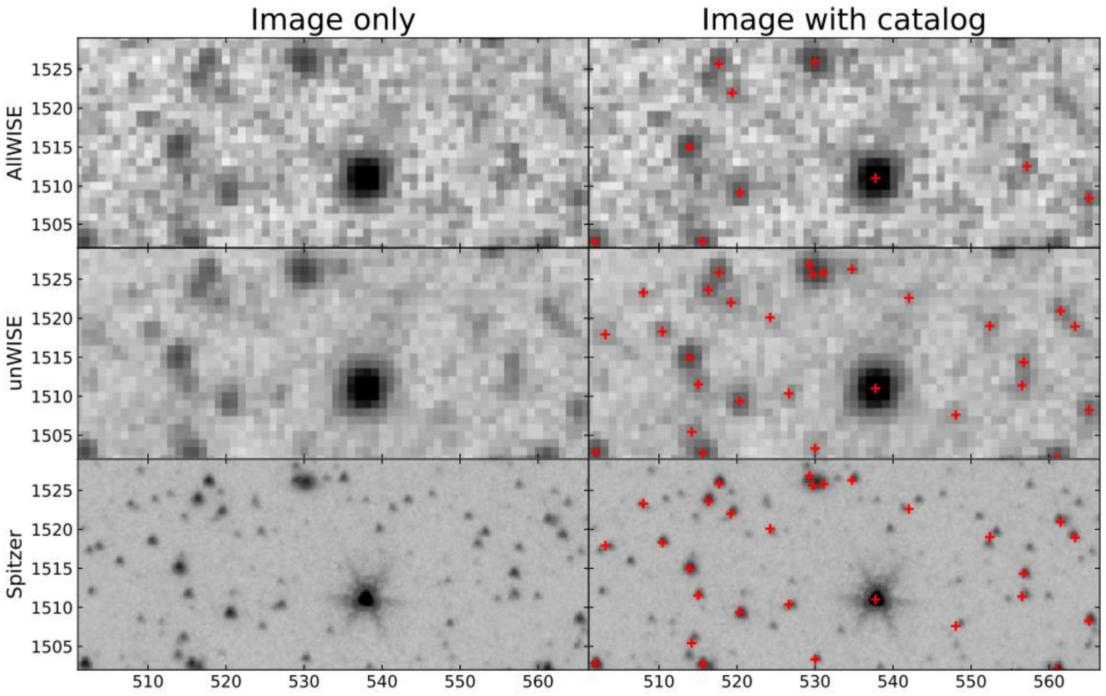
Results

- Estimated fluxes are biased.
- Uncertainties are underestimated.

This is true even with perfect models and data!

First reaction:

- N... ise: photo... hard.
- M... rce)
- Blend...
- Bac...
- ... sou...
- D... ion limits/sele... effects
- Etc.

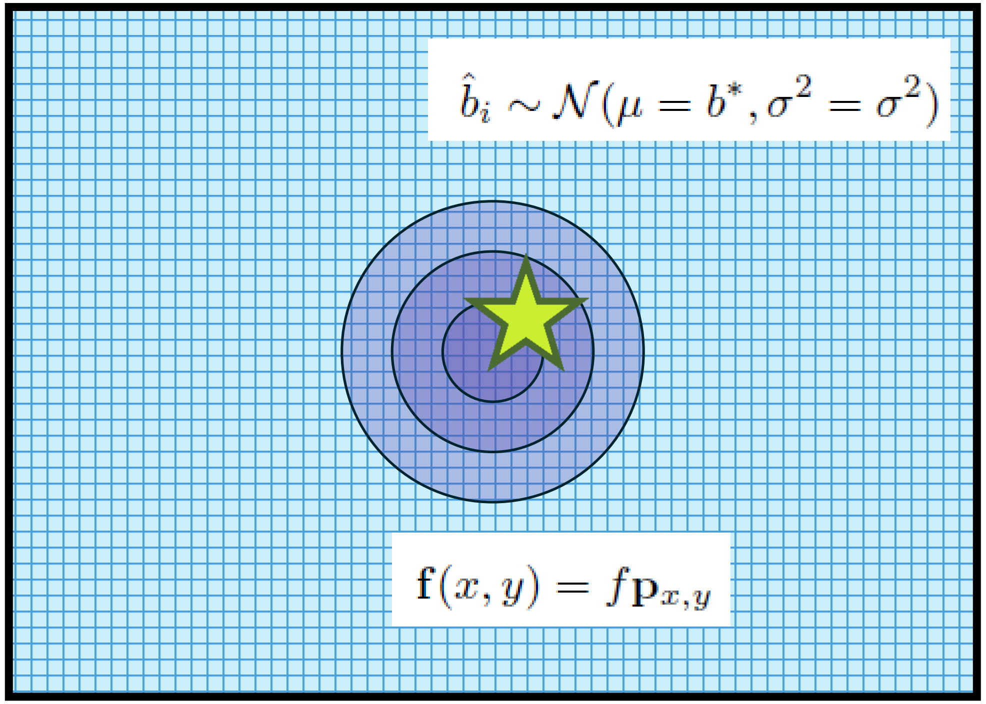


unWISE: Schlafly et al. (2019)

Starting Point

- Single, isolated **point source** in **one band** with PSF known and Gaussian background noise.

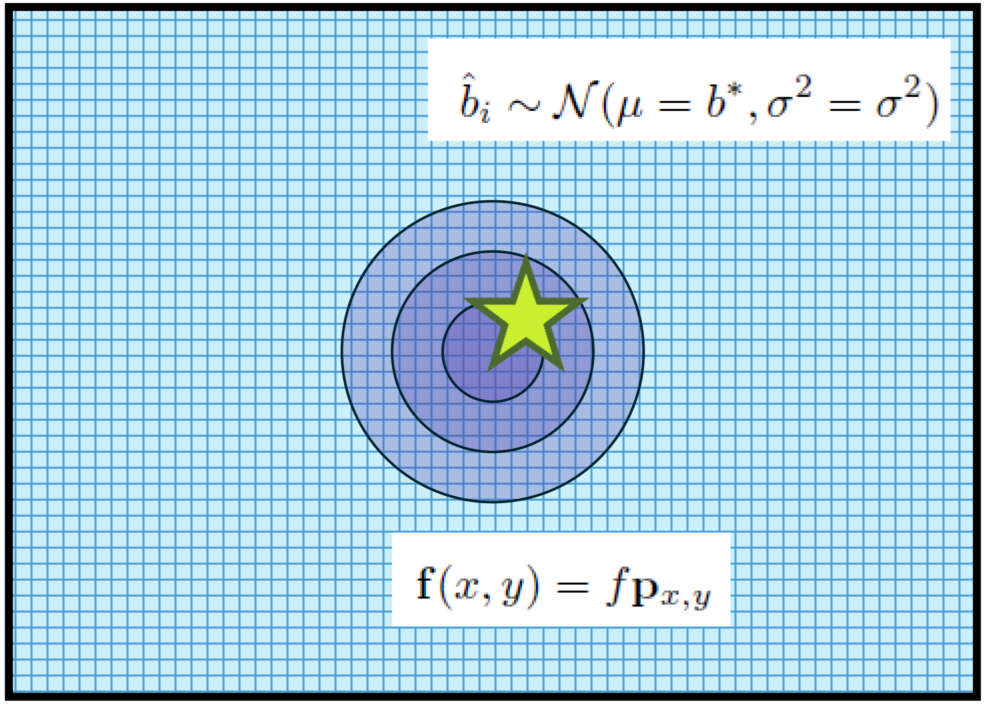
$n \times m$ footprint



Starting Point

- Single, isolated **point source** in **one band** with PSF known and Gaussian background noise.

$n \times m$ footprint



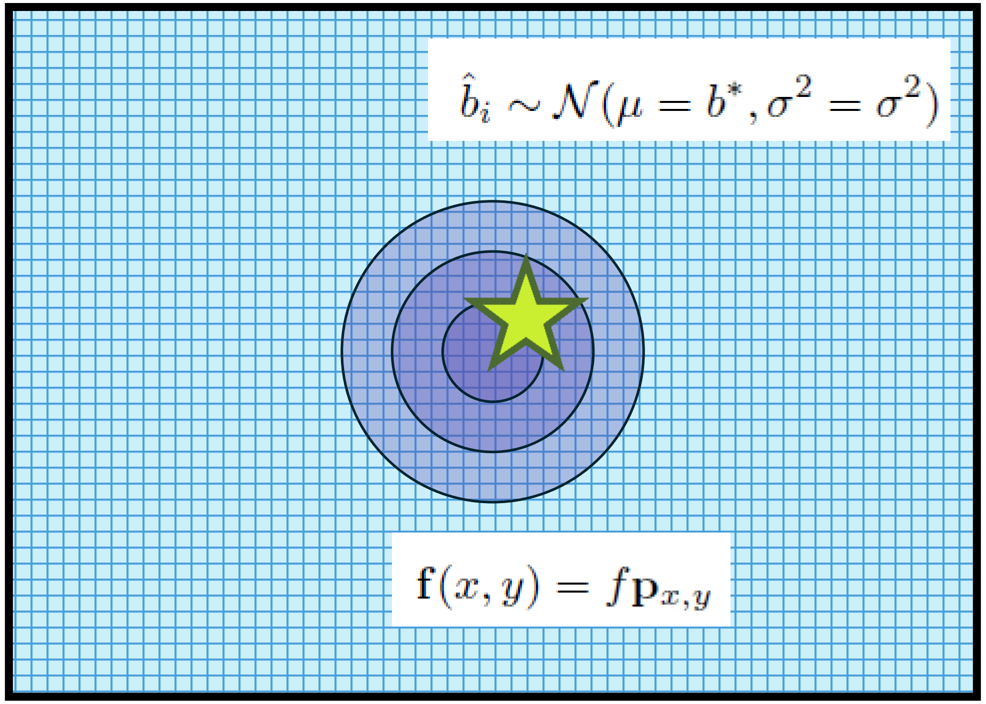
Likelihood

$$\ln \mathcal{L}(x, y, f, b)$$

Starting Point

- Single, isolated **point source** in **one band** with PSF known and Gaussian background noise.

$n \times m$ footprint



Likelihood

$$\ln \mathcal{L}(x, y, f, b)$$

Maximum-Likelihood Solution

$$\partial_f \ln \mathcal{L}(x, y, f, b) = 0$$

Starting Point

- Normal log-likelihood:

$$\ln \mathcal{L}(x, y, f, b) = -\frac{nm}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (\hat{f}_i - f p_i(x, y) - b)^2$$

Normalization **Residual**

Starting Point

- Normal log-likelihood:

$$\ln \mathcal{L}(x, y, f, b) = -\frac{nm}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i(\hat{f}_i - f p_i(x, y) - b)^2$$

Normalization **Residual**

- **Flux** and error:

$$f_{\text{ML}}(x, y, b) = \frac{\sum_i(\hat{f}_i - b)p_i(x, y)}{\sum_i p_i^2(x, y)}$$

“Naïve” error

$$\tilde{\sigma}_f^2(x, y) \equiv -\left(\partial_f^2 \ln \mathcal{L}\right)^{-1} = \frac{\sigma^2}{\sum_i p_i^2(x, y)} \equiv \boxed{A_{\text{psf}}(x, y) \times \sigma^2}$$

$4\pi s^2$ for Gaussian

Starting Point

- Normal log-likelihood:

$$\ln \mathcal{L}(x, y, f, b) = -\frac{nm}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (\hat{f}_i - f p_i(x, y) - b)^2$$

Normalization **Residual**

- **Position** error:

PSF "Smoothness" **Residual**

"Naïve" error

$$\tilde{\sigma}_x^2(x_{\rm ML}, y, f, b) = \frac{\sigma^2}{f^2}\left(\sum_i (\partial_x p_i(x_{\rm ML}, y))^2 - \frac{1}{f}\left((\hat{f}_i - b) - f p_i(x_{\rm ML}, y)\right)\partial_x^2 p_i(x_{\rm ML}, y)\right)^{-1}$$

SNR **PSF "Variability"**

Starting Point

- Normal log-likelihood:

$$\ln \mathcal{L}(x, y, f, b) = -\frac{nm}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (\hat{f}_i - f p_i(x, y) - b)^2$$

Normalization **Residual**

- **Position** error:

“Naïve” error

$$\tilde{\sigma}_x^2(x_{\mathrm{ML}}, y, f, b) \approx \frac{1}{f^2} \frac{\sigma^2}{\sum_i \left(\partial_x p_i(x_{\mathrm{ML}}, y)\right)^2} \equiv \boxed{S_{\mathrm{psf}}(x_{\mathrm{ML}}, y) \times \left(\frac{f^2}{\sigma^2}\right)^{-1}}$$

$8\pi s^4 = 2s^2 A_{\mathrm{psf}}$ for Gaussian

Starting Point

- Normal log-likelihood:

$$\ln \mathcal{L}(x, y, f, b) = -\frac{nm}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_i (\hat{f}_i - f p_i(x, y) - b)^2$$

Normalization **Residual**

- **Background** and error:

$$b_{\mathrm{ML}}(x, y, f) = \frac{1}{nm} \sum_i \hat{f}_i - f p_i(x, y)$$

$$\boxed{\tilde{\sigma}_b^2 = \frac{\sigma^2}{nm} \equiv \frac{\sigma^2}{A}}$$

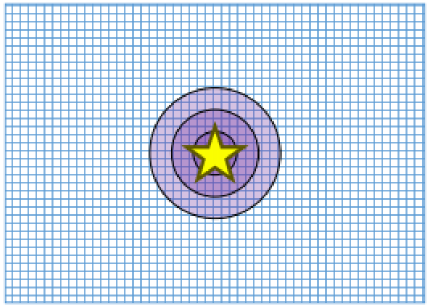
“Naïve” error

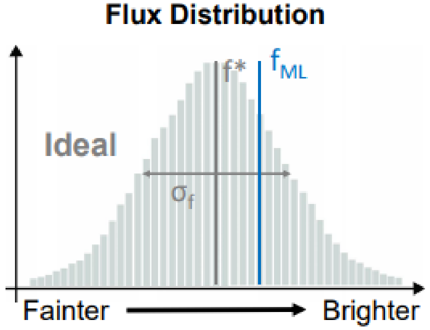
Biases in PSF Photometry

- Assume background known

Position Known

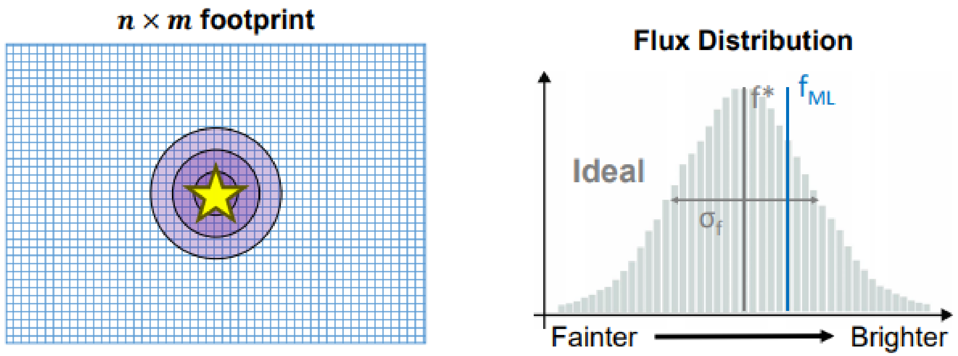
$n \times m$ footprint



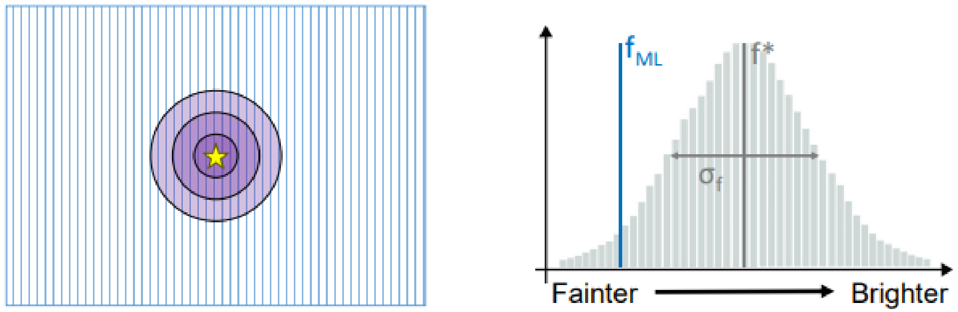


Random noise sometimes leads to high fluctuations.

Position **Known**

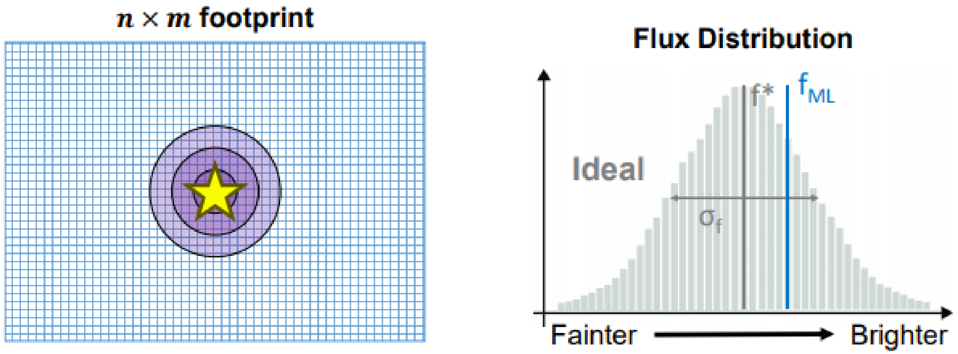


Random noise sometimes leads to high fluctuations.

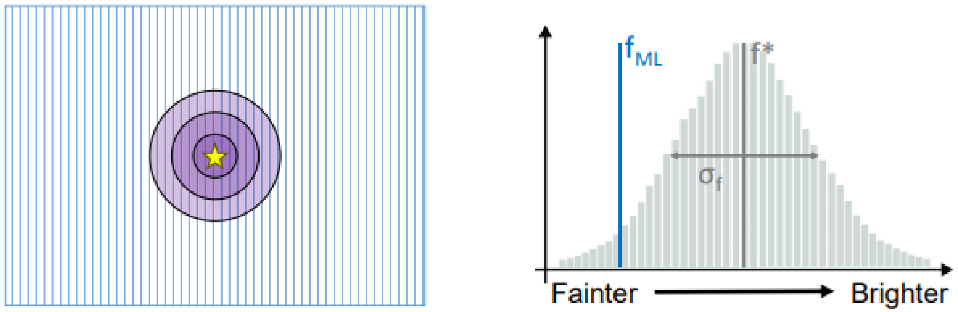


It also leads to low fluctuations with equal probability.

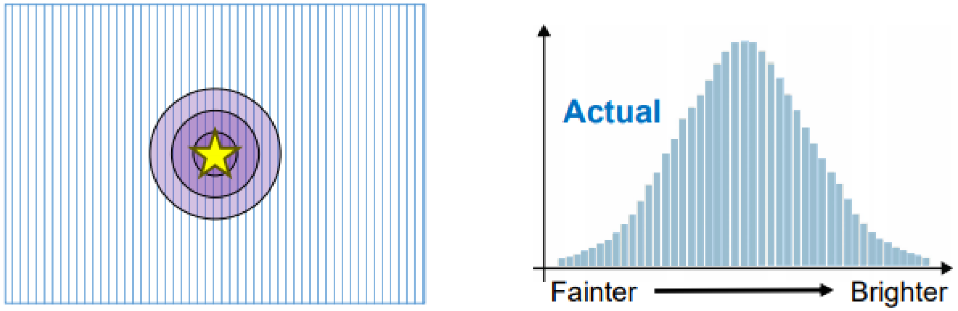
Position Known



Random noise sometimes leads to high fluctuations.

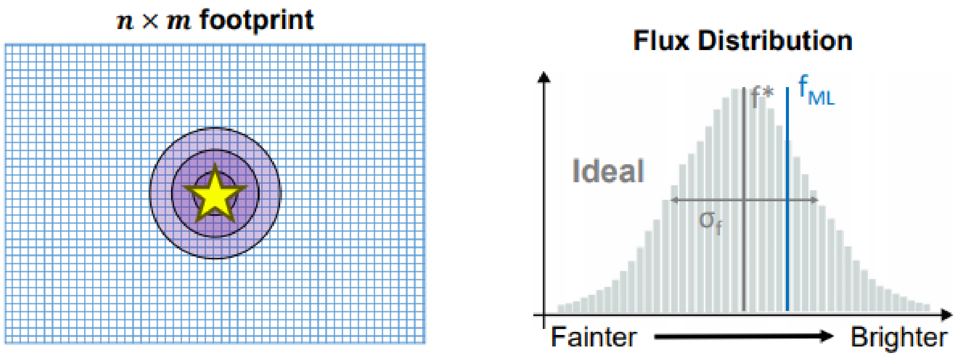


It also leads to low fluctuations with equal probability.

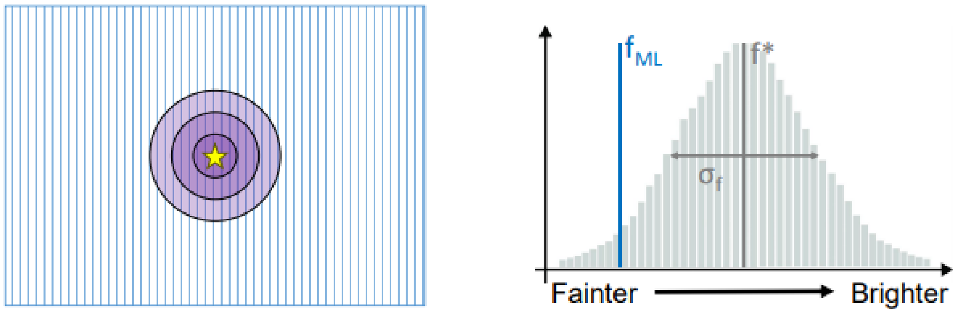


Because fluctuations are symmetric,
the maximum-likelihood estimate is unbiased.

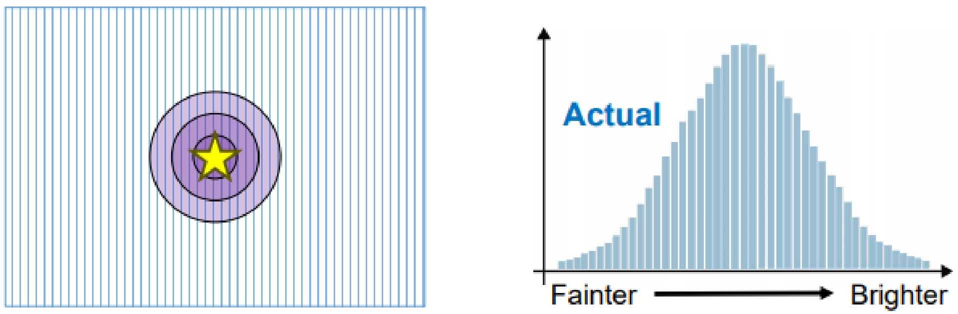
Position Known



Random noise sometimes leads to high fluctuations.

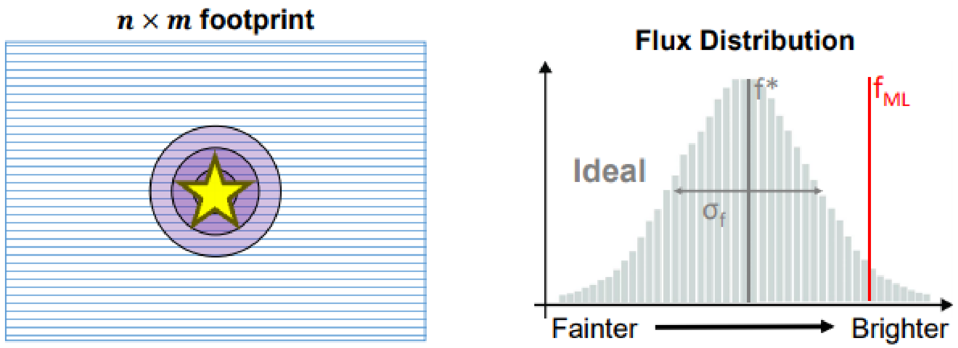


It also leads to low fluctuations with equal probability.



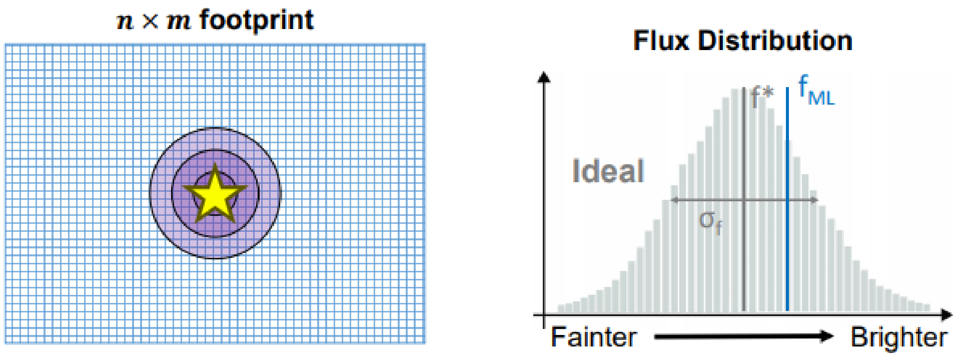
Because fluctuations are symmetric, the maximum-likelihood estimate is unbiased.

Position Unknown

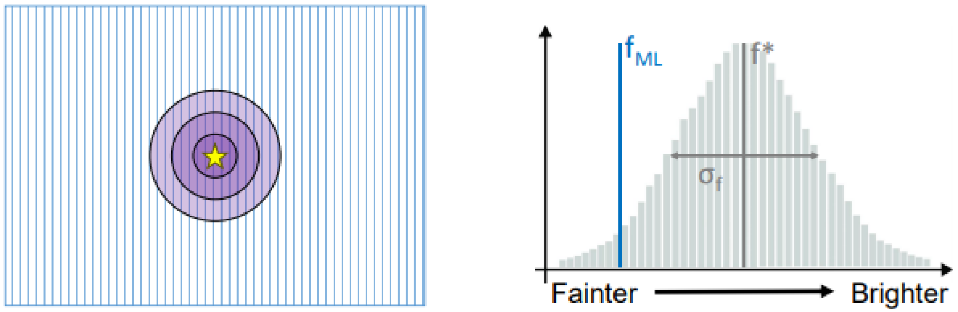


The position is centered at the true position when random noise generates high fluctuations…

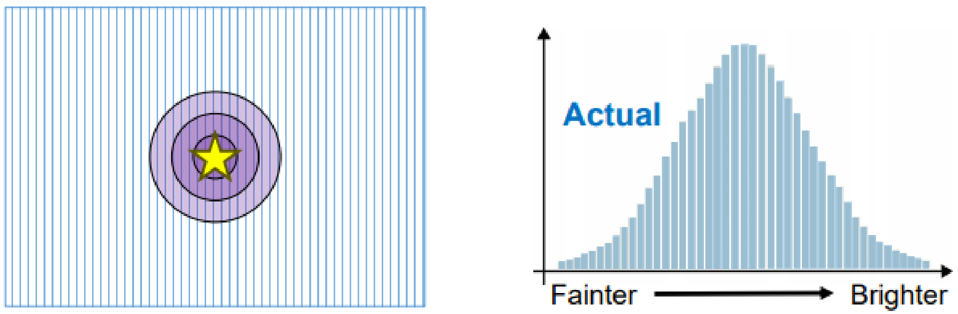
Position Known



Random noise sometimes leads to high fluctuations.

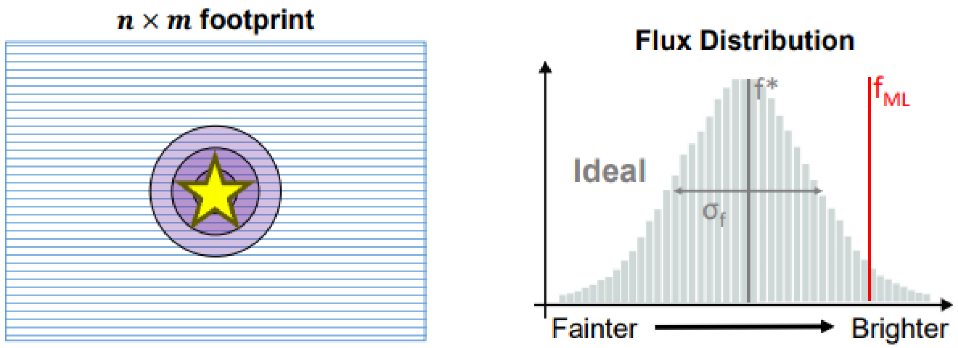


It also leads to low fluctuations with equal probability.

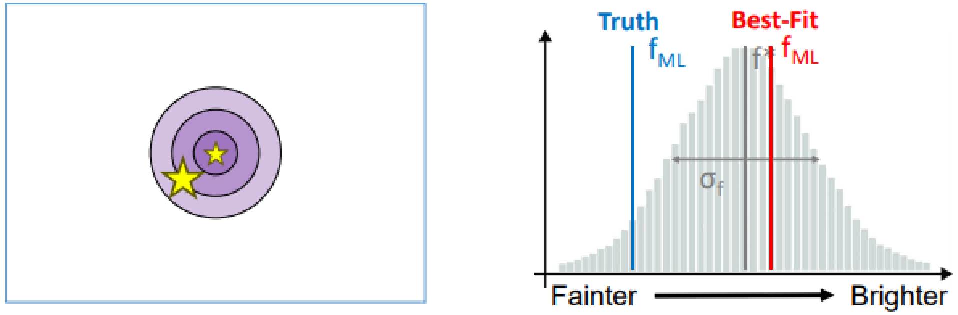


Because fluctuations are symmetric,
the maximum-likelihood estimate is unbiased.

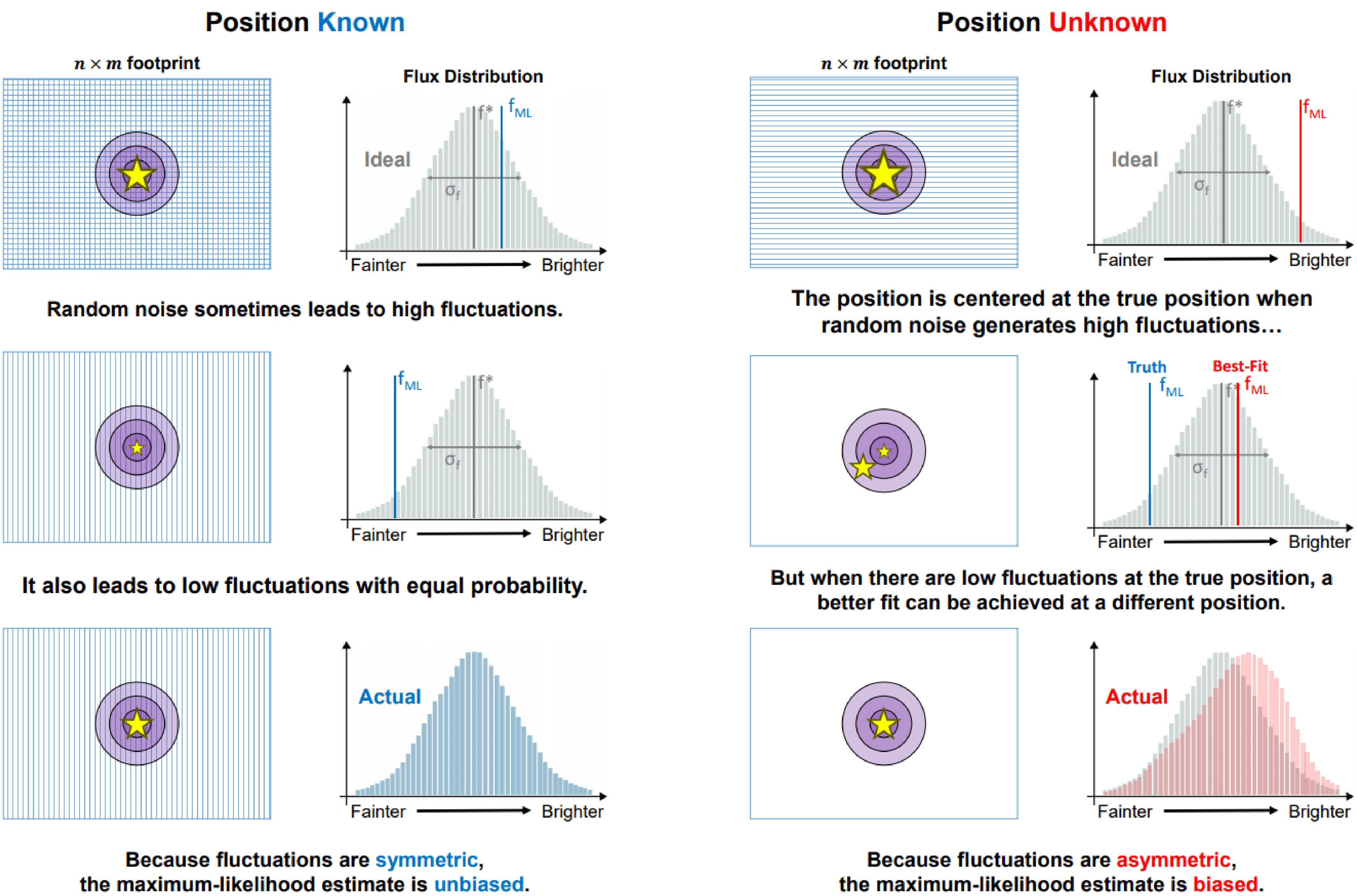
Position Unknown



The position is centered at the true position when random noise generates high fluctuations…



But when there are low fluctuations at the true position, a better fit can be achieved at a different position.



Position Known
n × m footprint
Flux Distribution
Ideal
f*
f_ML
σ_f
Fainter
Brighter
Random noise sometimes leads to high fluctuations.
It also leads to low fluctuations with equal probability.
Actual
Because fluctuations are symmetric,
the maximum-likelihood estimate is unbiased.
Position Unknown
The position is centered at the true position when
random noise generates high fluctuations…
Truth
Best-Fit
But when there are low fluctuations at the true position, a
better fit can be achieved at a different position.
Because fluctuations are asymmetric,
the maximum-likelihood estimate is biased.

Biases in PSF Photometry

- While the MLE is *consistent* (unbiased as N → infinity), not necessarily unbiased.

- Recast the problem with **random variables**:

$$\hat{\mathbf{f}} = f^*\mathbf{p}(x^*, y^*) + \mathbf{b}^* + \mathbf{C}^{1/2}\mathbf{Z}$$

Standard Normal Random Numbers

Biases in PSF Photometry: Ideal Case

- Assume we know position (x*, y*) and background b*.

$$\ln \mathcal{L}(x^*, y^*, f, b^*) = -\frac{nm}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i ((f^* - f)p_i(x^*, y^*) + \sigma Z_i)^2$$

$$f_{\rm ML}(x^*, y^*, b^*) = f^* + \frac{\sum_i Z_i p_i(x^*, y^*)/\sigma}{\sum_i p_i^2(x^*, y^*)/\sigma^2}$$

Biases in PSF Photometry: Ideal Case

- Assume we know position (x*, y*) and background b*.

$$\ln \mathcal{L}(x^*, y^*, f, b^*) = -\frac{nm}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i ((f^* - f)p_i(x^*, y^*) + \sigma Z_i)^2$$

$$f_{\rm ML}(x^*, y^*, b^*) = f^* + \frac{\sum_i Z_i p_i(x^*, y^*)/\sigma}{\sum_i p_i^2(x^*, y^*)/\sigma^2}$$

$$\boxed{f_{\rm ML}(x^*, y^*, b^*) \sim \mathcal{N}(f^*, \sigma_f^2(x^*, y^*))}$$

Biases in PSF Photometry: General Case

- In general, if we knew the truth and simulated the data, we'd end up with:

$$\ln \mathcal{L}(x^*, y^*, f^*, \mathbf{b}^*) = -\frac{1}{2}\ln(\det(2\pi\mathbf{C})) - \frac{1}{2}\sum_i Z_i^2$$

$$\boxed{\sum_{i=1}^{nm} Z_i^2 \sim \chi^2_{nm}}$$

Biases in PSF Photometry: General Case

- When you introduce parameters, they "absorb" some of the noise:

$$(\hat{\mathbf{f}} - \mathbf{f}_{\boldsymbol{\theta}_{\mathrm{ML}}})^{\mathrm{T}} \mathbf{C}^{-1} (\hat{\mathbf{f}} - \mathbf{f}_{\boldsymbol{\theta}_{\mathrm{ML}}}) \sim \chi^2_{nm-p}$$

- The variation in the parameters contains the missing noise:

$$(\boldsymbol{\theta}^* - \boldsymbol{\theta}_{\mathrm{ML}})^{\mathrm{T}} \mathbf{C}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\theta}_{\mathrm{ML}}) \sim \chi^2_p$$

Biases in PSF Photometry: w/o Background

- We can exploit this to relate our two estimators in distribution:

Biases in PSF Photometry: w/o Background

- We can exploit this to relate our two estimators in distribution:

MLE residual with position fixed

$$(\hat{\mathrm{f}} - f^*_{\mathrm{ML}} \mathrm{P}_{x^*,y^*})^{\mathrm{T}} \mathrm{C}^{-1} (\hat{\mathrm{f}} - f^*_{\mathrm{ML}} \mathrm{P}_{x^*,y^*})$$

Biases in PSF Photometry: w/o Background

- We can exploit this to relate our two estimators in distribution:

Equal in distribution

MLE residual with position free **MLE residual with position fixed**

$$(\hat{\mathbf{f}} - f_{\mathrm{ML}}\mathbf{p}_{x_{\mathrm{ML}},y_{\mathrm{ML}}})^{\mathrm{T}}\mathbf{C}^{-1}(\hat{\mathbf{f}} - f_{\mathrm{ML}}\mathbf{p}_{x_{\mathrm{ML}},y_{\mathrm{ML}}}) + X_2^2 \sim (\hat{\mathbf{f}} - f^*_{\mathrm{ML}}\mathbf{p}_{x^*,y^*})^{\mathrm{T}}\mathbf{C}^{-1}(\hat{\mathbf{f}} - f^*_{\mathrm{ML}}\mathbf{p}_{x^*,y^*})$$

$\chi^2(\mathrm{dof} = 2)$
random variable

Biases in PSF Photometry: w/o Background

- We can use this to get a first-order bias correction:

$$\mathbb{E}[f^*_{\mathrm{ML}}] \approx f_{\mathrm{ML}} \left[1 - \frac{\tilde{\sigma}^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}} \right]$$

Biases in PSF Photometry: w/o Background

- We can use this to get a first-order bias correction:

$$\mathbb{E}[f^*_{\mathrm{ML}}] \approx f_{\mathrm{ML}} \left[1 - \frac{\tilde{\sigma}^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}}\right]$$

- Applying this correction increases the variance:

$$\tilde{\sigma}_{f^*_{\mathrm{ML}}} = \sqrt{\tilde{\sigma}^2_{f_{\mathrm{ML}}} + \mathbb{V}[f^*_{\mathrm{ML}}]} \approx \tilde{\sigma}_{f_{\mathrm{ML}}} \left(1 + \frac{1}{2}\frac{\mathbb{V}[f^*_{\mathrm{ML}}]}{\tilde{\sigma}^2_{f_{\mathrm{ML}}}}\right) = \tilde{\sigma}_{f_{\mathrm{ML}}} \left(1 + \frac{1}{2}\frac{\tilde{\sigma}^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}}\right)$$

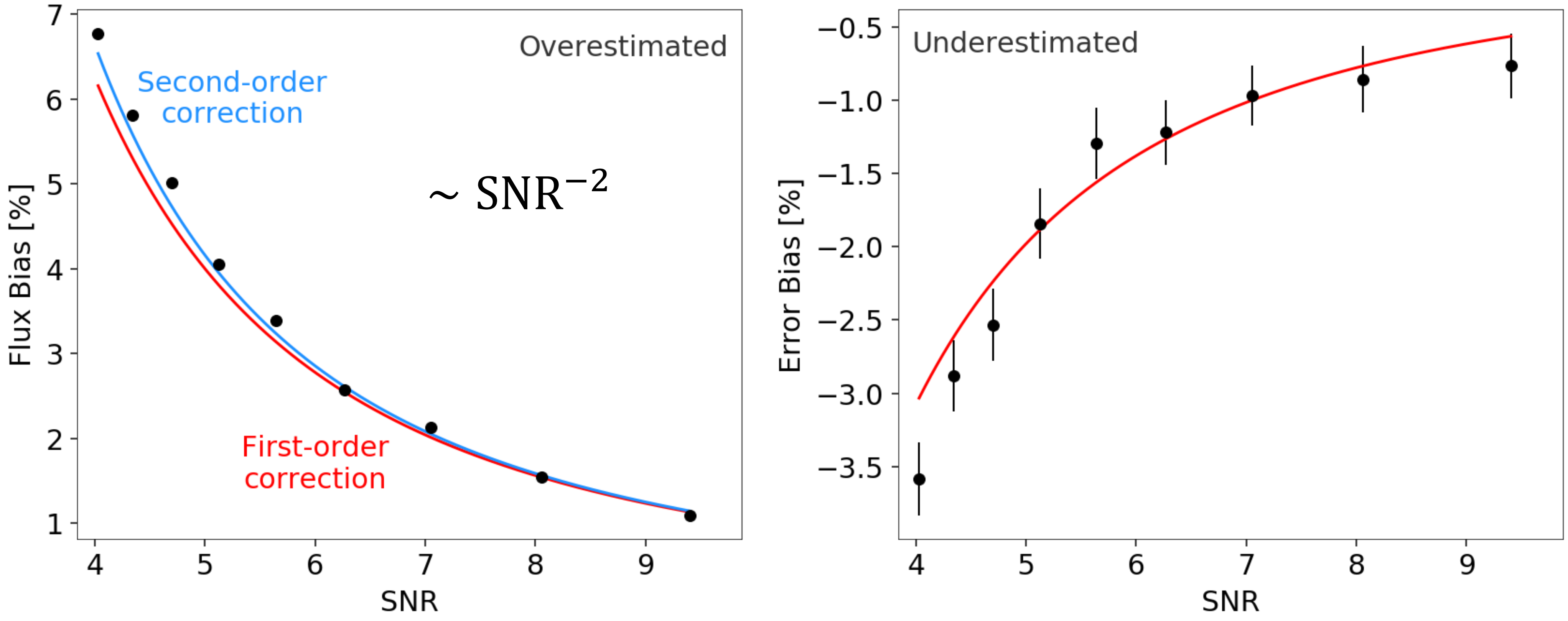
- This is an example of the **bias-variance trade-off**.

10-sigma source:
- Flux: +1% bias.

Biases in PSF Photometry

- Position **unknown**, background known

Background fixed

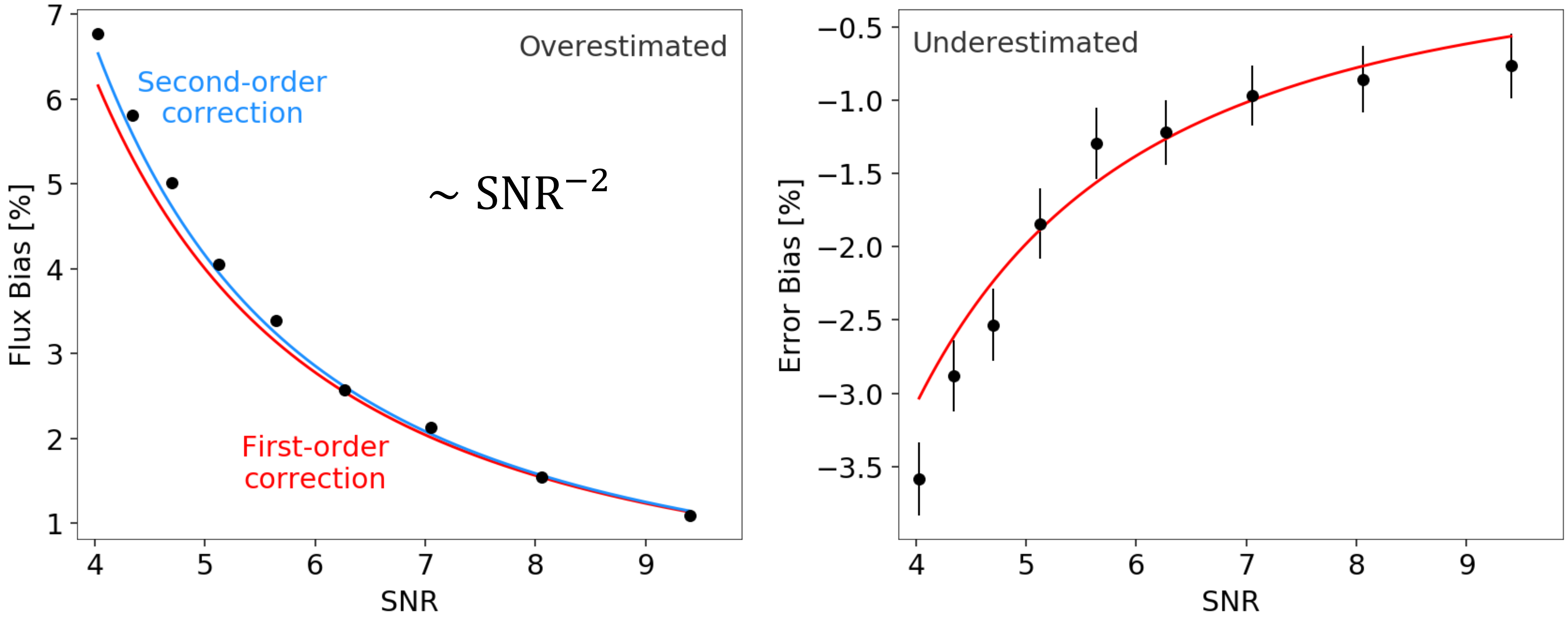


10-sigma source:
- Flux: +1% bias.

Biases in PSF Photometry

- Position **unknown**, background **unknown**

Background fixed

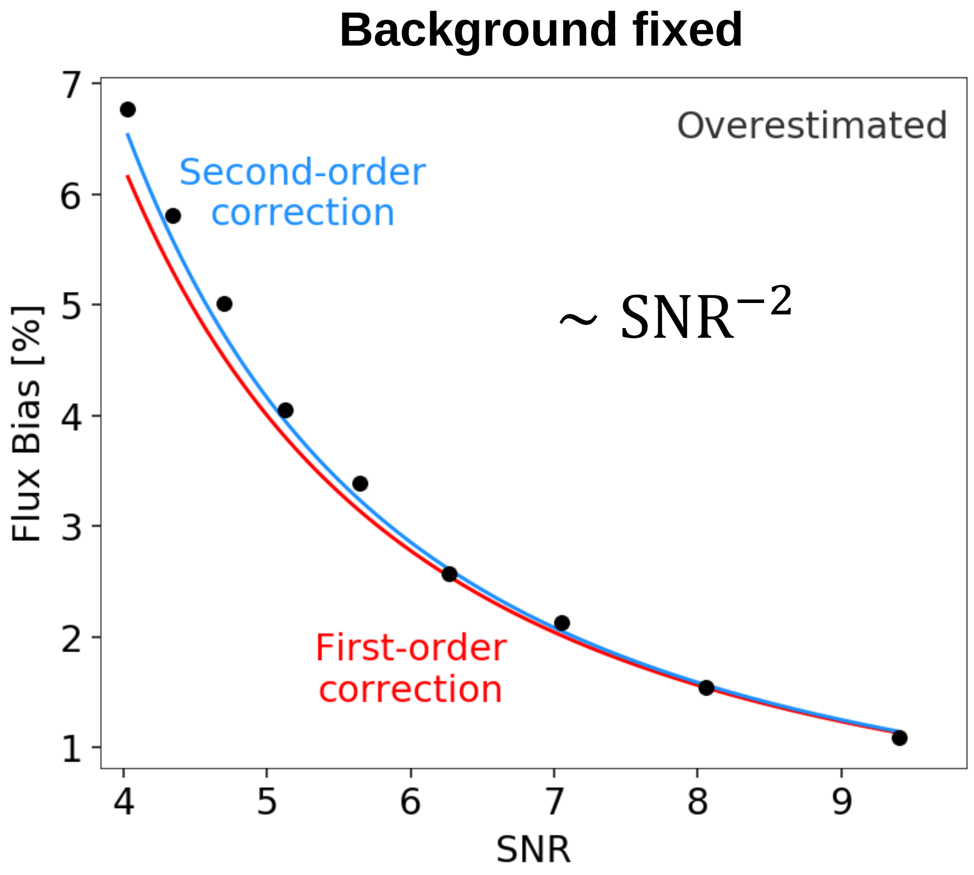


10-sigma source:

- Flux: +1% bias.

Biases in PSF Photometry

- Position **unknown**, background **unknown**



- Need to account for covariance between background and other parameters.

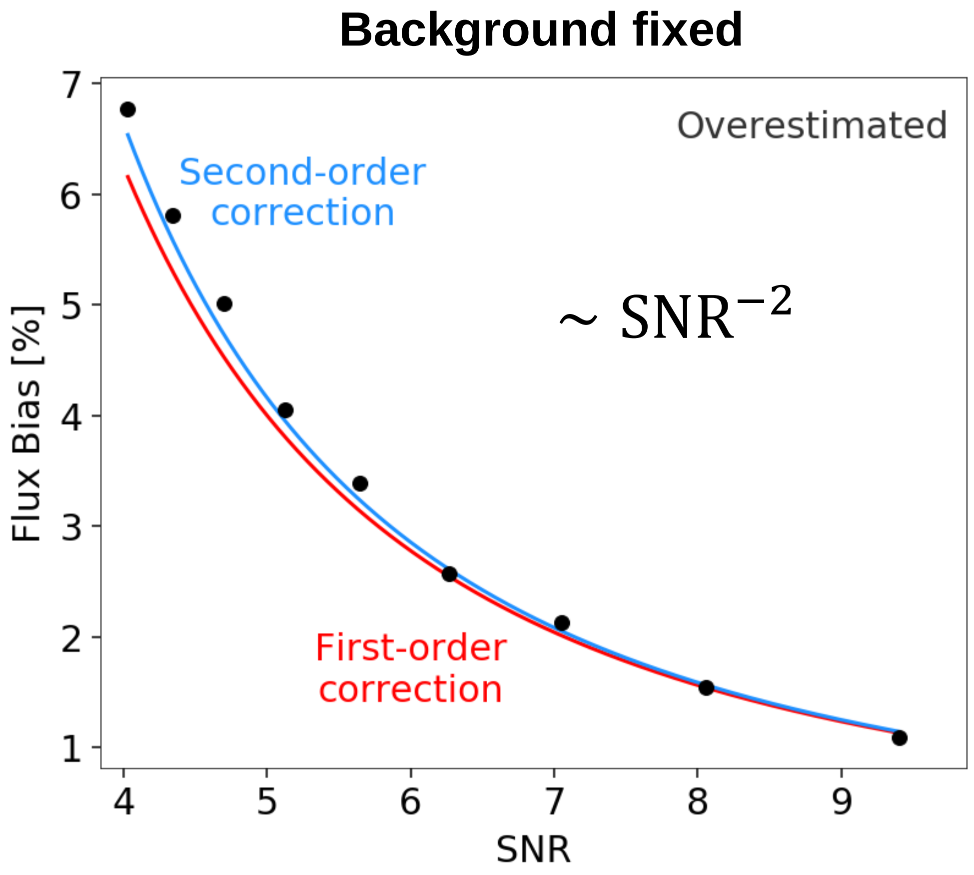
$$\mathrm{C}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{\mathrm{ML}}) \approx -\mathbb{E}_{\mathbf{D}}\left[\partial^2_{\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta}_{\mathrm{ML}})|\boldsymbol{\theta}_{\mathrm{ML}}\right]^{-1}$$

$$\equiv \left(\boldsymbol{\mathcal{F}}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_{\mathrm{ML}})\right)^{-1} = -\left(\partial^2_{\boldsymbol{\theta}} \ln \mathcal{L}(\boldsymbol{\theta}_{\mathrm{ML}})\right)^{-1}$$

10-sigma source:
- Flux: +1% bias.

Biases in PSF Photometry

- Position **unknown**, background **unknown**



- Need to account for covariance between background and other parameters.

$$\partial_f \partial_b \ln \mathcal{L}(x,y) = -\frac{1}{\sigma^2}\sum_i p_i(x,y) = -\frac{1}{\sigma^2}$$

$$\partial_x \partial_b \ln \mathcal{L}(x,y) = -\frac{f}{\sigma^2}\sum_i \partial_x p_i(x,y) \approx 0$$

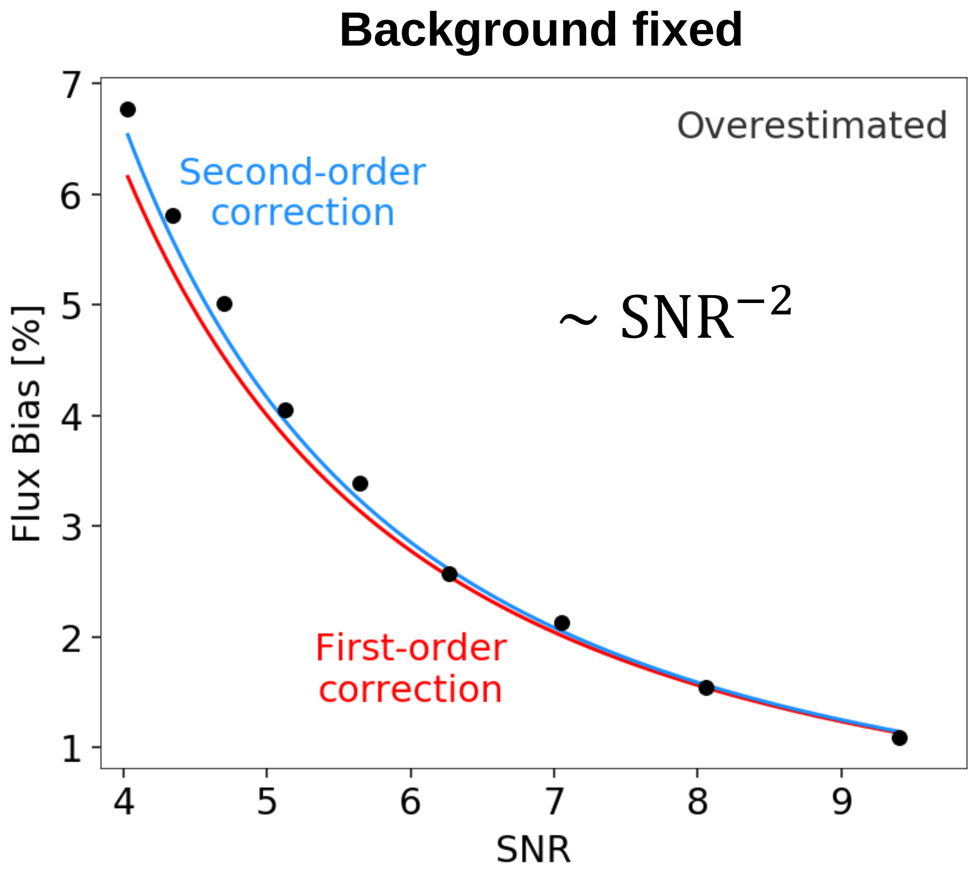
Oversampled limit: can approximate sum with integral

10-sigma source:

- Flux: +1% bias.

Biases in PSF Photometry

- Position **unknown**, background **unknown**



- Need to account for covariance between background and other parameters.

$$\partial_f \partial_b \ln \mathcal{L}(x,y) = -\frac{1}{\sigma^2} \sum_i p_i(x,y) = -\frac{1}{\sigma^2}$$

$$\partial_x \partial_b \ln \mathcal{L}(x,y) = -\frac{f}{\sigma^2} \sum_i \partial_x p_i(x,y) \approx 0$$

$$\sigma_f^2(x,y) = \boxed{\frac{A}{A - A_{\rm psf}(x,y)}} \times \tilde{\sigma}_f^2$$

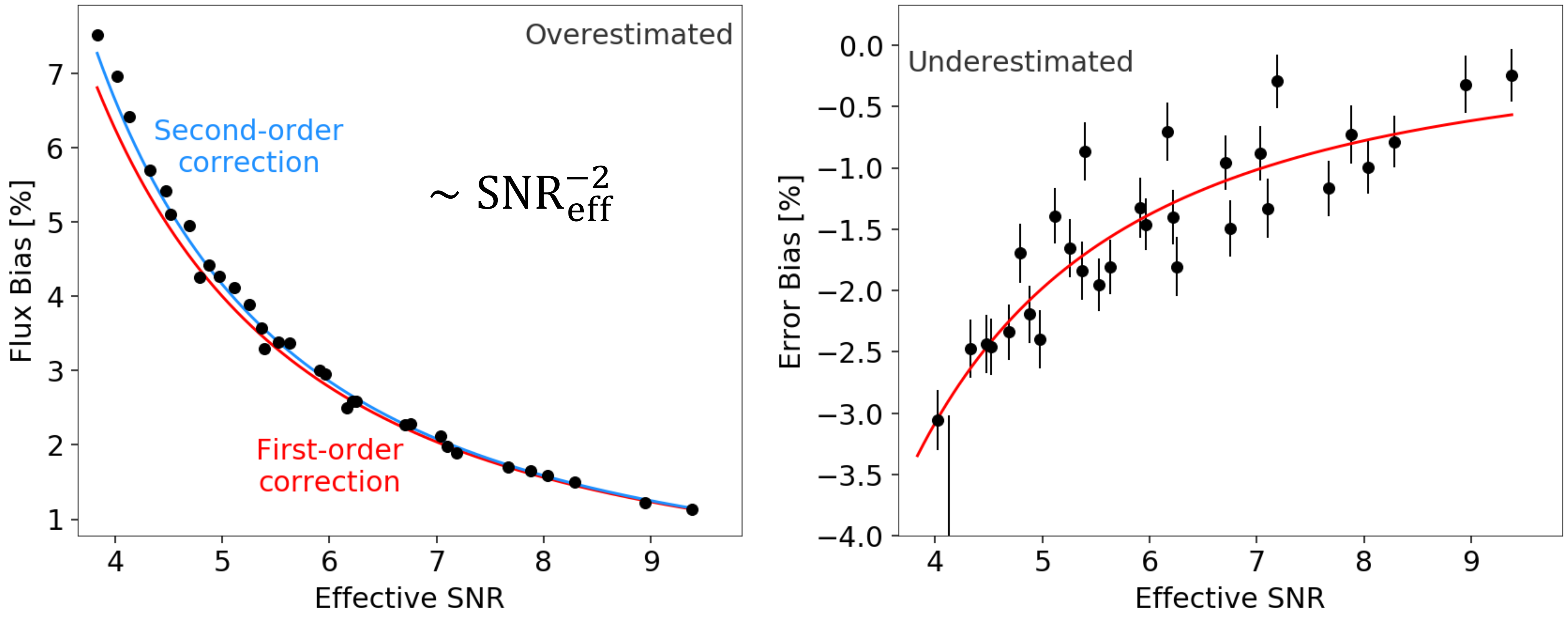
10-sigma source:

- **Flux: +1% bias.**
- **Variance: -0.1% bias.**

Biases in PSF Photometry

- Position **unknown**, background **unknown**

Background variable



Biases in PSF Photometry

- Position **unknown**, background **unknown**

Flux Bias

$$\frac{\delta_{f_{\mathrm{ML}}}}{f_{\mathrm{ML}}} \approx \frac{\sigma^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}}$$

Extra degrees of freedom allows fit to chase noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x, y) = -\frac{A_{\mathrm{psf}}(x, y)}{A}$$

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry

Flux Bias

$$\frac{\delta_{f_{\mathrm{ML}}}}{f_{\mathrm{ML}}} \approx \frac{\sigma^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}}$$

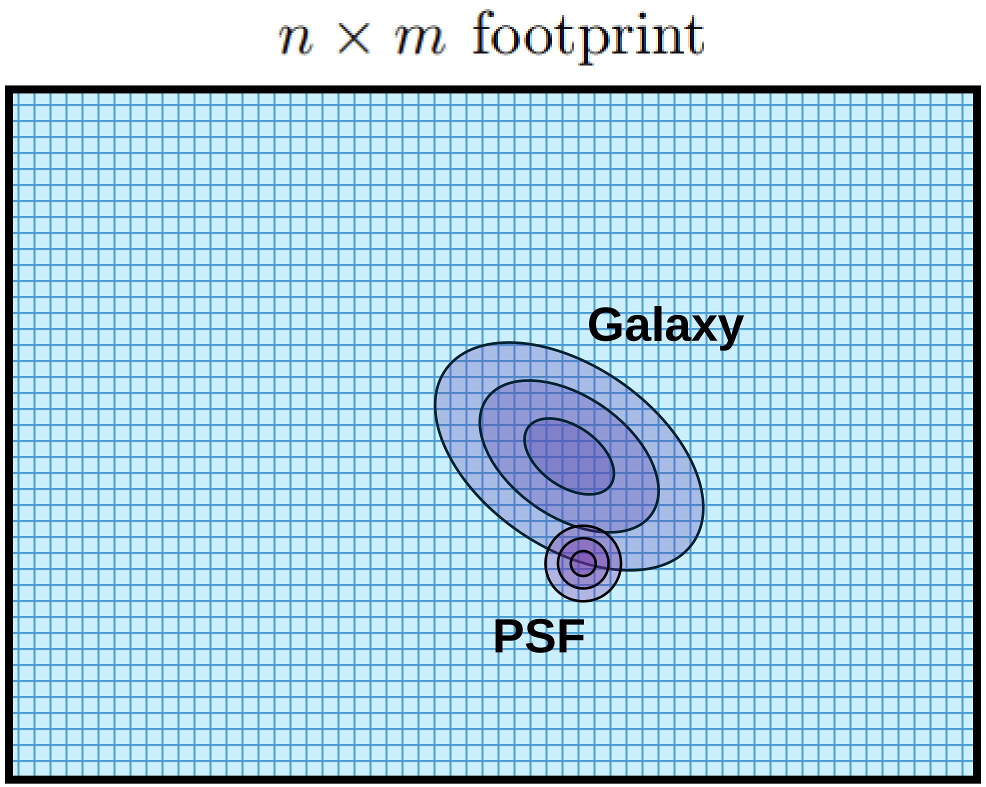
Extra degrees of freedom allows fit to chase noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x,y) = -\frac{A_{\mathrm{psf}}(x,y)}{A}$$

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry



Biases in Extended Source Photometry

Flux Bias

$$\frac{\delta_{f_{\mathrm{ML}}}}{f_{\mathrm{ML}}} \approx \frac{\sigma^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}}$$

Extra degrees of freedom allows fit to chase noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x, y) = -\frac{A_{\mathrm{psf}}(x, y)}{A}$$

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry

- More parameters

Flux Bias

$$\frac{\delta_{f_{\rm ML}}}{f_{\rm ML}} \approx \frac{\sigma^2_{f_{\rm ML}}}{f^2_{\rm ML}}$$

Extra degrees of freedom allows fit to chase noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x,y) = -\frac{A_{\rm psf}(x,y)}{A}$$

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry

- More parameters

Flux Bias

$$\frac{\delta_{f_{\mathrm{ML}}}}{f_{\mathrm{ML}}} \approx \boxed{\frac{p-1}{2}} \frac{\sigma^2_{f_{\mathrm{ML}}}}{f^2_{\mathrm{ML}}}$$

Extra degrees of freedom allows fit to chase noise.

$\boldsymbol{p = 3 \rightarrow 6 - 10}$

Shape parameters soak up noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x, y) = -\frac{A_{\mathrm{psf}}(x, y)}{A}$$

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry

- More parameters, more covariances

Flux Bias

$$\frac{\delta_{f_{\rm ML}}}{f_{\rm ML}} \approx \boxed{\frac{p-1}{2}} \frac{\sigma^2_{f_{\rm ML}}}{f^2_{\rm ML}}$$

Extra degrees of freedom allows fit to chase noise.

$\boldsymbol{p = 3 \rightarrow 6 - 10}$

Shape parameters soak up noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x, y) = -\frac{A_{\rm psf}(x, y)}{A}$$

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry

- More parameters, more covariances

Flux Bias

$$\frac{\delta_{f_{\rm ML}}}{f_{\rm ML}} \approx \frac{p-1}{2}\frac{\sigma^2_{f_{\rm ML}}}{f^2_{\rm ML}}$$

Extra degrees of freedom allows fit to chase noise.

$p = 3 \rightarrow 6 - 10$

Shape parameters soak up noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x,y) = -\frac{2A_{\rm psf}(x,y)}{A}$$

Ignoring covariances underestimates errors.

Shape parameters add covariances.

Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area

Flux Bias

$$\frac{\delta_{f_{\rm ML}}}{f_{\rm ML}} \approx \frac{p-1}{2}\frac{\sigma^2_{f_{\rm ML}}}{f^2_{\rm ML}}$$

Extra degrees of freedom allows fit to chase noise.

$p = 3 \rightarrow 6 - 10$

Shape parameters soak up noise.

Error Bias

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x,y) = -\frac{2A_{\rm psf}(x,y)}{A}$$

Ignoring covariances underestimates errors.

Shape parameters add covariances.

Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area

Flux Bias

$$\frac{\delta_{f_{\rm ML}}}{f_{\rm ML}} \approx \frac{p-1}{2}\frac{\sigma^2_{f_{\rm ML}}}{f^2_{\rm ML}}$$

Extra degrees of freedom allows fit to chase noise.

$p = 3 \rightarrow 6 - 10$

Shape parameters soak up noise.

Error Bias

$A_{\rm gal+psf}$

$$\frac{\delta_{\tilde{\sigma}^2_f}}{\sigma^2_f}(x,y) = -\frac{2A_{\rm psf}(x,y)}{A}$$

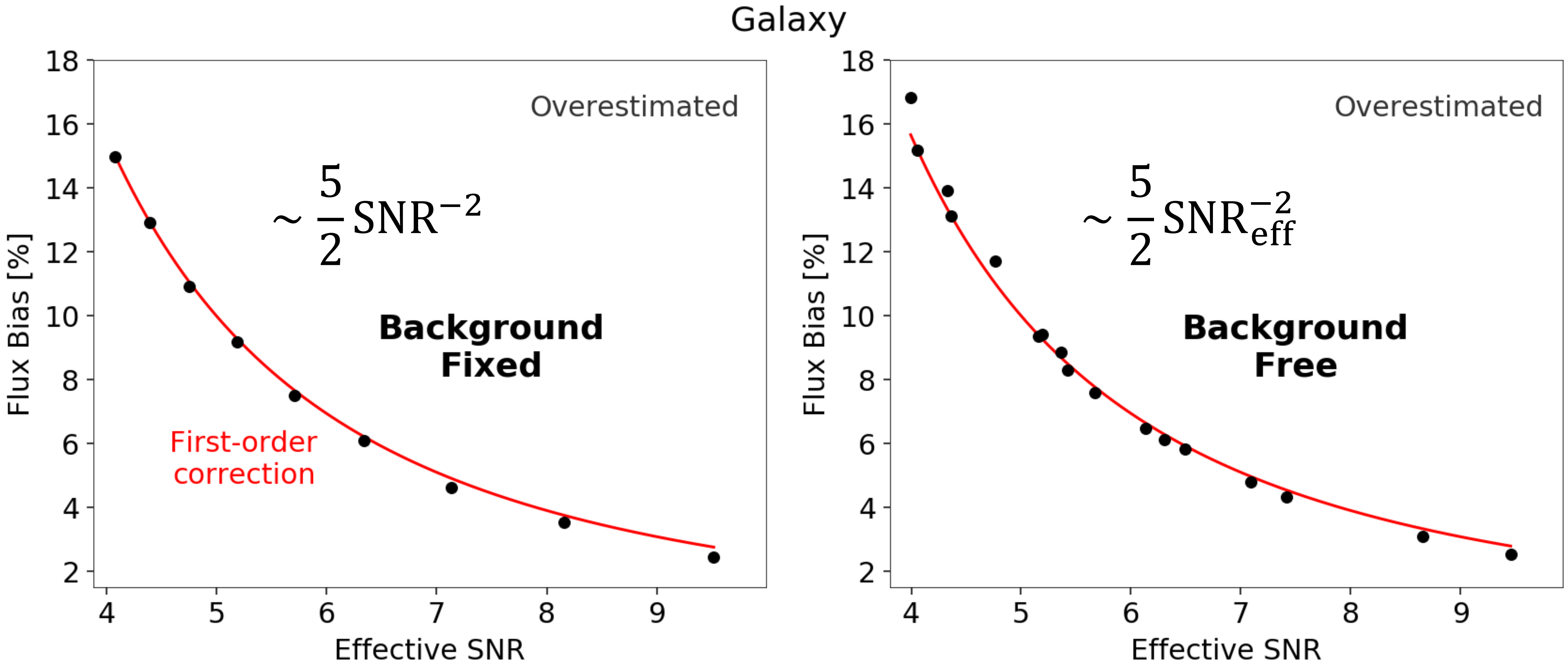
Ignoring covariances underestimates errors.

Shape parameters add covariances.

Extended shape impedes background estimation.

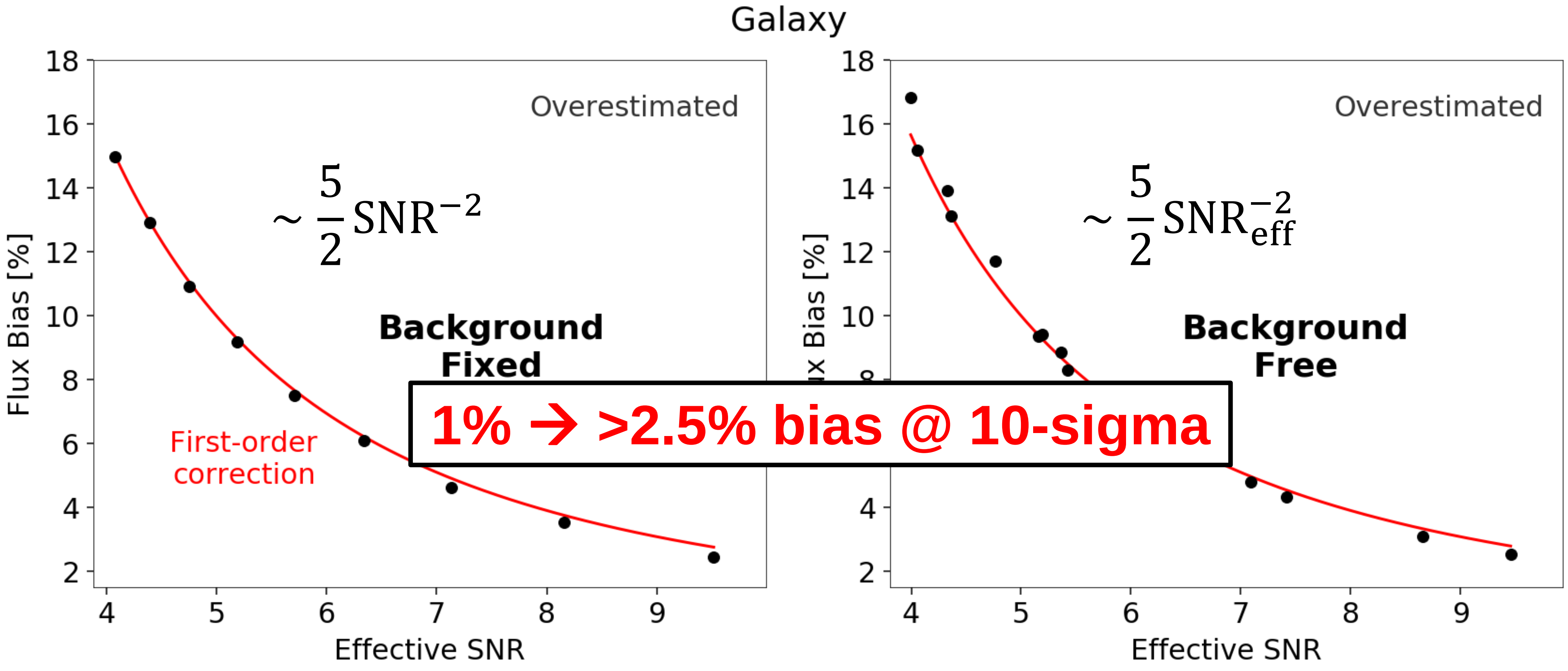
Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area



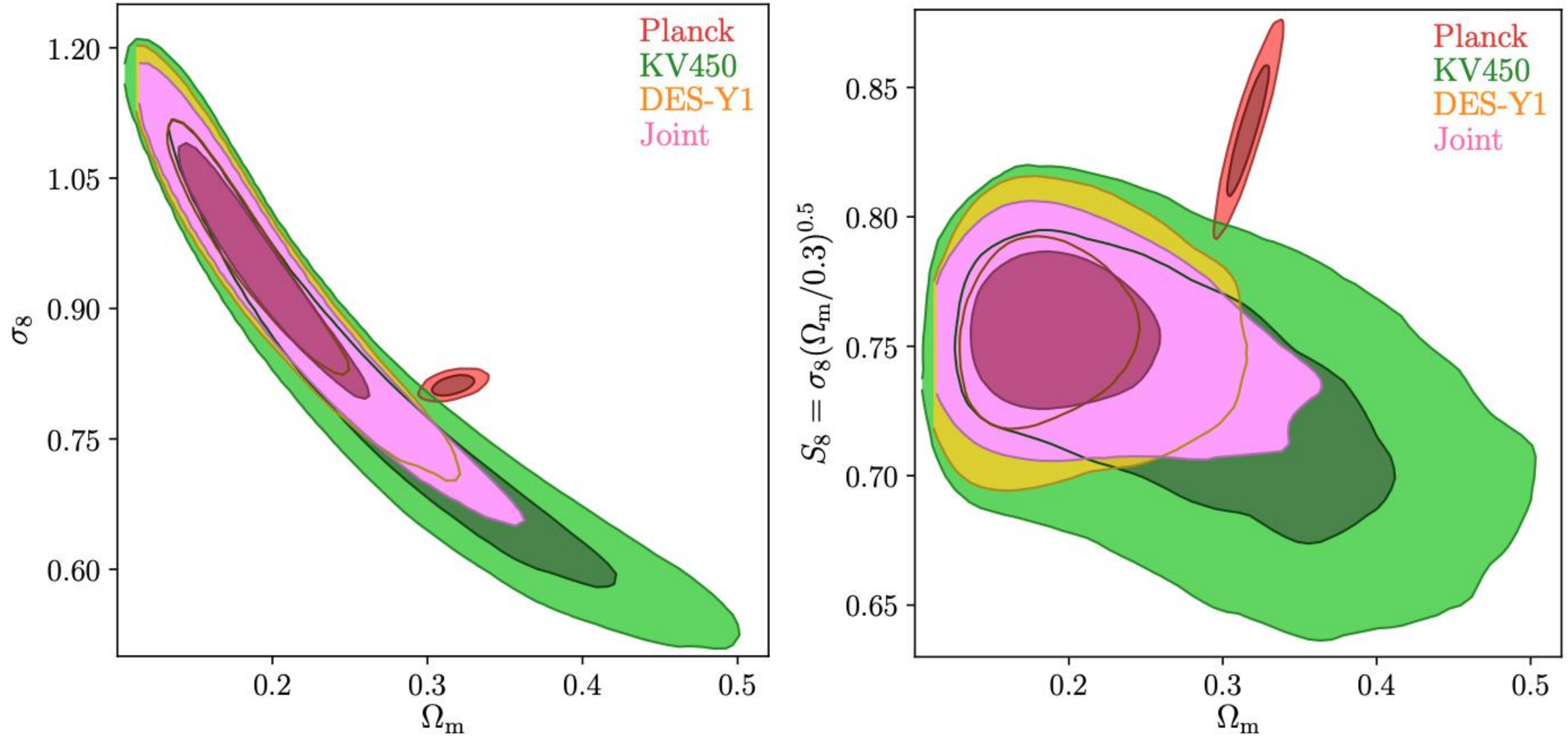
Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area



Impacts

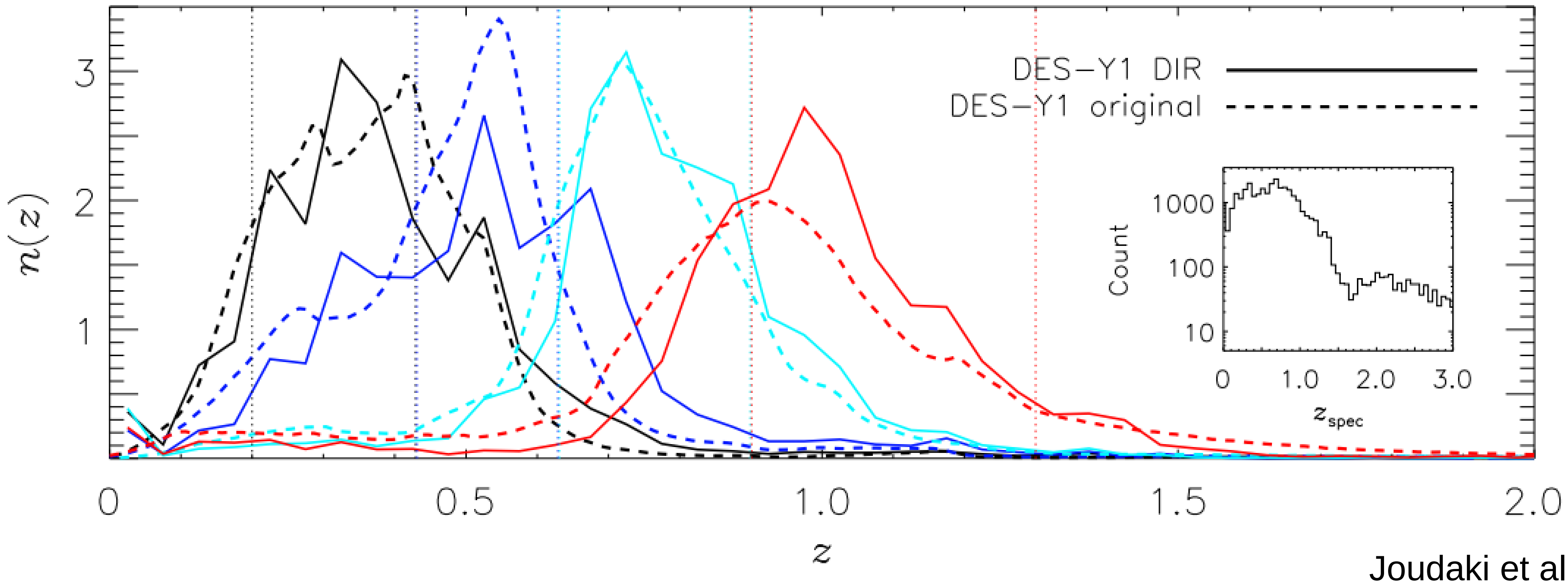
- Emerging tension in large-scale clustering between Cosmic Microwave Background (CMB) and weak lensing.



Asgari et al. (2019)

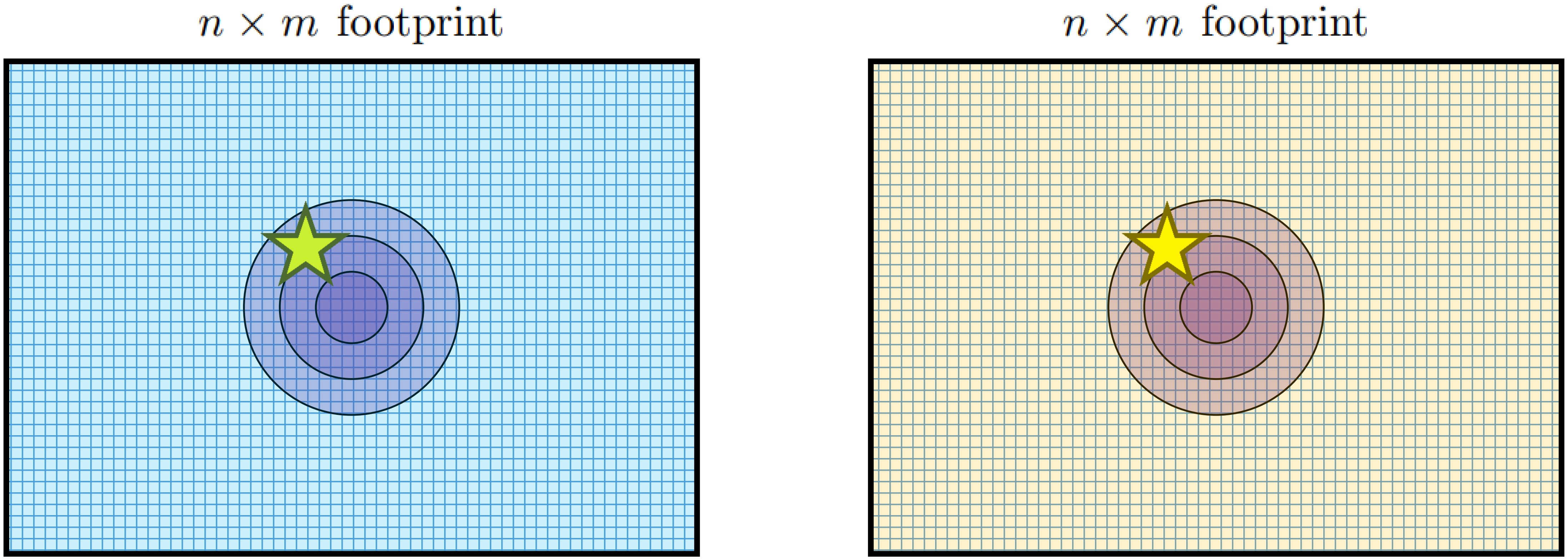
Impacts

- Offsets due to differences in inferred redshift (distance) distribution of galaxies from photometry.
  - Flux overestimated → slightly closer → smaller redshift → population bias.



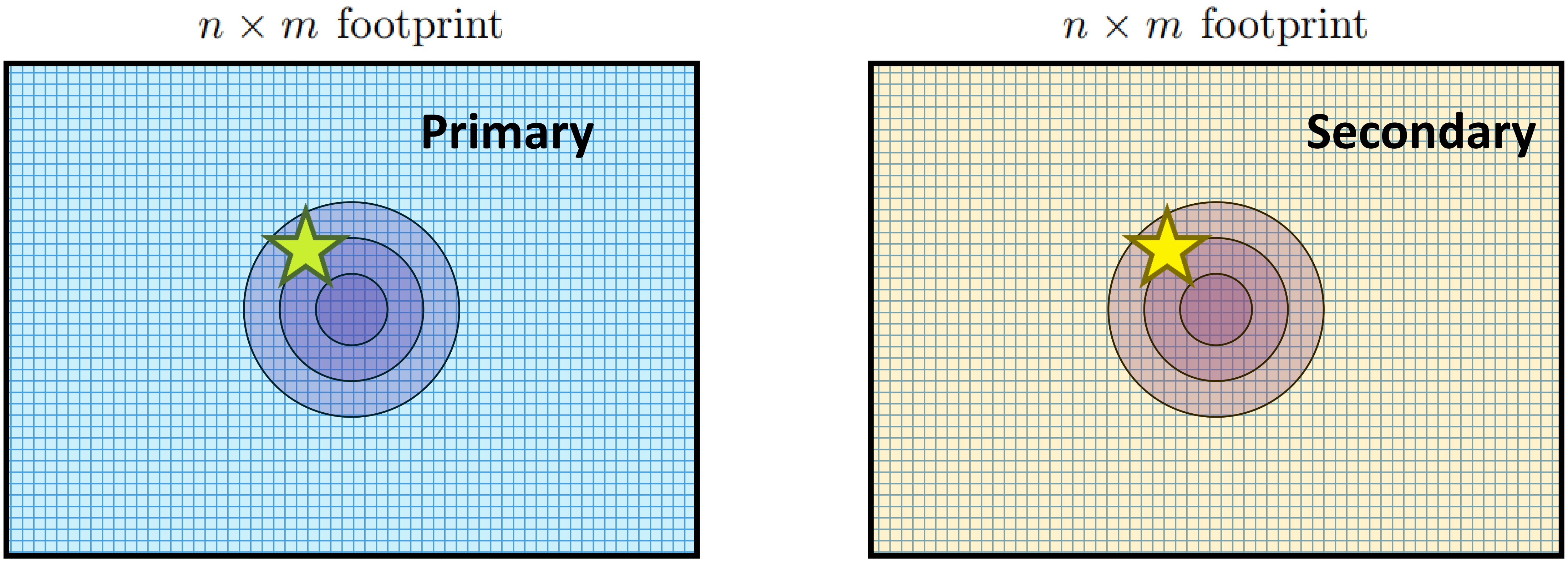
Joudaki et al. (2019)

Biases in Multi-band Photometry



Biases in Forced Multi-band Photometry

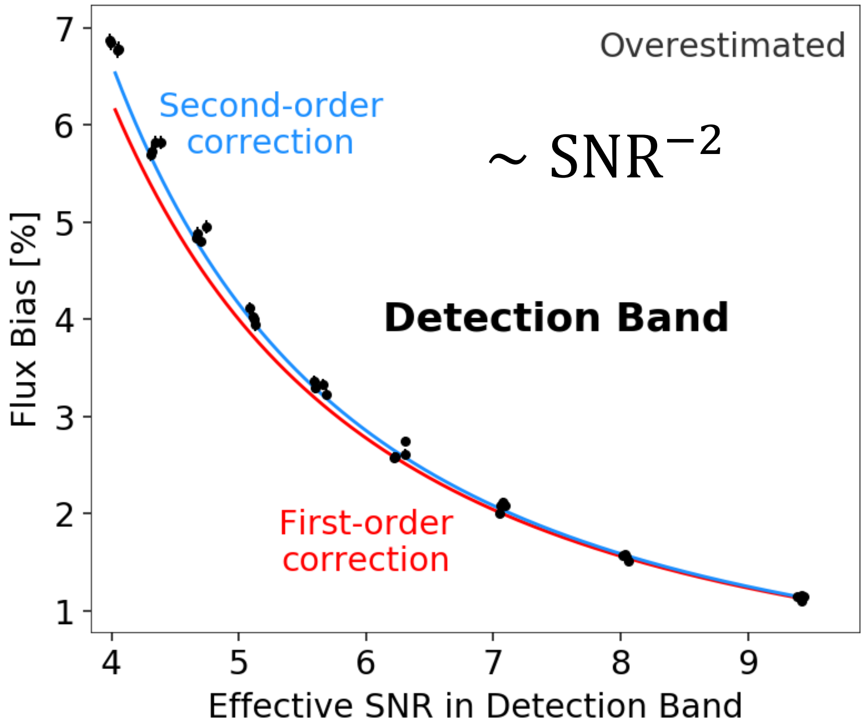
- Detect in one band, fix position, force photometry in others



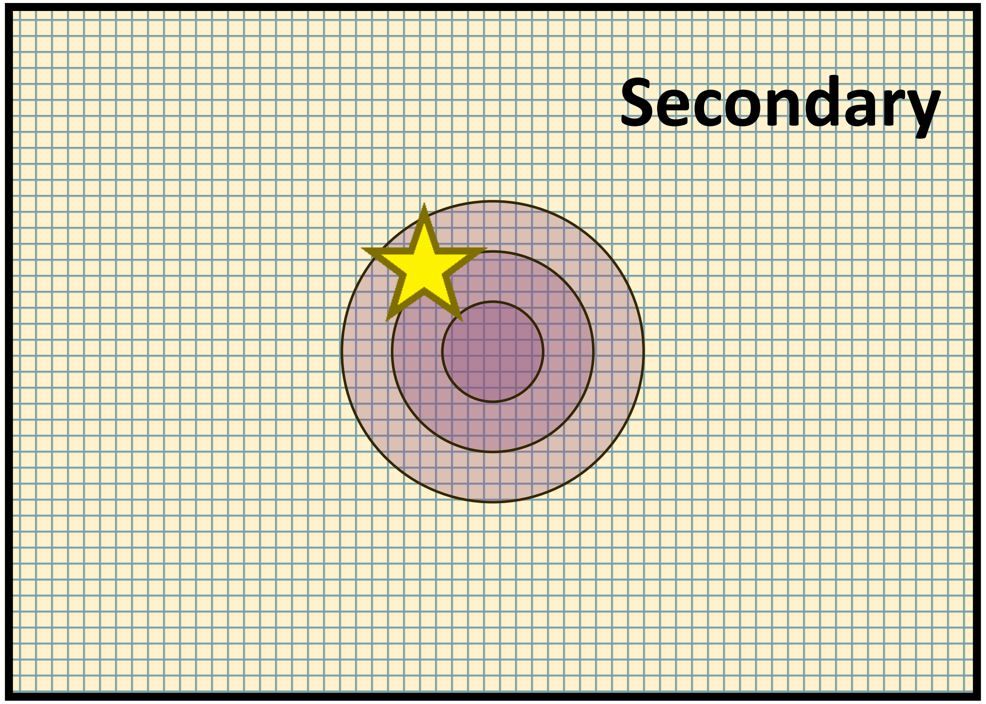
Biases in Forced Multi-band Photometry

- Detect in one band, fix position, force photometry in others

$n \times m$ footprint



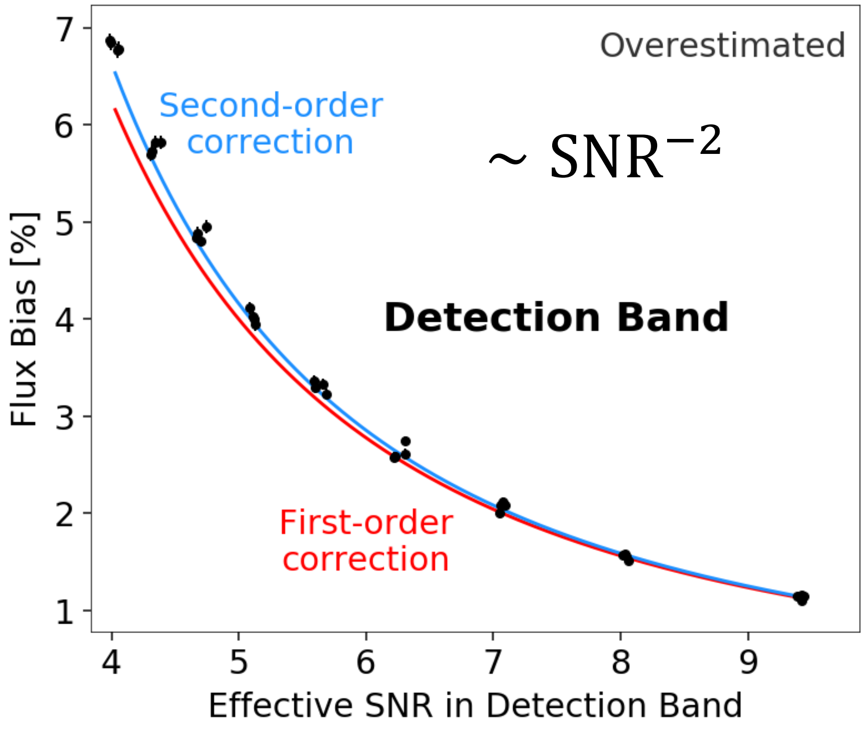
$n \times m$ footprint



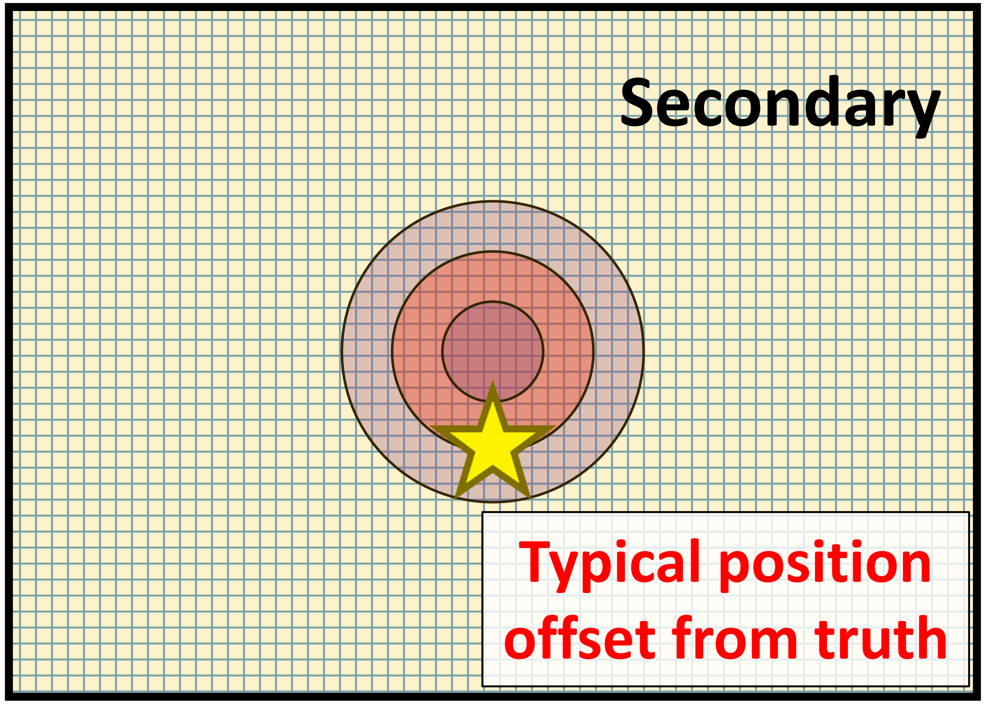
Biases in Forced Multi-band Photometry

- Detect in one band, fix position, force photometry in others

$n \times m$ footprint

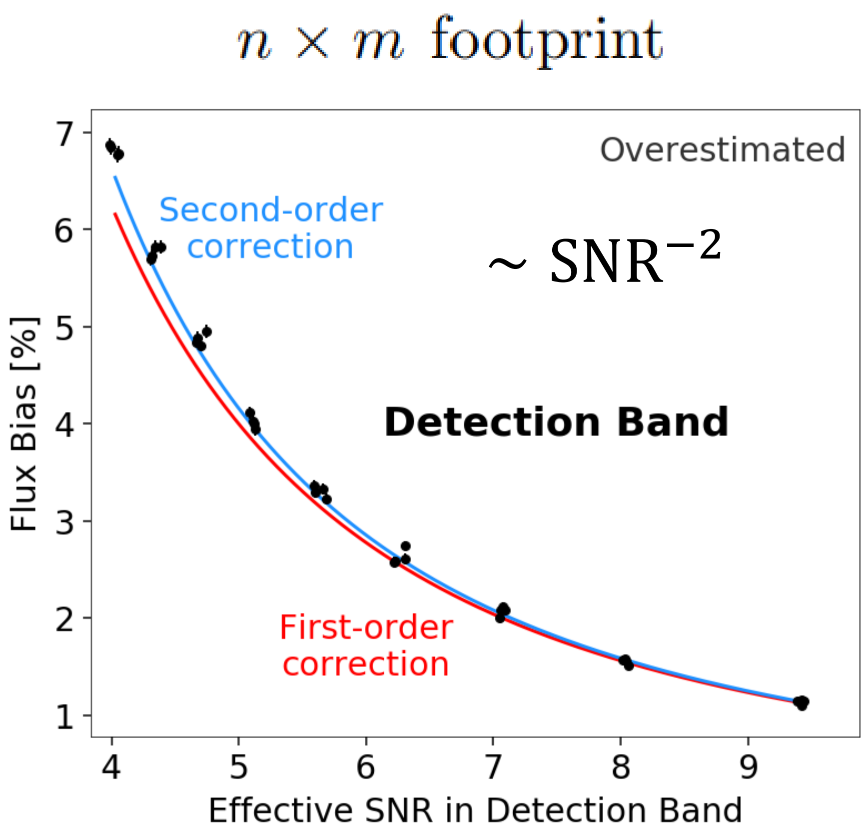


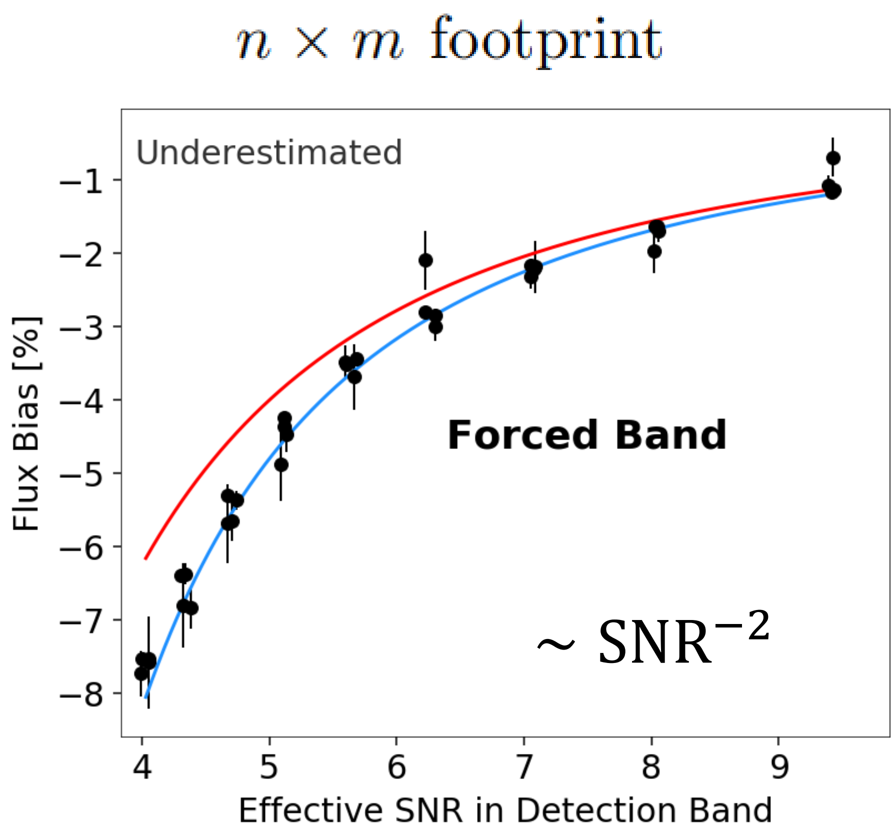
$n \times m$ footprint



Biases in Forced Multi-band Photometry

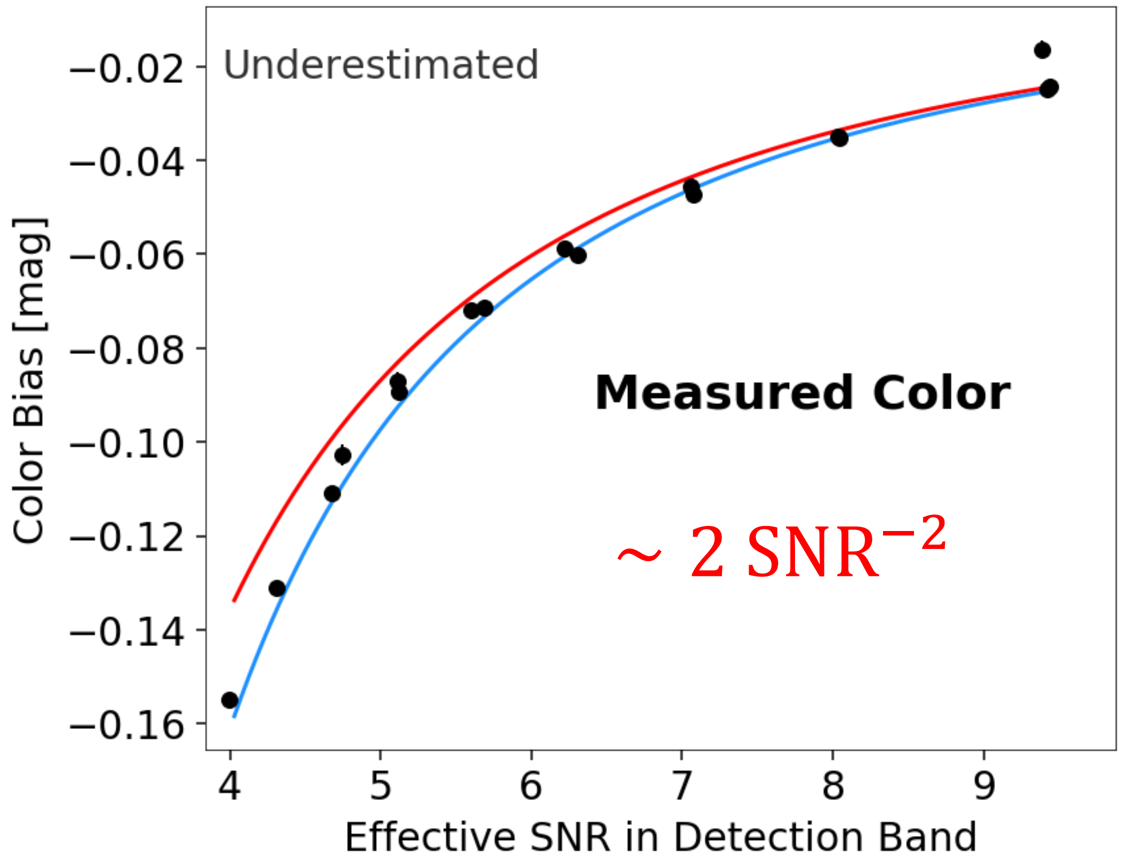
- Detect in one band, fix position, force photometry in others





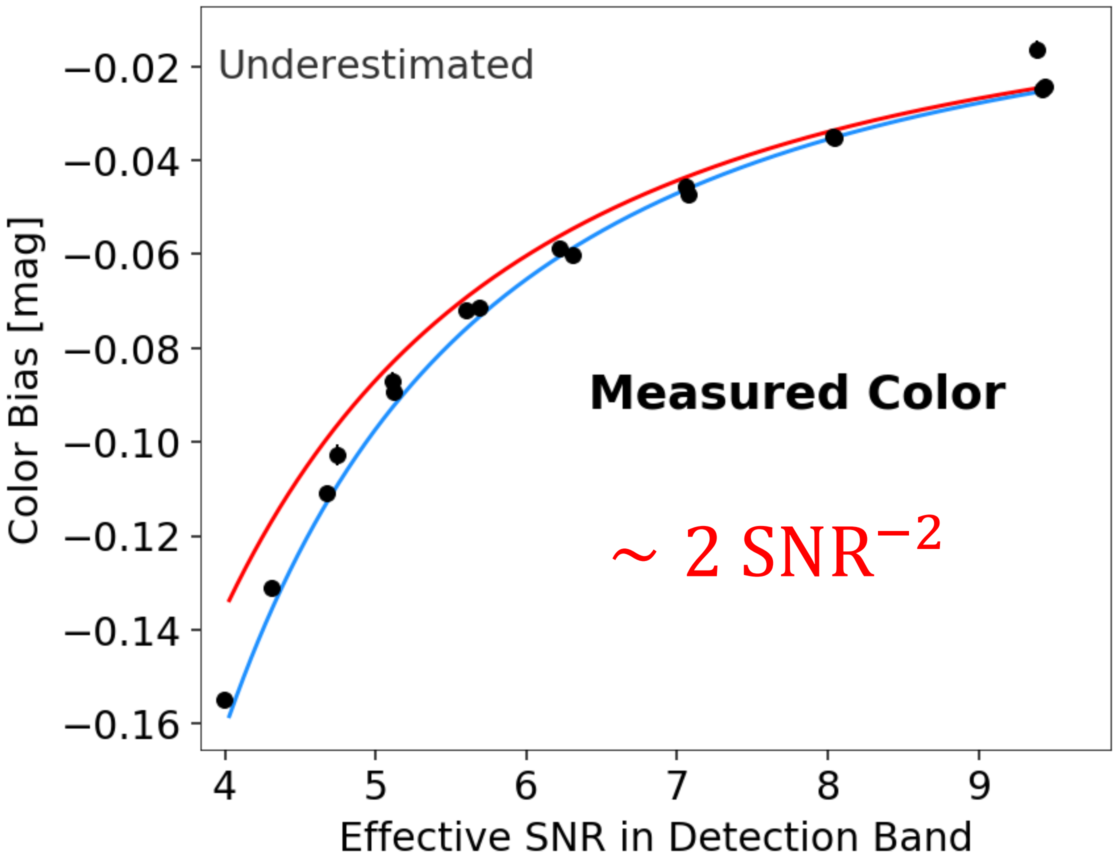
Biases in Forced Multi-band Photometry

- Detect in one band, fix position, force photometry in others



Biases in Forced Multi-band Photometry

- Detect in one band, fix position, force photometry in others

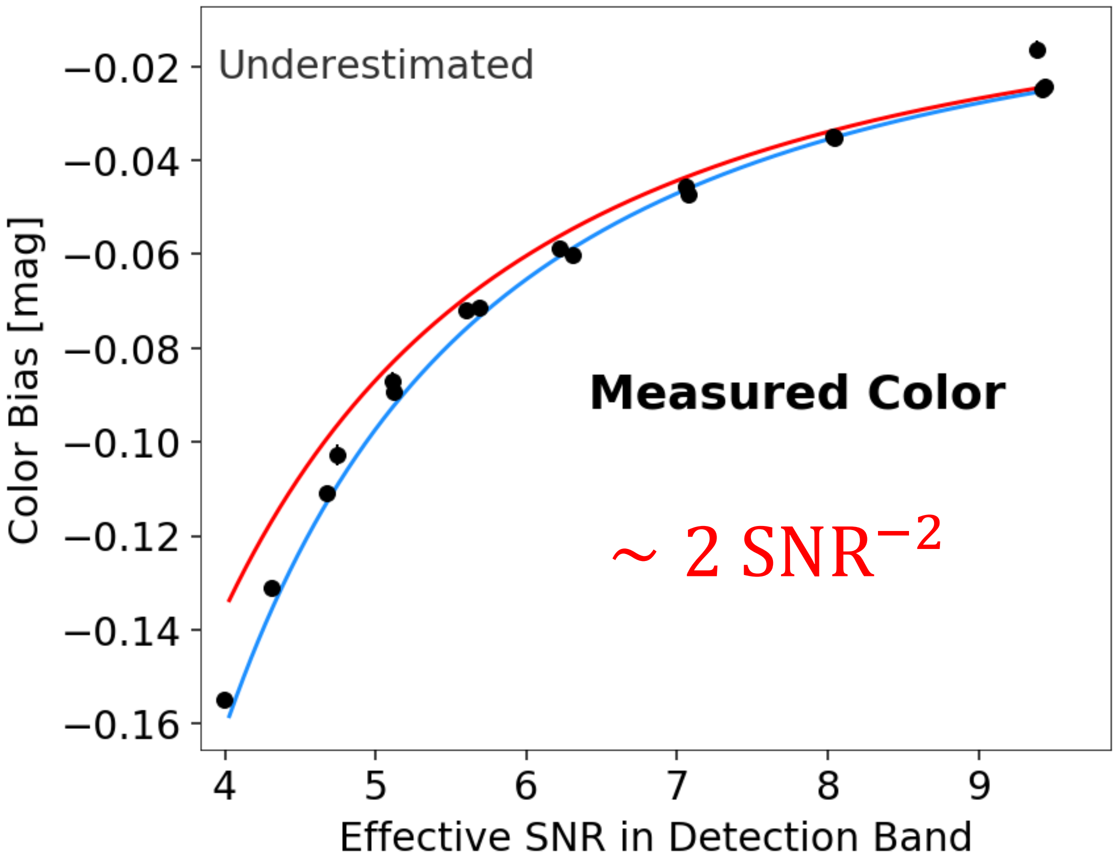


Star @ 10-sigma:

- 1% flux bias
- 0.02 mag color bias.

Biases in Forced Multi-band Photometry

- Detect in one band, fix position, force photometry in others



Star @ 10-sigma:

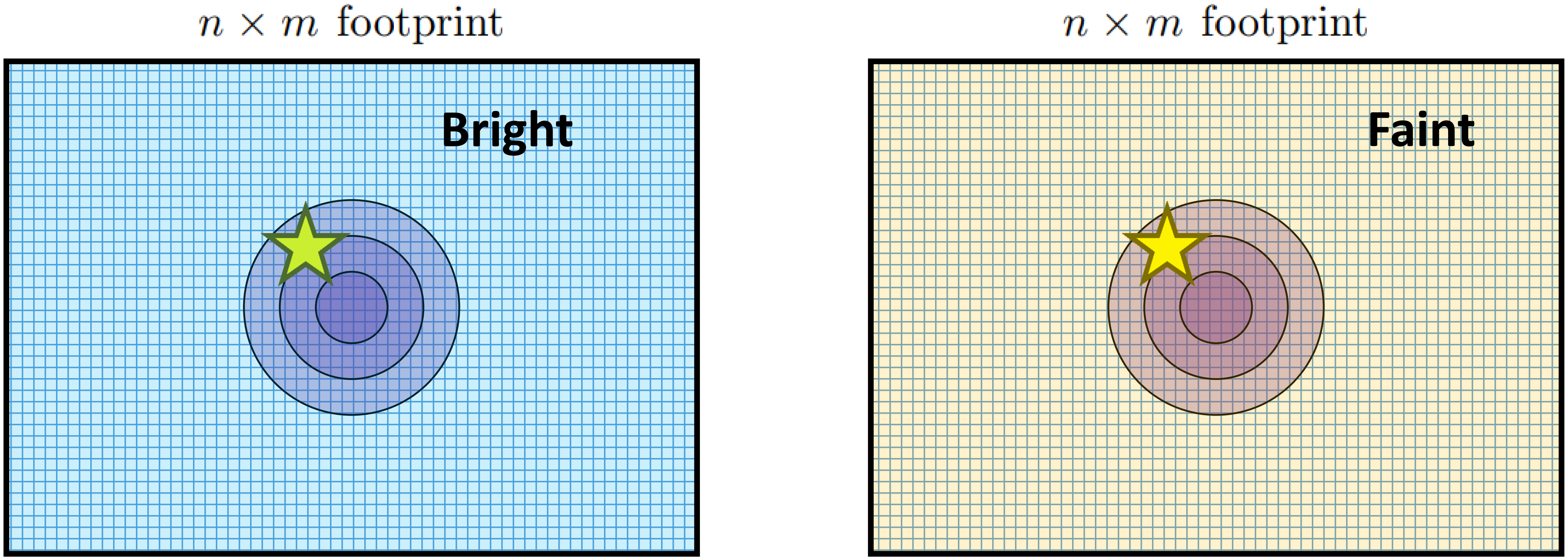
- 1% flux bias
- 0.02 mag color bias.

Galaxy @ 10-sigma:

- 2.5% flux bias
- **0.05 mag color bias.**

Biases in Joint Multi-band Photometry

- Fit all bands simultaneously (~ detect on stack, force in bands)



Biases in Joint Multi-band Photometry

- Fit all bands simultaneously (~ detect on stack, force in bands)

D. MAXIMUM-LIKELIHOOD BIASES USING BIAS TENSORS

As discussed in §5, ML estimators have a bias δ which tends to zero as the signal-to-noise ratio (SNR) increases. Cox & Snell (1968) found that the leading-order bias term for any parameter s can be found with

$$\delta_s(\boldsymbol{\theta}_{\rm ML}) = \sum_{r,t,u} (\mathcal{F}^{-1}(\boldsymbol{\theta}_{\rm ML}))_{rs}\, (\mathcal{F}^{-1}(\boldsymbol{\theta}_{\rm ML}))_{tu}\, (\mathcal{B}(\boldsymbol{\theta}_{\rm ML}))_{rtu} \tag{D34}$$

where

$$(\mathcal{B}(\boldsymbol{\theta}_{\rm ML}))_{rtu} \equiv \mathbb{E}_{\mathbf{D}}\left[\frac{1}{2}\partial_r\partial_t\partial_u \ln \mathcal{L}(\boldsymbol{\theta}_{\rm ML}) + (\partial_t \ln \mathcal{L}(\boldsymbol{\theta}_{\rm ML}))(\partial_r\partial_u \ln \mathcal{L}(\boldsymbol{\theta}_{\rm ML}))\,\middle|\, \boldsymbol{\theta}_{\rm ML}\right] \tag{D35}$$

is the bias tensor and $\mathbb{E}_{\mathbf{D}}[\cdot|\boldsymbol{\theta}_{\rm ML}]$ is the expectation value with respect to the data $\mathbf{D}$ for $\boldsymbol{\theta}_{\rm ML}$ fixed.

Biases in Joint Multi-band Photometry

- Fit all bands simultaneously (~ detect on stack, force in bands)

D. MAXIMUM-LIKELIHOOD BIASES USING BIAS TENSORS

As discussed in §5, ML estimators have a bias δ which tends to zero as the signal to noise ratio (SNR) increases. Cox & Snell (1968) found ... d with

$$\frac{\delta_{f_i}}{f_i} = \frac{\sigma^2_{f_i}}{f_i^2}\left(\sum_j \frac{f_j^2}{f_i^2}\frac{s_i^2\sigma^2_{f_i}}{s_j^2\sigma^2_{f_j}}\right)^{-1}$$

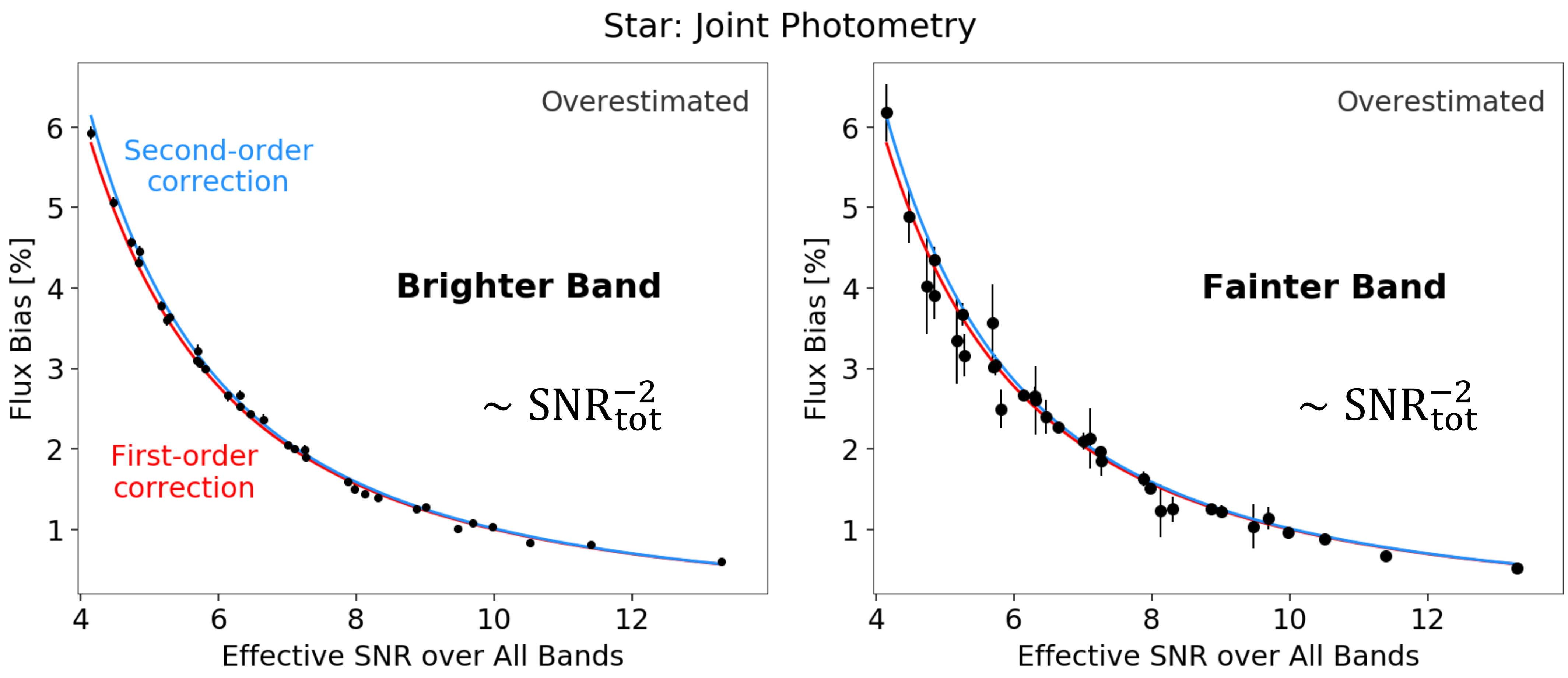
(D34)

where

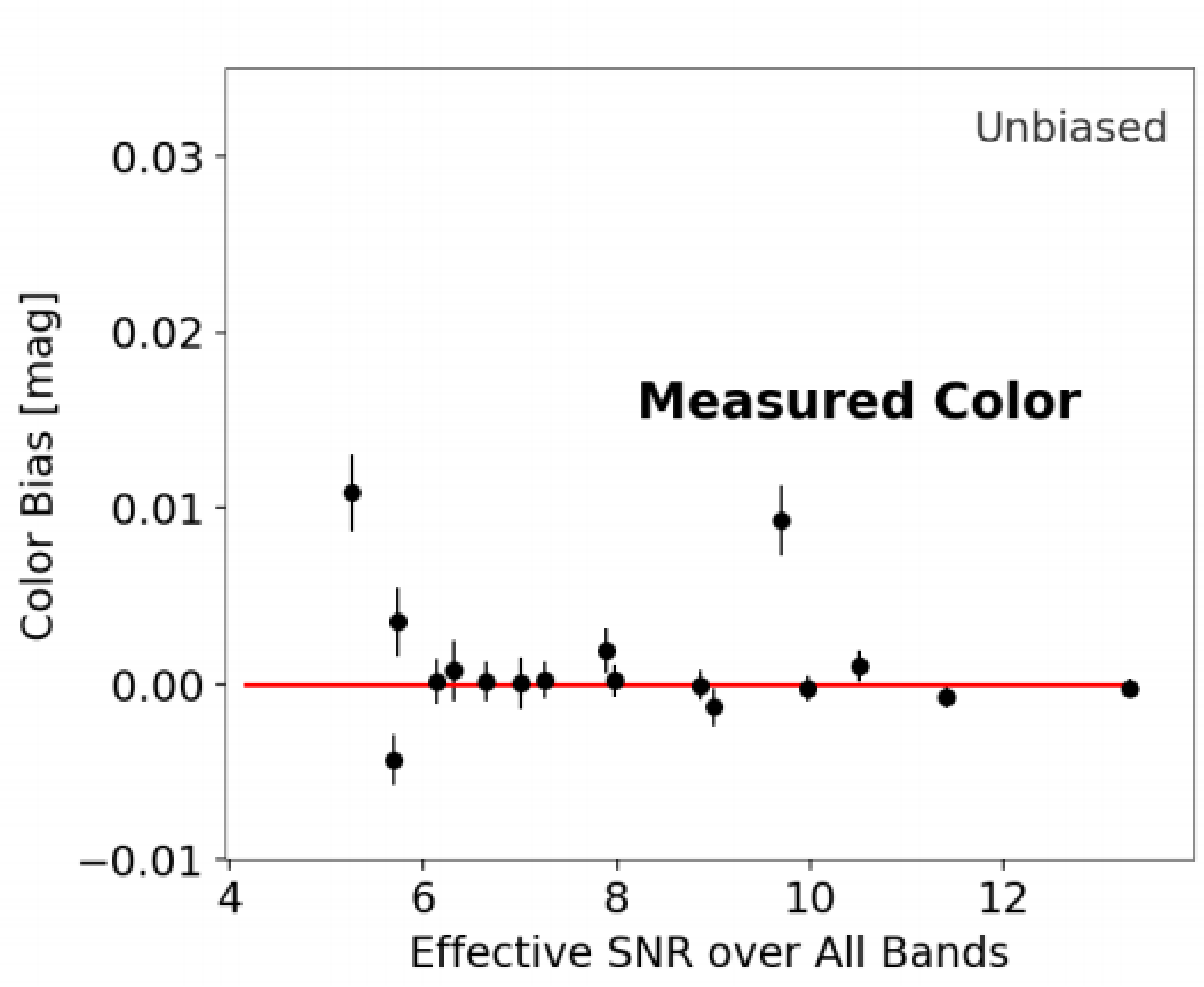
$(\boldsymbol{\mathcal{B}}(\boldsymbol{\theta}_{\rm M}$ ... $\Big|\boldsymbol{\theta}_{\rm ML}\Big]$ (D35)

is the bias tensor and $\mathbb{E}_{\rm D}$ ... L fixed.

Biases in Joint Multi-band Photometry



Biases in Joint Multi-band Photometry



Biases in Multi-band Photometry

Forced Photometry

Positive bias in detection band.
Negative bias in forced bands.
Doubly-biased colors.

Joint Photometry

Positive bias evenly spread across all bands.
Unbiased colors.

Proof?

Proof?

- Two test cases:
    - HSC SynPipe (mock data, real pipeline)
    - SDSS Stripe 82 (real data, real pipeline)

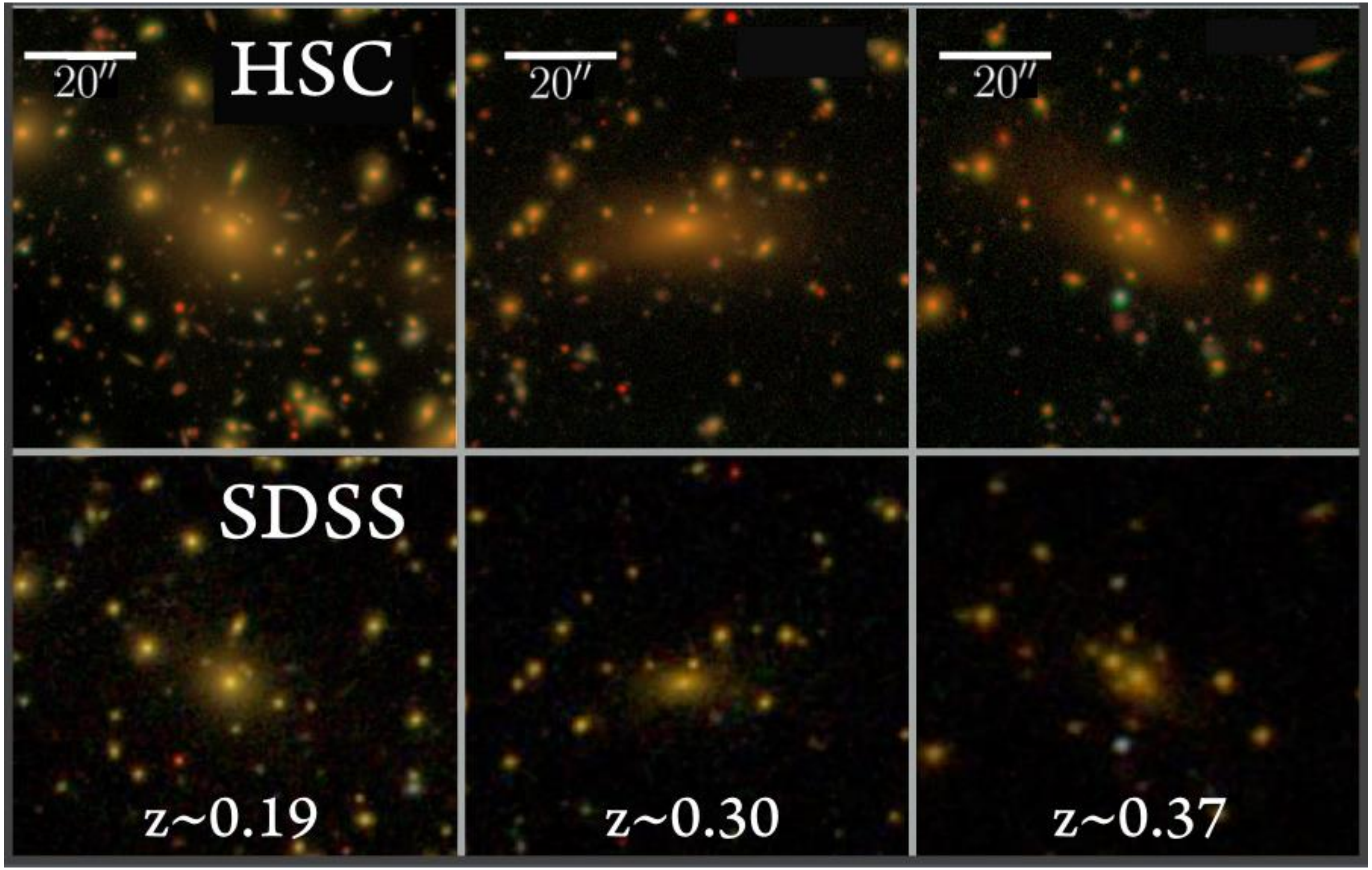
Mock Data: HSC SynPipe

Prime Focus
Instrument on Subaru
(8.2 m)

- Fov (1.5° in diameter)
  10xSuprime-Cam
- Primary science driver : weak lensing
- 2014 - 2019, 300 nights
- Deep multi-band imaging
  (grizy, i~26, y~24) over 1500 deg^2.
  About 940xCOSMOS!
- Excellent seeing (0.6”- 0.7”)
- In hand : 100 deg^2 to full depth

Credit: Alexie Leauthaud

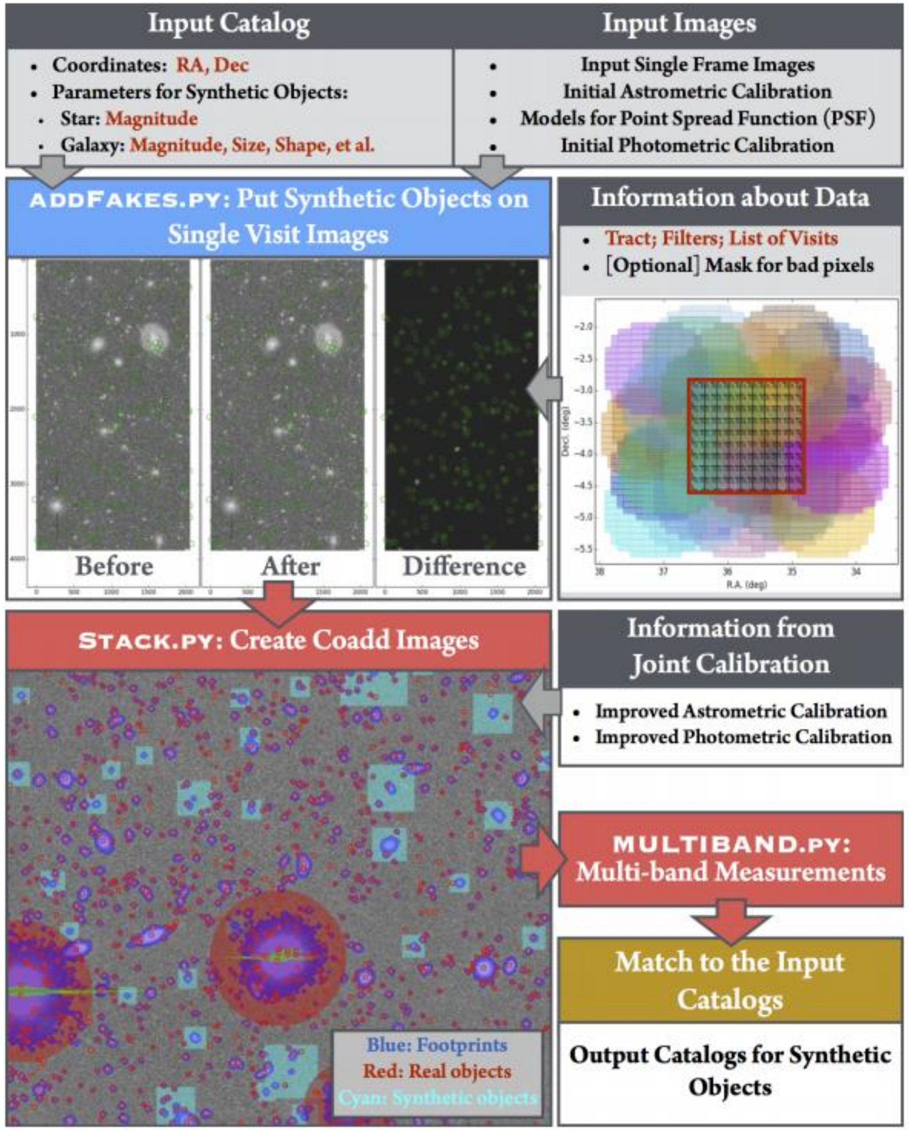
Mock Data: HSC SynPipe



Credit: Song Huang

Mock Data: HSC SynPipe

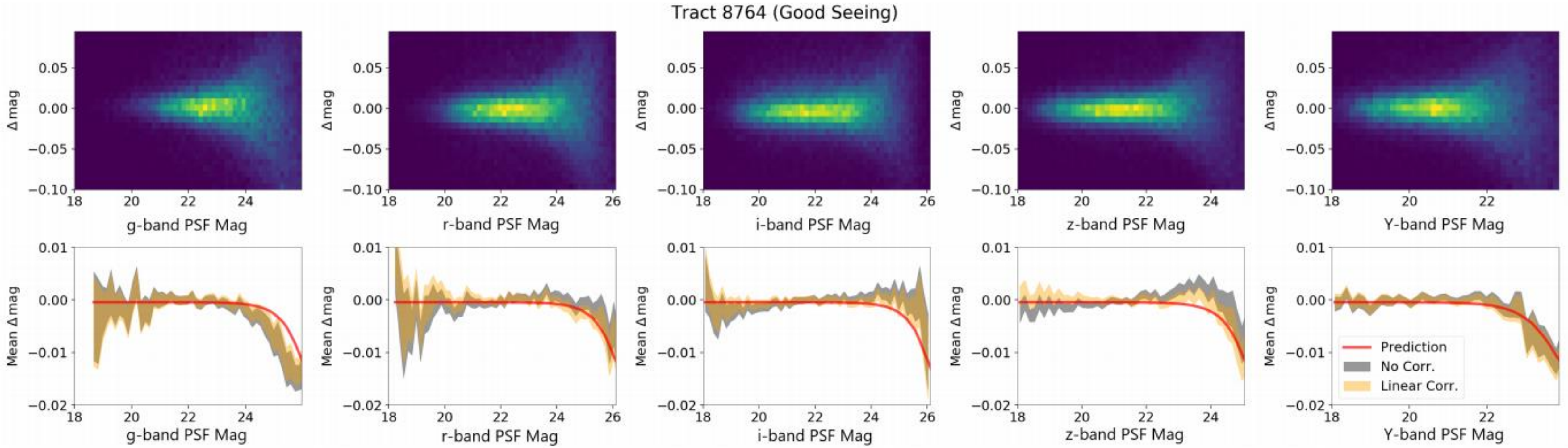
- **Fake object pipeline**: inject **fake** objects into **real** images drawn from realistic SED distributions.
- **“Forced” photometry**: detect in i-band, force in others
- **PSF magnitudes**



Credit: Song Huang

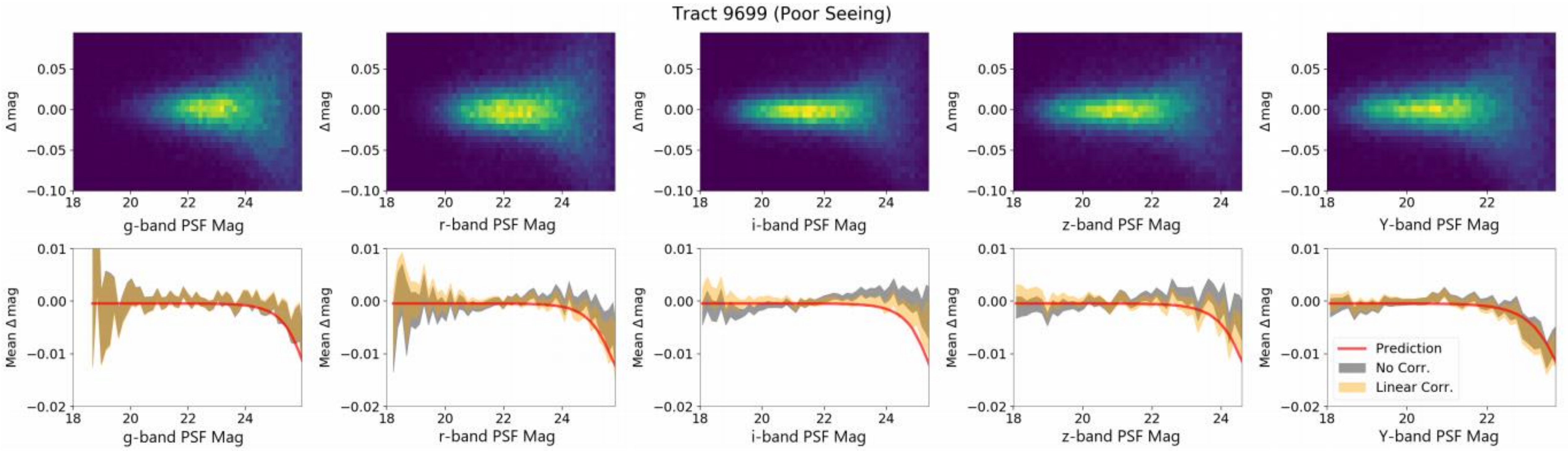
Mock Data: HSC SynPipe

Good seeing: ~0.6’’



Mock Data: HSC SynPipe

Poor seeing: ~1.2''



Mock Data: HSC SynPipe

- **Fake object pipeline**: inject **fake** objects into **real** images drawn from realistic SED distributions.
- **“Forced” photometry**: detect in i-band, force in others
- **PSF magnitudes**

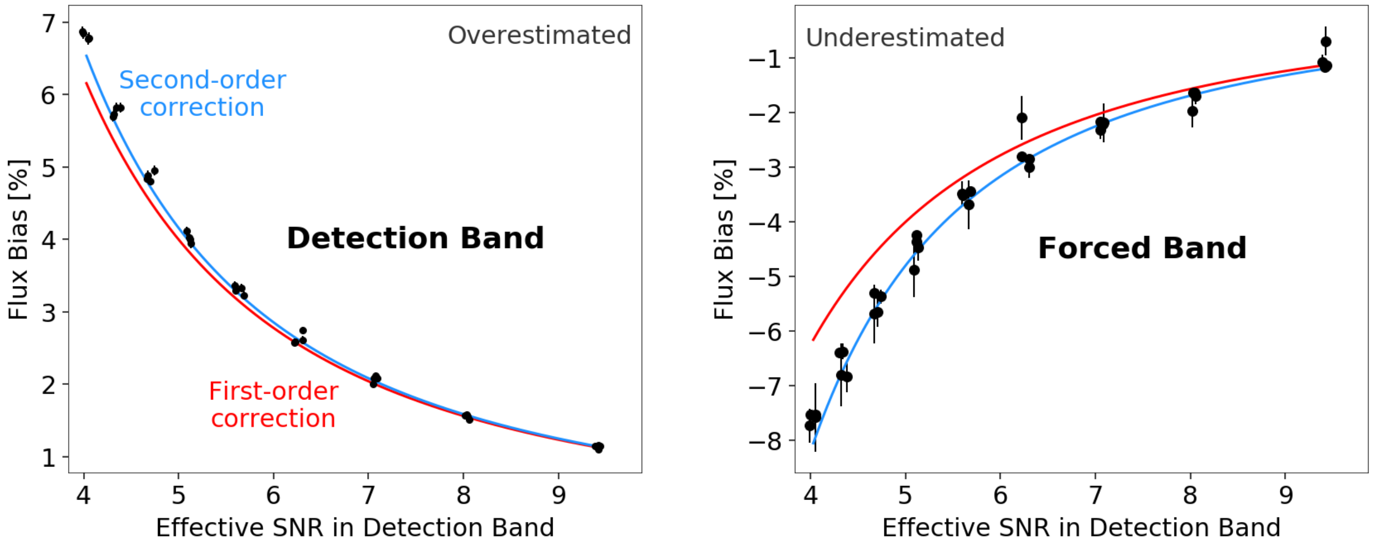
Mock Data: HSC SynPipe

- **Fake object pipeline**: inject **fake** objects into **real** images drawn from realistic SED distributions.
- **"Forced" photometry**: detect in i-band, force in others
- **PSF magnitudes** **???**

Mock Data: HSC SynPipe

- **Fake object pipeline**: inject **fake** objects into **real** images drawn from realistic SED distributions.
- **“Forced” photometry**: detect in i-band, force in others
- **PSF magnitudes** ???

Results look like single-band fits!



CENTER FOR ASTROPHYSICS HARVARD & SMITHSONIAN

DiRAC

The Devil’s in the Details: Photometric Biases in Modern Surveys

Stephen Portillo[1,2,*] and **Josh Speagle**[1,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington)

[*]Equal contribution

Portillo, Speagle, & Finkbeiner (**PSF**) subm., arxiv:1902.02374



The Devil's in the Details: Photometric Biases in Modern Surveys

Stephen Portillo[1,2,*] and **Josh Speagle**[1,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington)

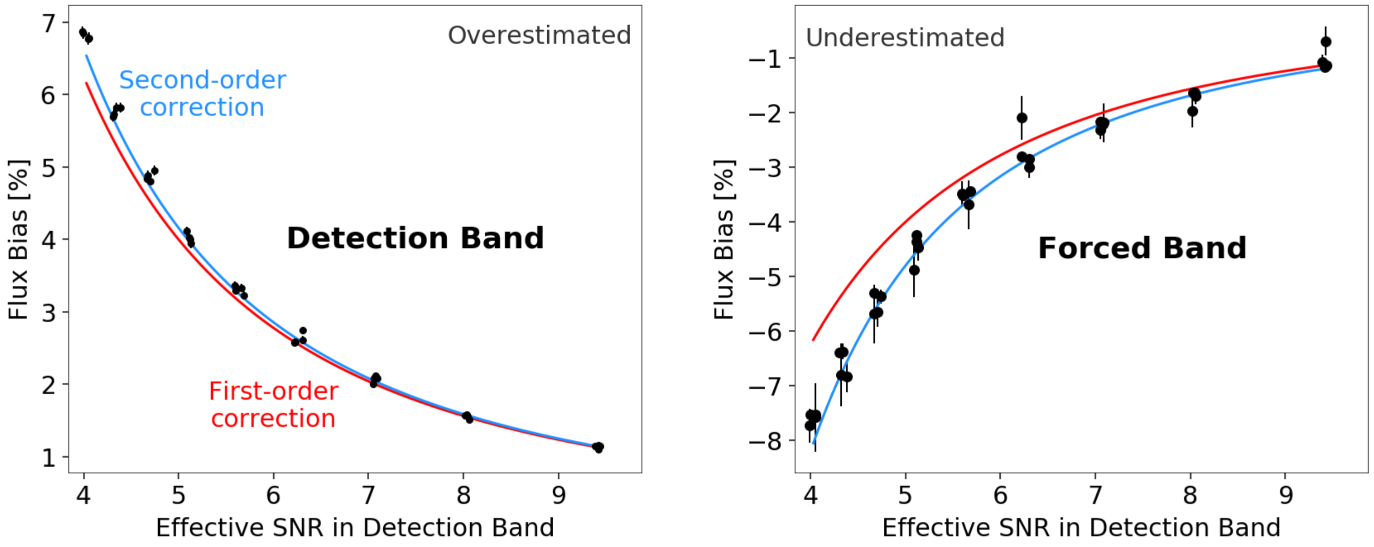
[*]Equal contribution

Portillo, Speagle, & Finkbeiner (**PSF**) subm., arxiv:1902.02374

Mock Data: HSC SynPipe

- **Fake object pipeline**: inject **fake** objects into **real** images drawn from realistic SED distributions.
- **“Forced” photometry**: detect in i-band, force in others
- **PSF magnitudes** ???

Results look like single-band fits!



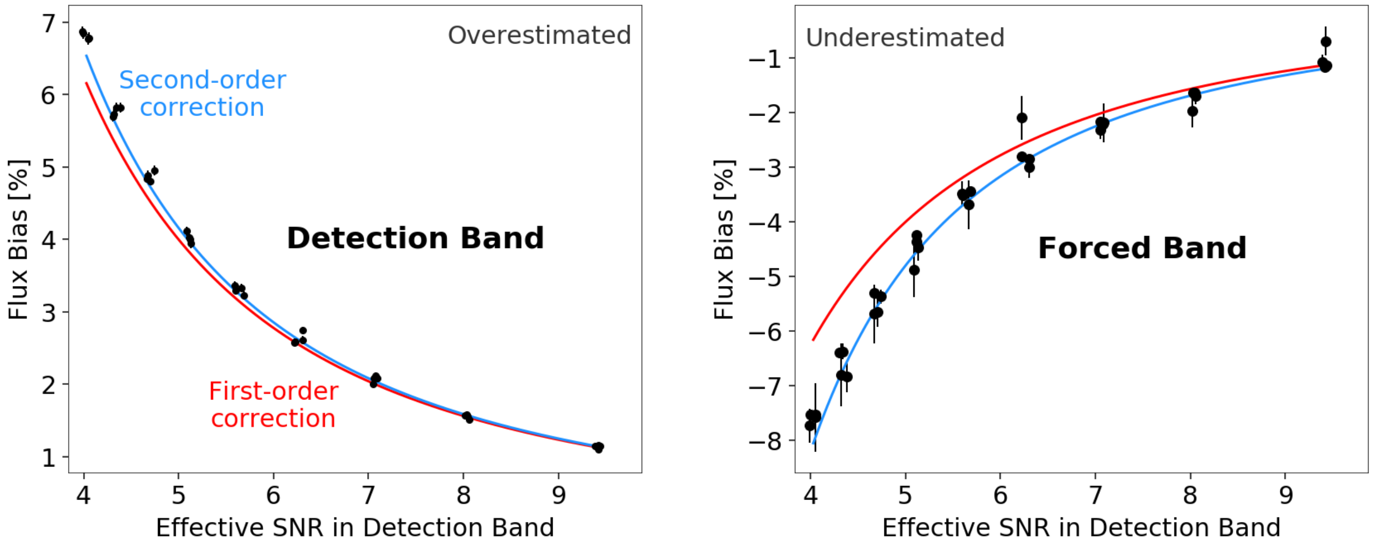
Mock Data: HSC SynPipe

- **Fake object pipeline**: inject **fake** objects into **real** images drawn from realistic SED distributions.
- **"Forced" photometry**: detect in i-band, force in others
- **PSF magnitudes ???**

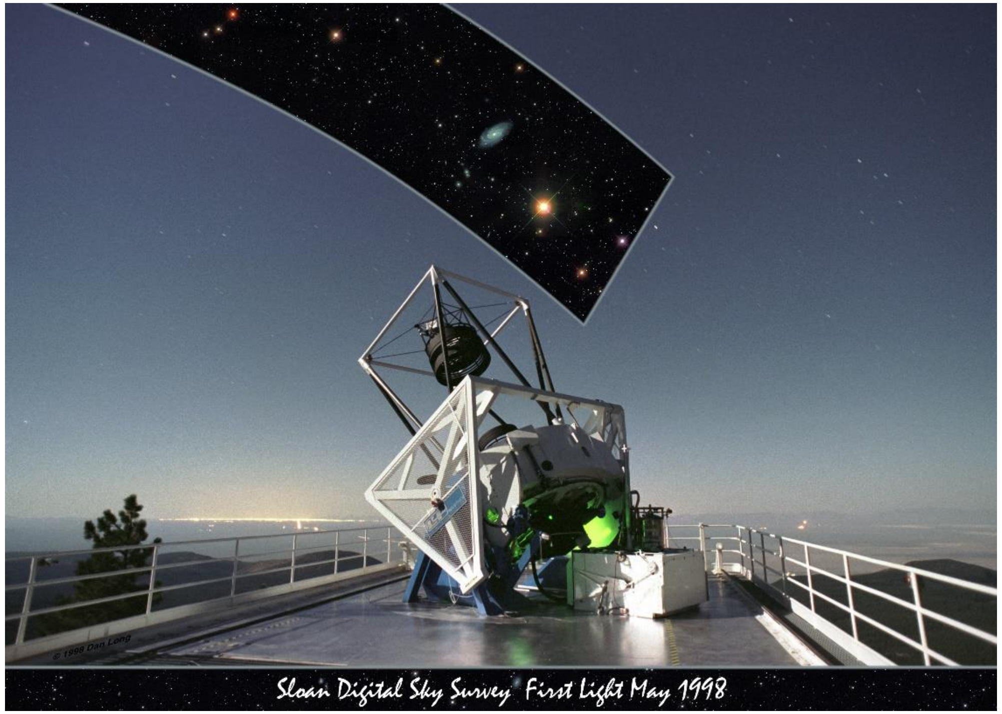
Results look like single-band fits!

Implementation-specific effect:

- HSC pipeline allows for **"local re-centering"** of forced position.
- Enough to undo forcing effect!



Real Data: Stripe 82



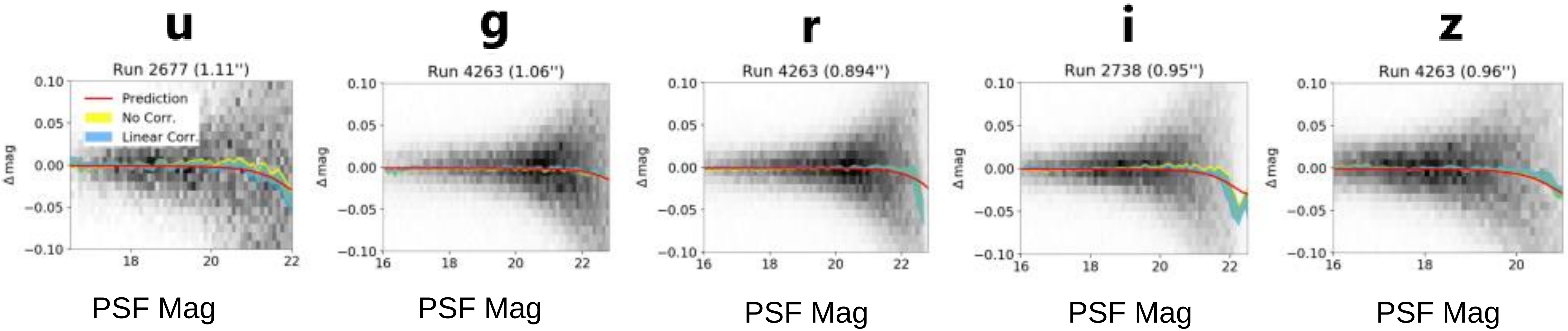
sdss.org

cfht.hawaii.org

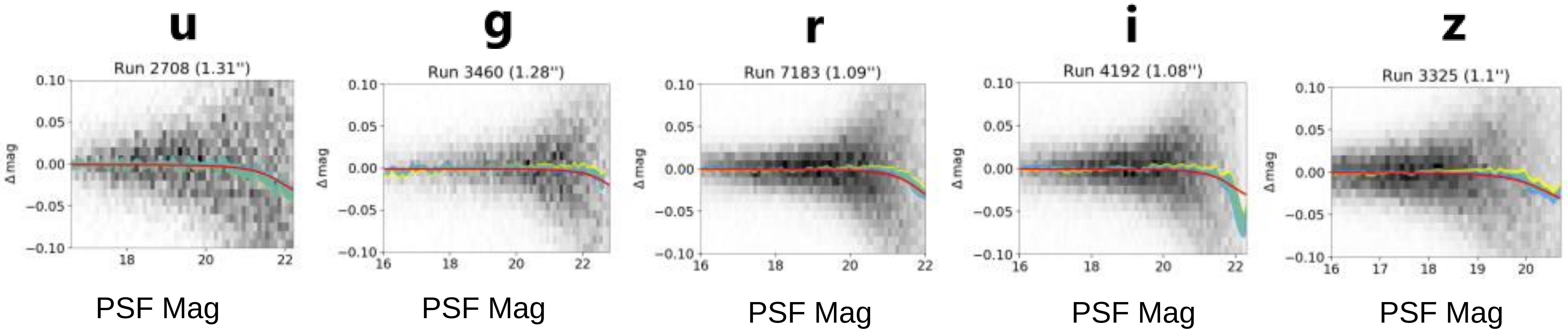
Real Data: Stripe 82

- **Repeated imaging**: compare catalog computed from a **"deep stack"** of all images ("truth") vs **individual runs** ("realization")
- **"Forced" photometry**: detect in r-band, force in others
- **PSF magnitudes**

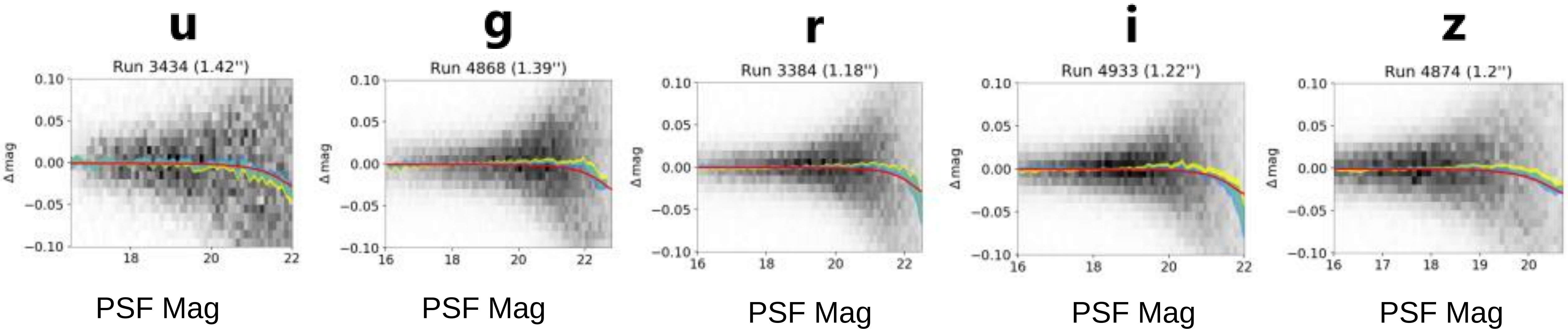
Real Data: Stripe 82



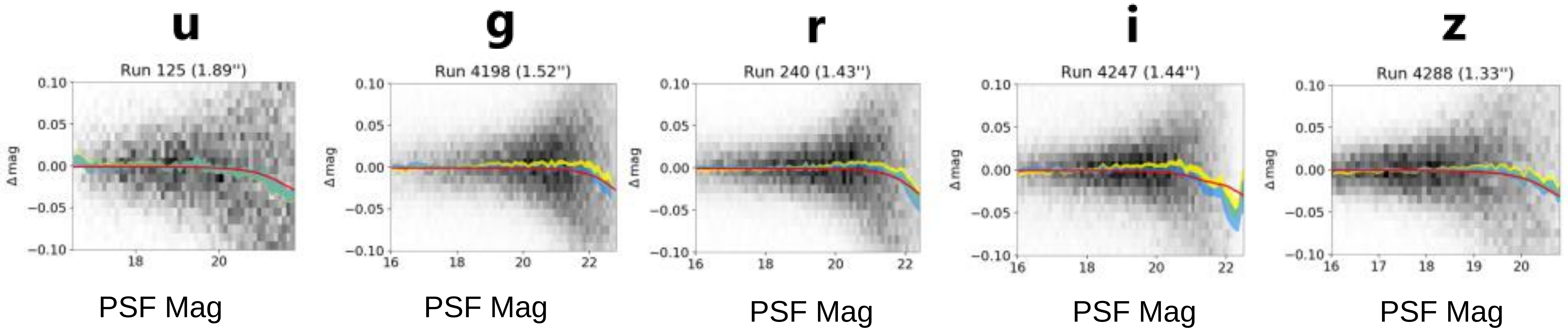
Real Data: Stripe 82



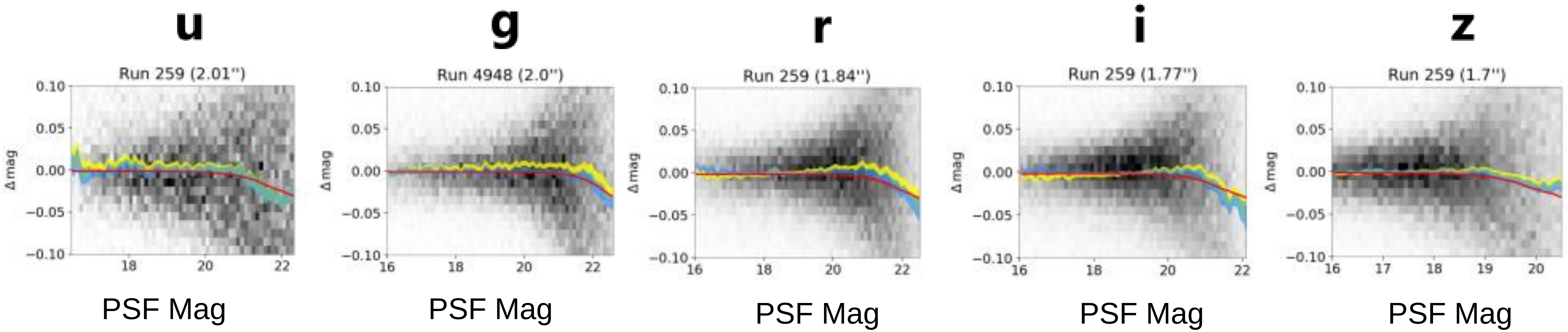
Real Data: Stripe 82



Real Data: Stripe 82



Real Data: Stripe 82



Real Data: Stripe 82

- **Repeated imaging**: compare catalog computed from a **“deep stack”** of all images (“truth”) vs **individual runs** (“realization”)
- **“Forced” photometry**: detect in r-band, force in others
- **PSF magnitudes**

Real Data: Stripe 82

- **Repeated imaging**: compare catalog computed from a **“deep stack”** of all images (“truth”) vs **individual runs** (“realization”)
- **“Forced” photometry**: detect in r-band, force in others
- **PSF magnitudes**

Real Data: Stripe 82

- **Repeated imaging**: compare catalog computed from a **“deep stack”** of all images (“truth”) vs **individual runs** (“realization”)
- **“Forced” photometry**: detect in r-band, force in others
- **PSF magnitudes**

- **HSC pipeline built off of SDSS pipeline.**
  Same implementation and effect.

Real Data: Stripe 82

- **Repeated imaging**: compare catalog computed from a **"deep stack"** of all images ("truth") vs **individual runs** ("realization")
- **"Forced" photometry**: detect in r-band, force in others
- **PSF magnitudes**

- **HSC pipeline built off of SDSS pipeline. Same implementation and effect.**
- **SDSS/HSC will also serve as basis for LSST pipeline.**

What about galaxies?

What about galaxies?

- **Much** more complicated:
  - Models are known to have deficiencies matching real data.
  - Models involve much more parameters.
  - Many more algorithmic choices involved/taken when fitting.

What about galaxies?

- **Much** more complicated:
  - Models are known to have deficiencies matching real data.
  - Models involve much more parameters.
  - Many more algorithmic choices involved/taken when fitting.

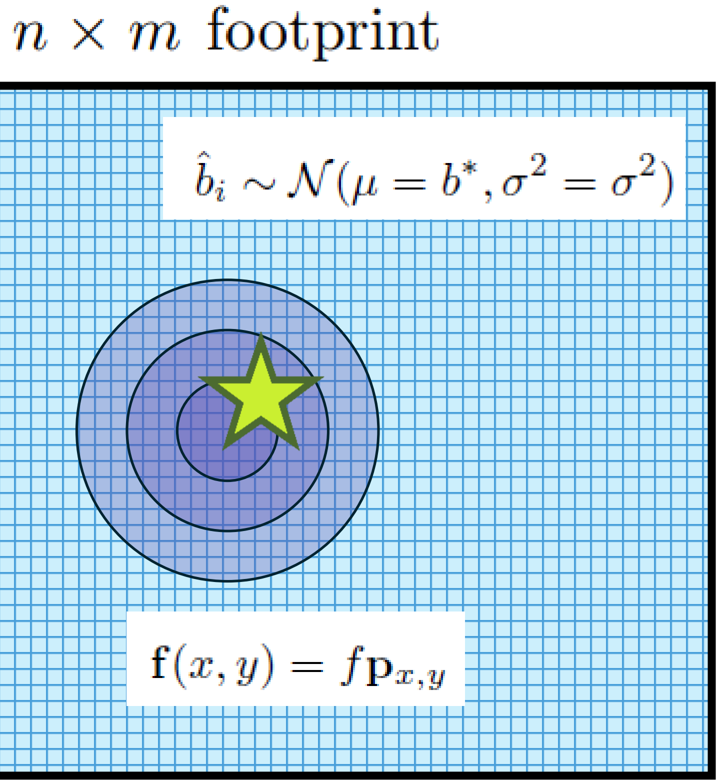
- Recommend directly calibrating from pipeline tests.

Cautionary note: Aperture photometry

- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?

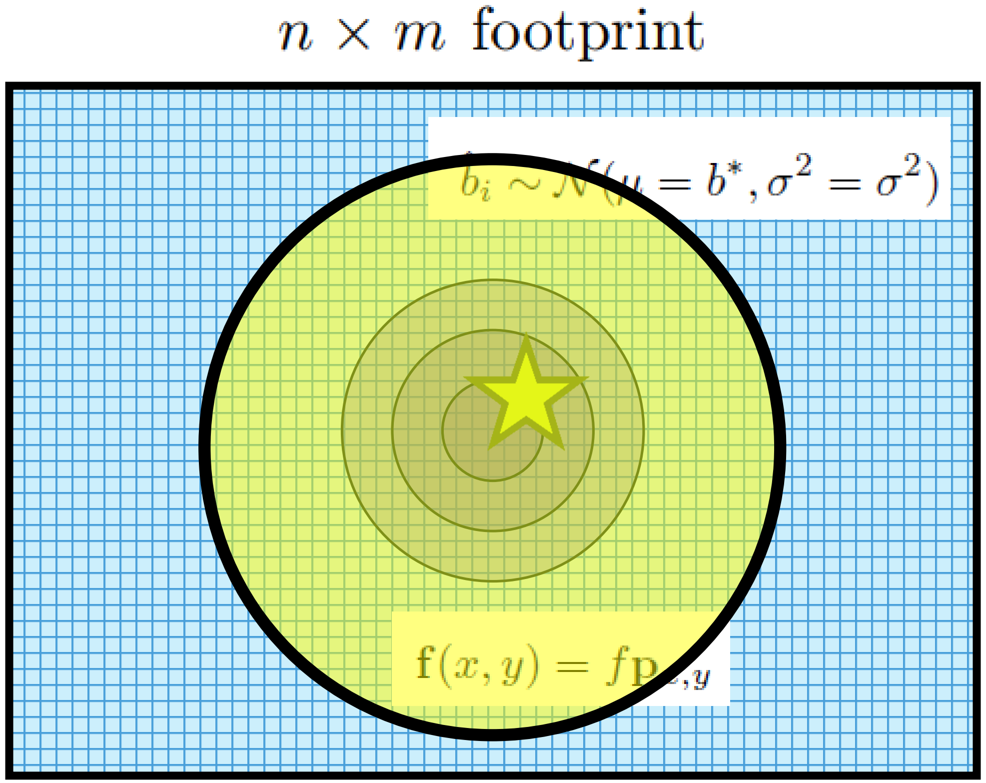
Cautionary note: Aperture photometry

- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?



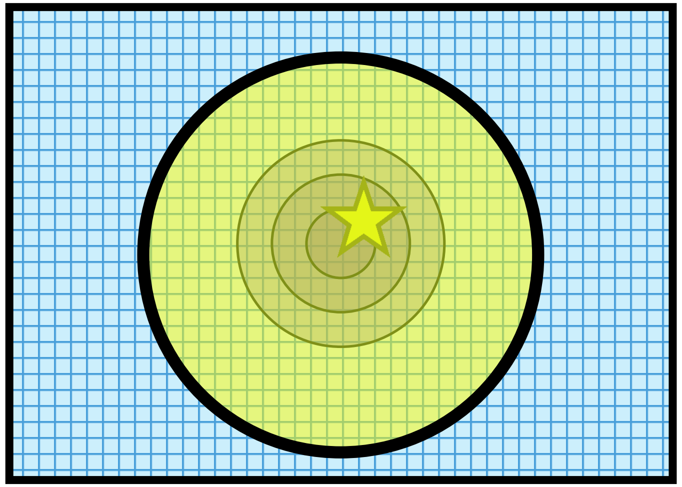
Cautionary note: Aperture photometry

- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?



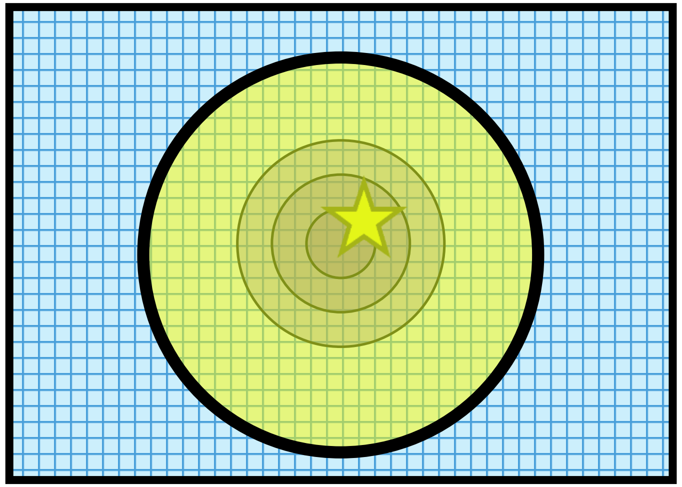
Cautionary note: Aperture photometry

- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?



Cautionary note: Aperture photometry

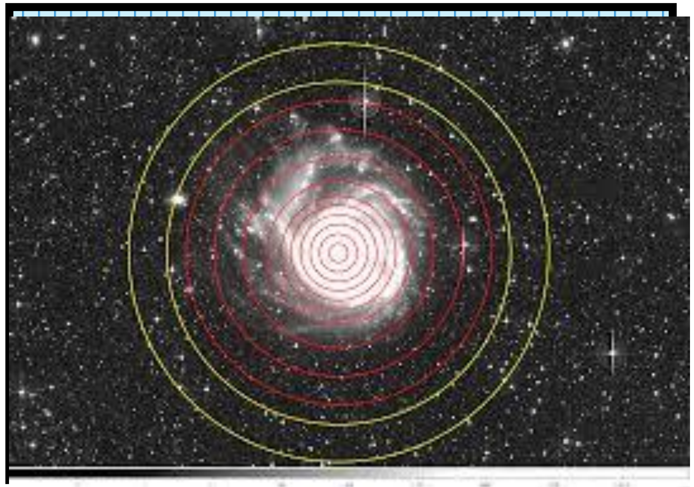
- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?
  1. Apertures will “miss” flux in *all* cases, requiring “**aperture corrections**” that dominate the error budget and are hard to estimate.
  2. Apertures are still subject to centering bias.
  3. SNR is lower due to much more background noise.
  4. Estimate degrades (MLE improves) with larger area.
  5. Unable to jointly model multiple images.
  6. Less amenable to statistical analysis than MLE.



Cautionary note: Aperture photometry

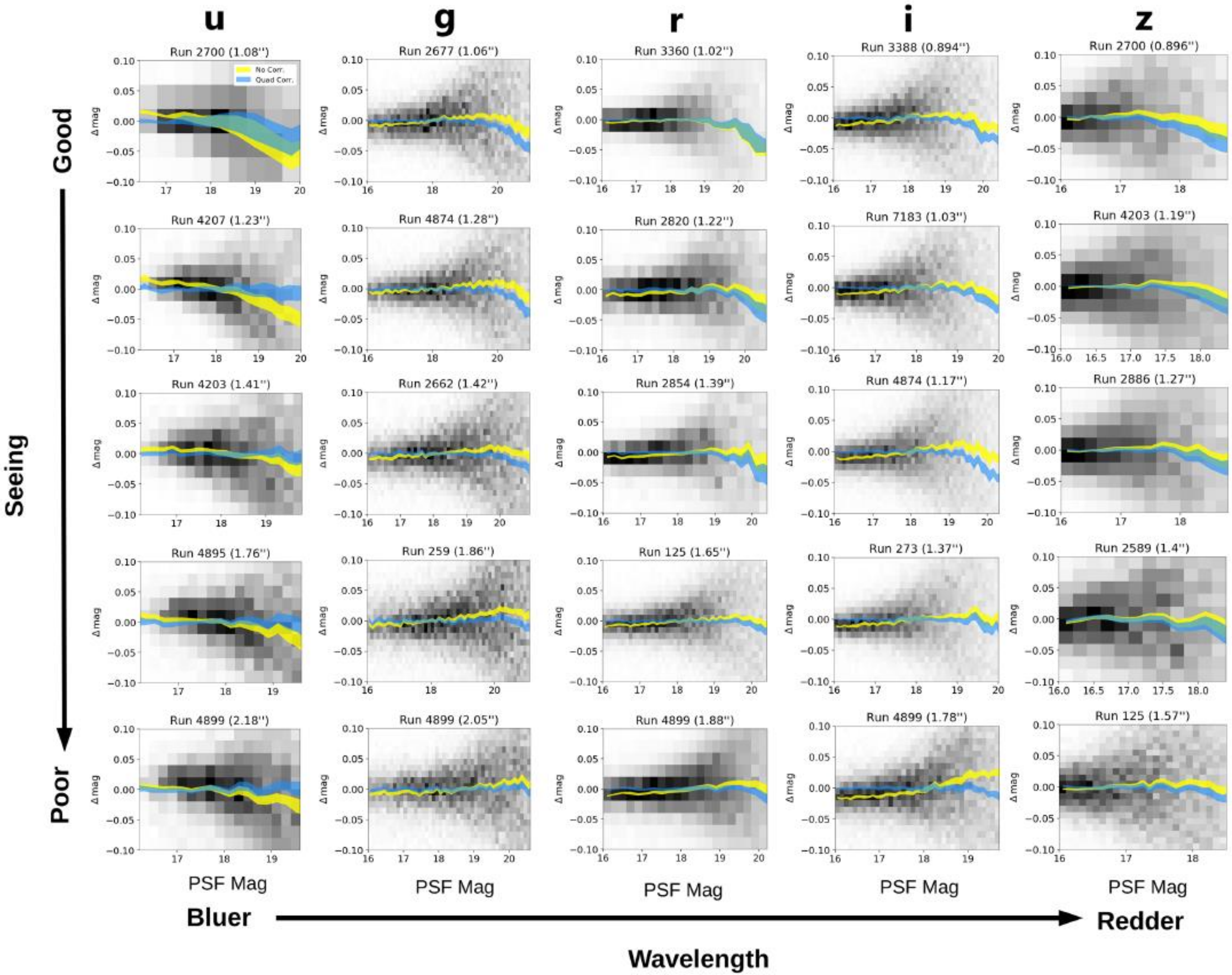
- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?
  1. Apertures will "miss" flux in *all* cases, requiring "**aperture corrections**" that dominate the error budget and are hard to estimate.
  2. Apertures are still subject to centering bias.
  3. SNR is lower due to much more background noise.
  4. Estimate degrades (MLE improves) with larger area.
  5. Unable to jointly model multiple images.
  6. Less amenable to statistical analysis than MLE.

burro.case.edu



Has a purpose, but should be used judiciously!

Aperture vs PSF Photometry: SDSS Stripe 82



Summary

- MLE photometry has a bias that goes as **SNR^-2**.
- Proportional to number of parameters of fit: **(p-1)/2**
- Naïve errors underestimated due to ignored covariances.

Summary

- MLE photometry has a bias that goes as **SNR^-2**.
- Proportional to number of parameters of fit: **(p-1)/2**
- Naïve errors underestimated due to ignored covariances.

- Mild effect for stars (1% @ SNR=10), **more severe for galaxies** (>2.5% @ SNR=10).
- **Forced photometry is dangerous.** Joint is better.

Summary

- MLE photometry has a bias that goes as **SNR^-2**.
- Proportional to number of parameters of fit: **(p-1)/2**
- Naïve errors underestimated due to ignored covariances.

- Mild effect for stars (1% @ SNR=10), **more severe for galaxies** (>2.5% @ SNR=10).
- **Forced photometry is dangerous.** Joint is better.

- Behavior is sensitive to implementation – **talk to pipeline teams**!
- These biases likely present in many modern photometry catalogs.