Astrostatistics Interest Group: Student Paper Award
Topic Contributed Papers Session (401) JSM 2020
August 5 2020

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

Stephen Portillo[1,2,*] and **Josh Speagle**[1,3,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington), [3]U. of Toronto

[*]Equal contribution

Portillo, Speagle, & Finkbeiner (**PSF**) 2020, 1902.02374

Astrostatistics Interest Group: Student Paper Award
Topic Contributed Papers Session (401) JSM 2020
August 5 2020

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

Stephen Portillo[1,2,*] and **Josh Speagle**[1,3,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington), [3]U. of Toronto

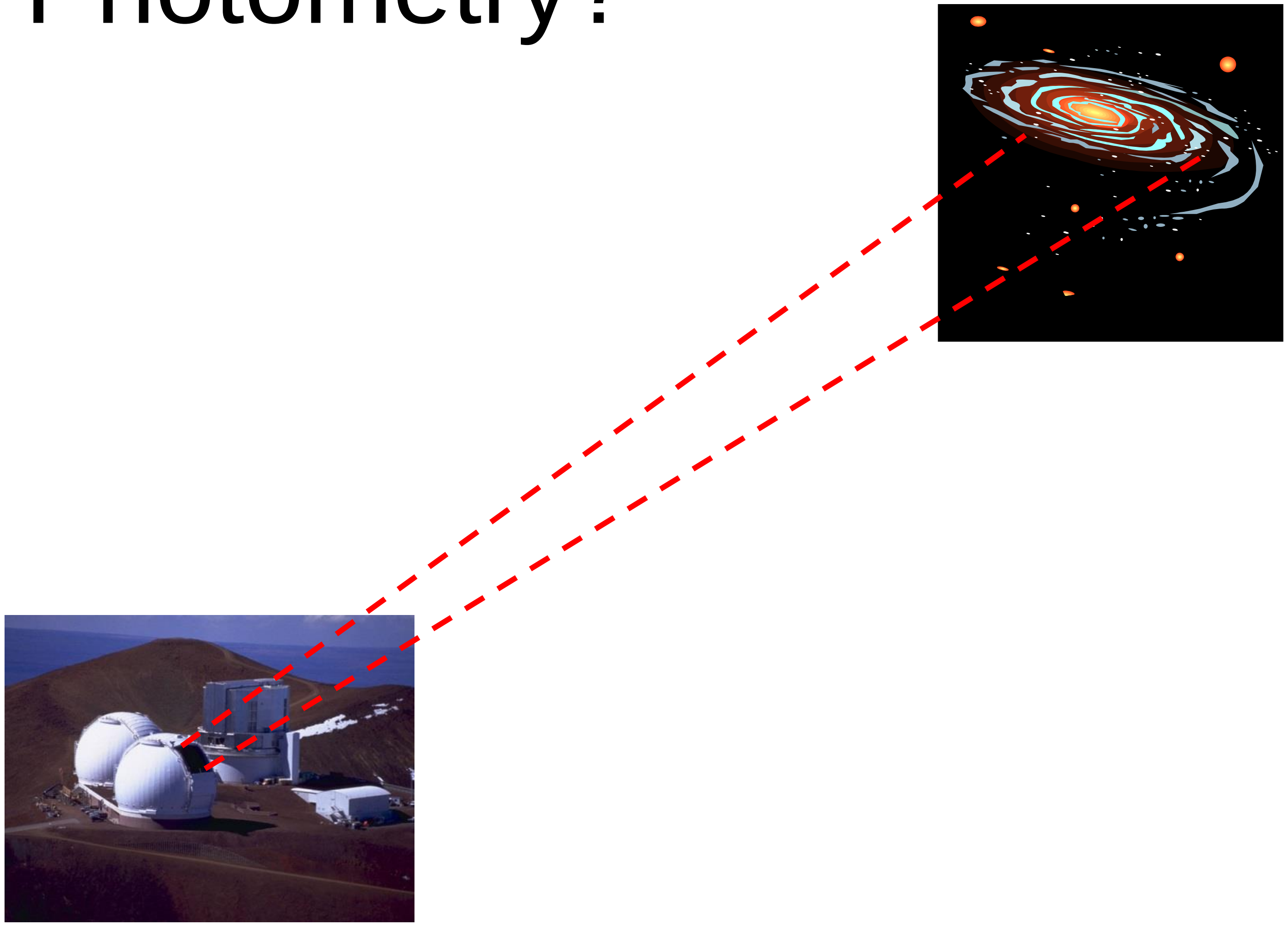
[*]Equal contribution



All data, code, and plots available online at https://github.com/joshspeagle/phot_bias

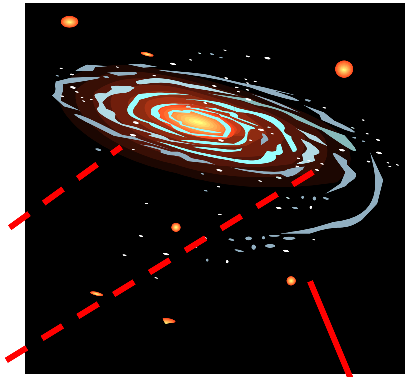
Portillo, Speagle, & Finkbeiner (**PSF**) 2020, 1902.02374

What is Photometry?

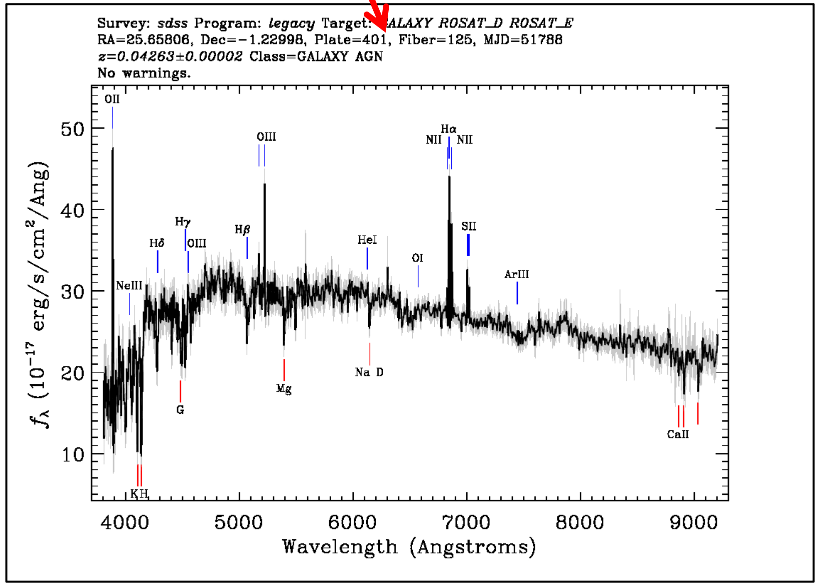


Keck Telecope (Hawaii)

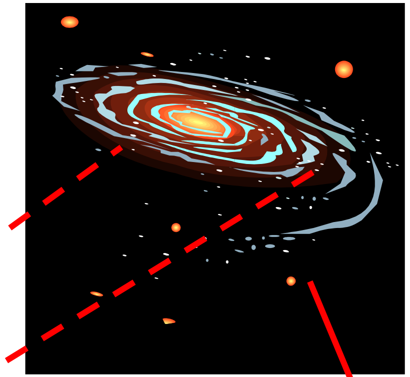
What is Photometry?



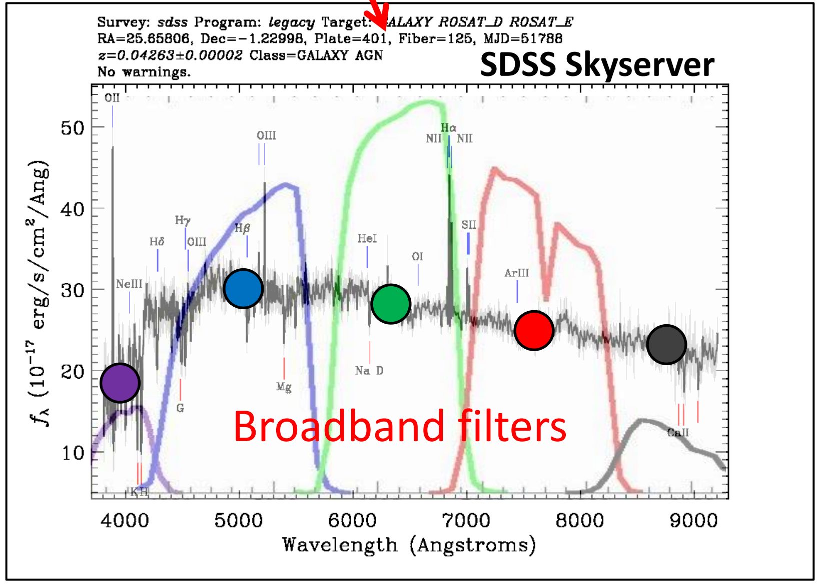
Keck Telecope (Hawaii)



What is Photometry?

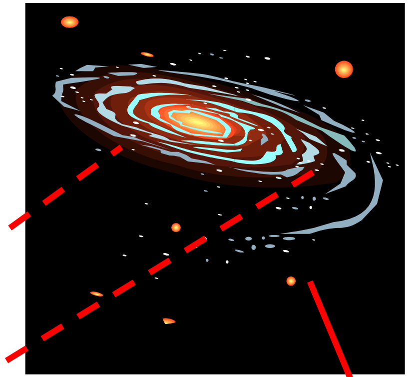


Keck Telecope (Hawaii)

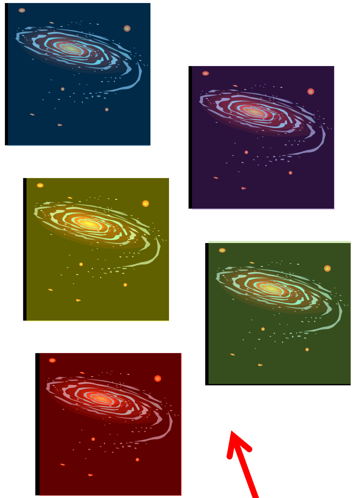


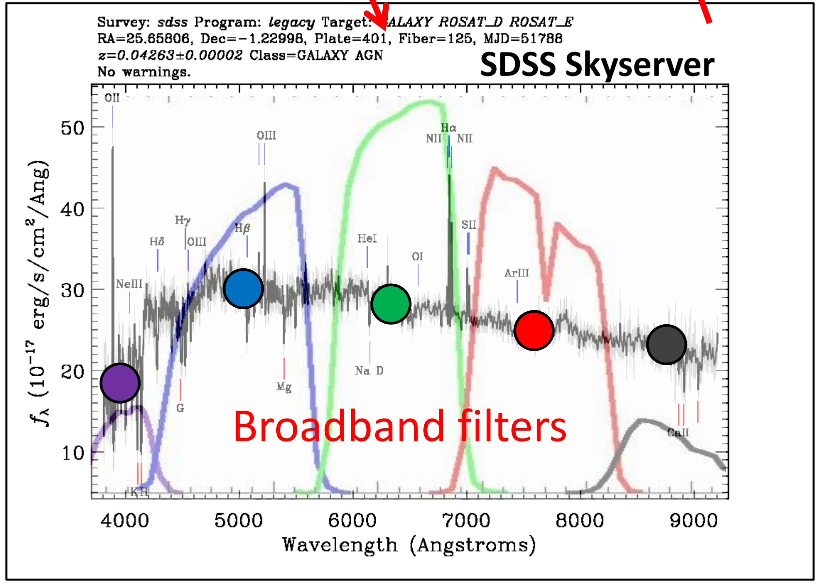
What is Photometry?

Keck Telecope (Hawaii)

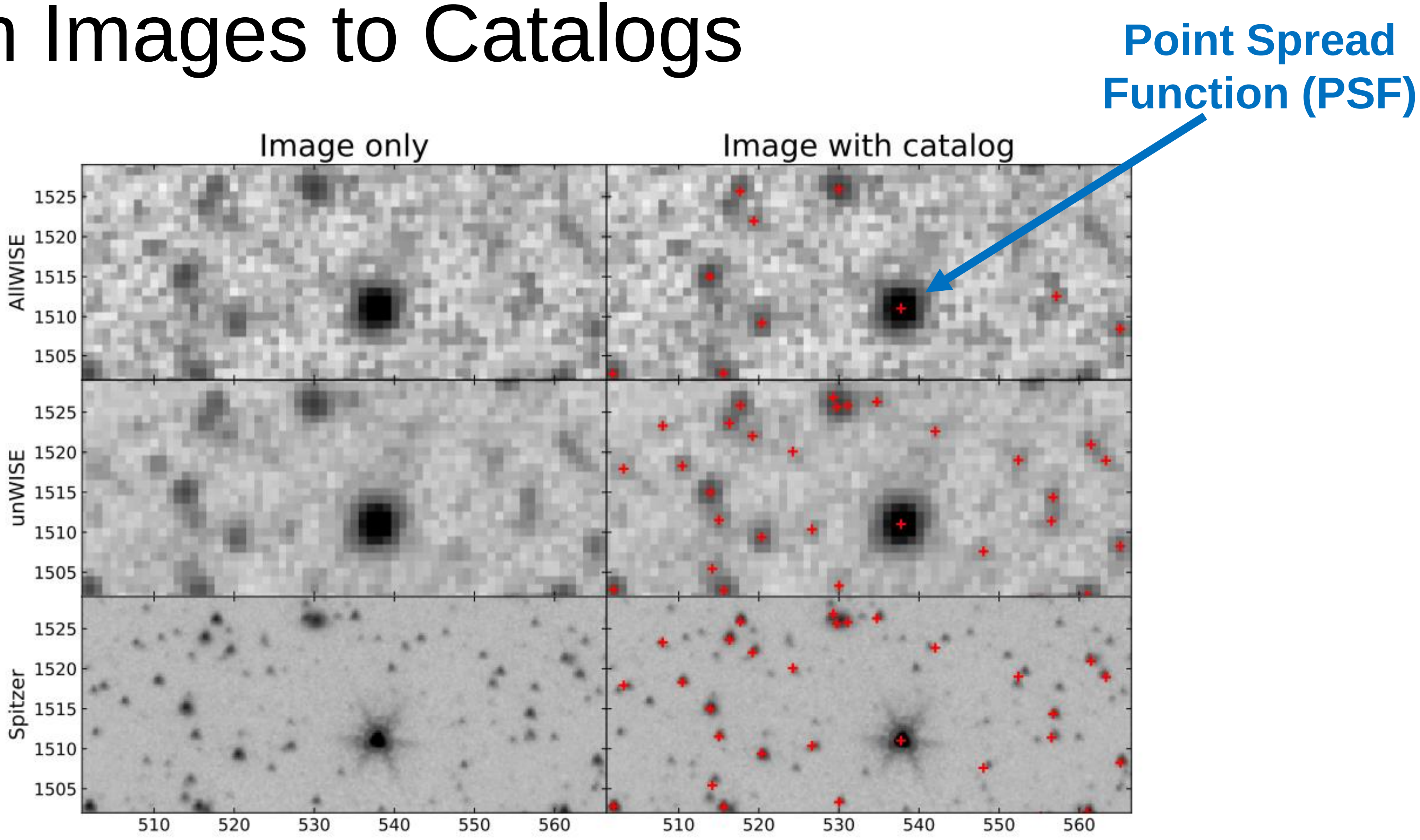


Spectral Energy Distribution (SED)





From Images to Catalogs



unWISE: Schlafly et al. (2019)

Motivation

“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.

Results

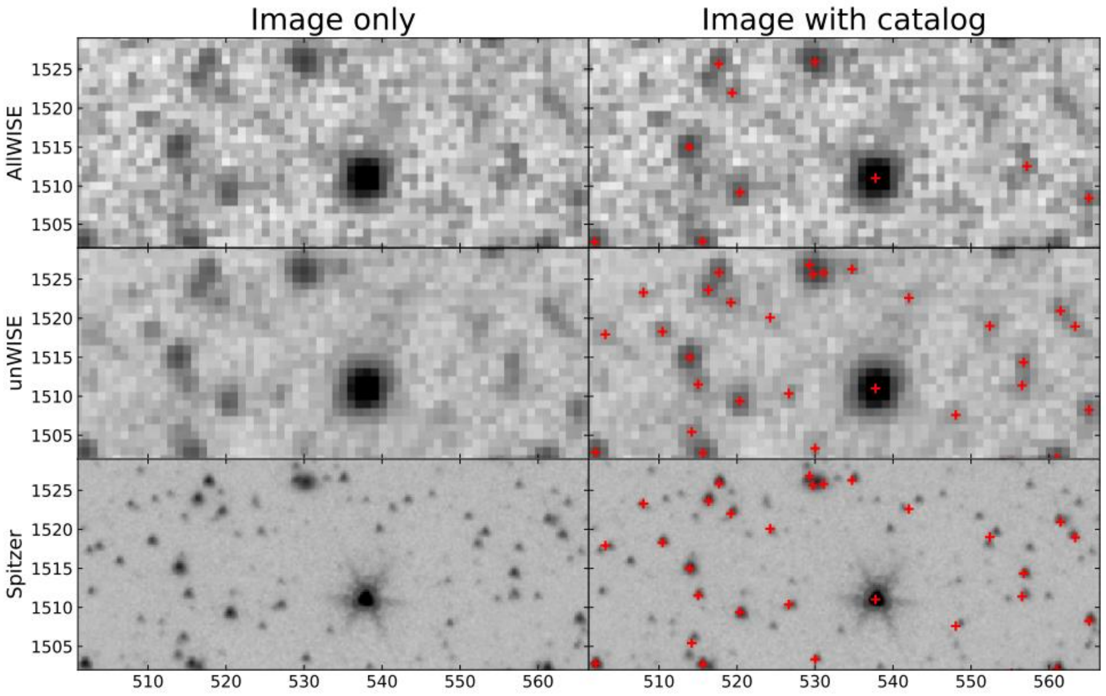
- Estimated fluxes are biased.
- Uncertainties are underestimated.

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.



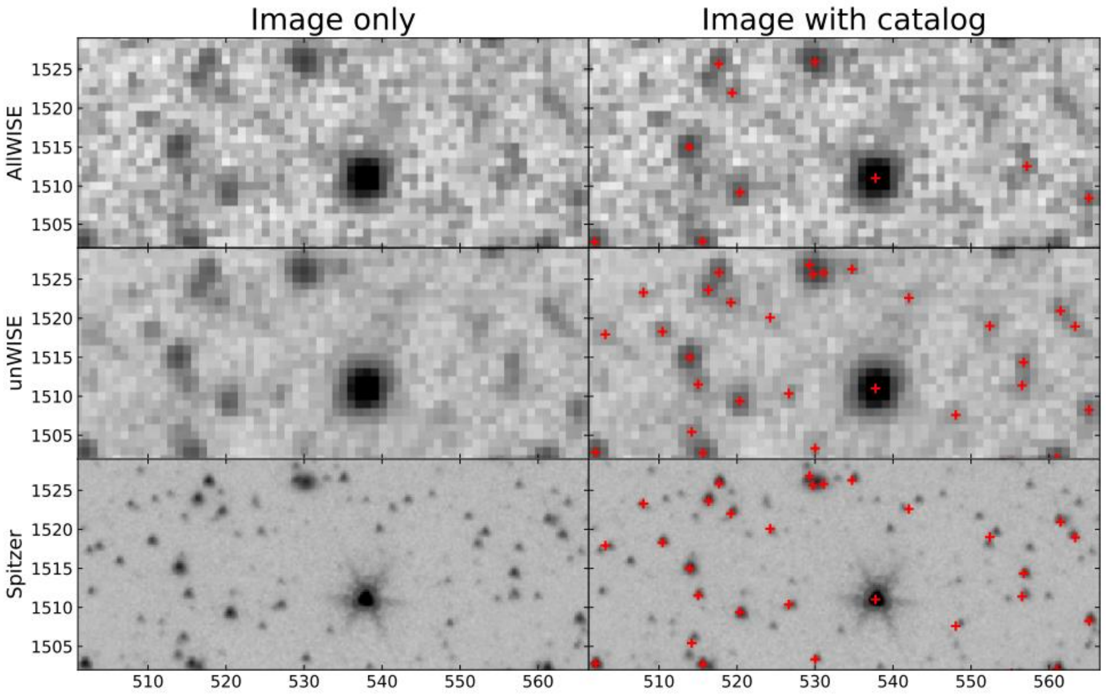
unWISE: Schlafly et al. (2019)

Results

- Estimated fluxes are biased.
- Uncertainties are underestimated.

First reaction:

- N ise: **photo s hard**.
- M t rce)
- Blend
- Bac
- sou
- D on limits/selec effects
- Etc.



unWISE: Schlafly et al. (2019)

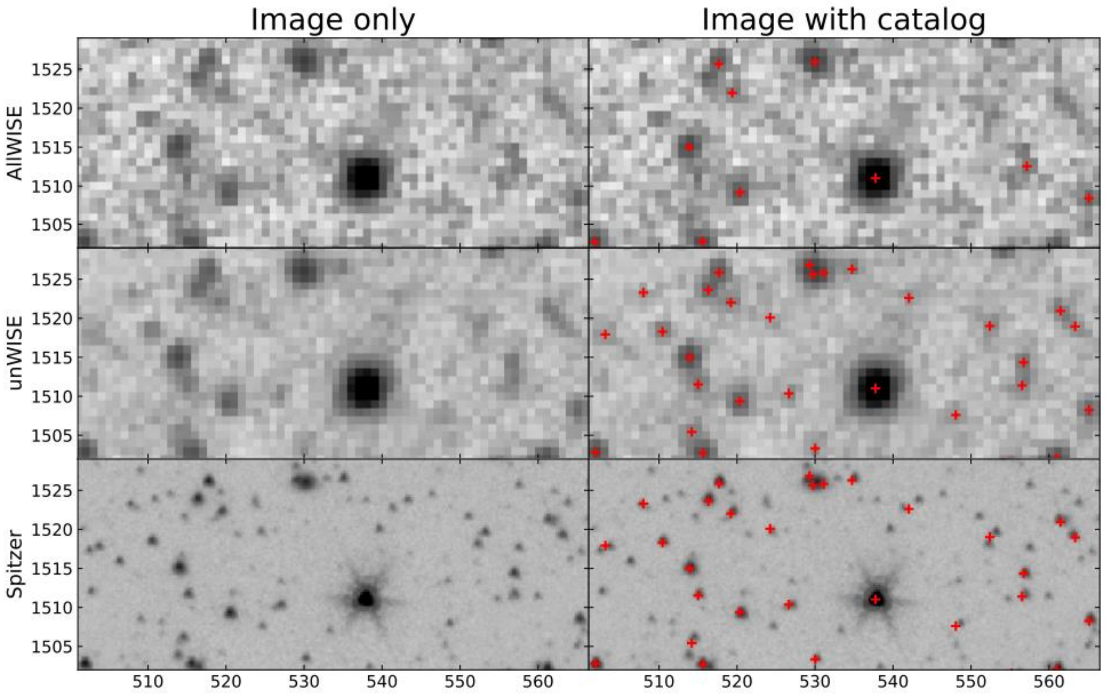
Results

- Estimated fluxes are biased.
- Uncertainties are underestimated.

This is true even with perfect models and data!

First reaction:

- N ise: **photo s hard**.
- M t rce)
- Blend
- Bac
- sou
- D on limits/selec effects
- Etc.



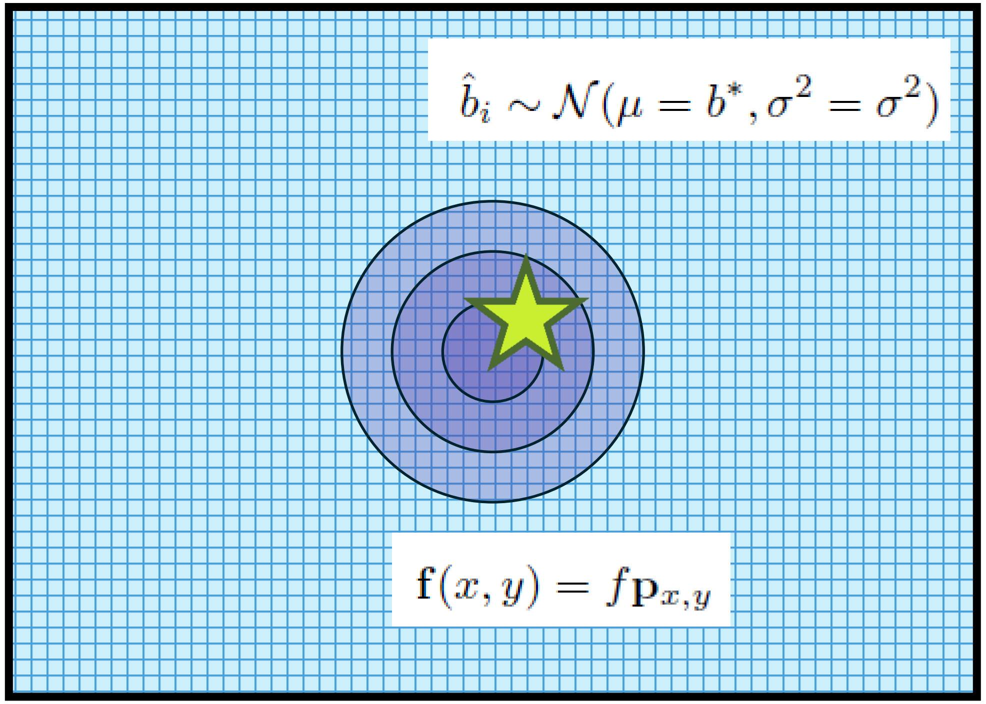
unWISE: Schlafly et al. (2019)

Starting Point

PSF = “point spread function”

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

$n \times m$ footprint

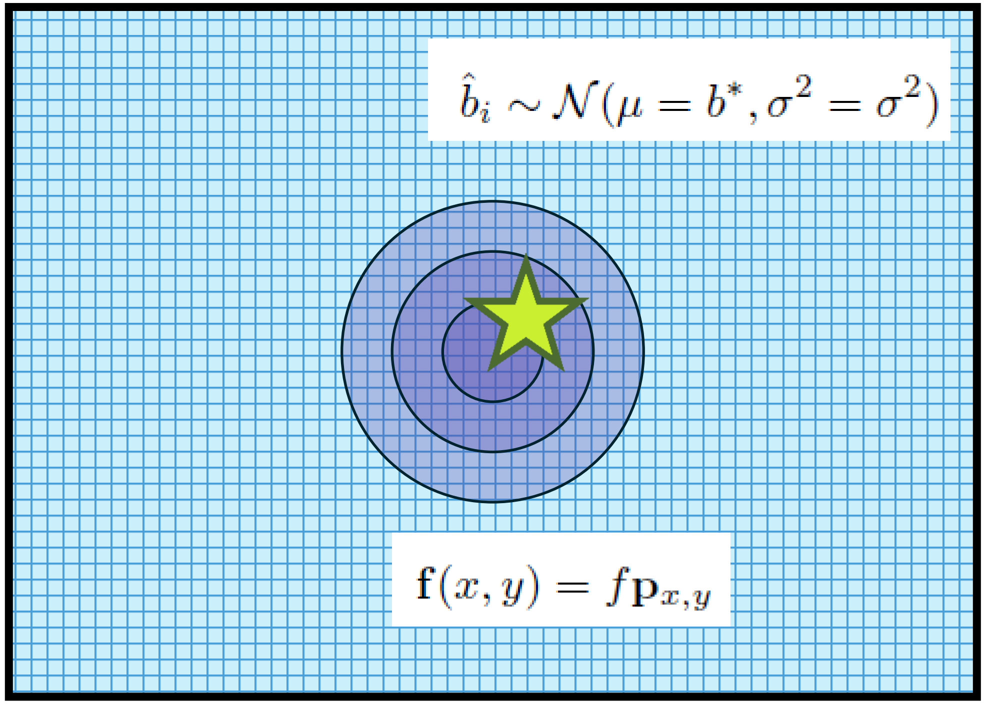


Starting Point

PSF = "point spread function"

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

$n \times m$ footprint



Likelihood

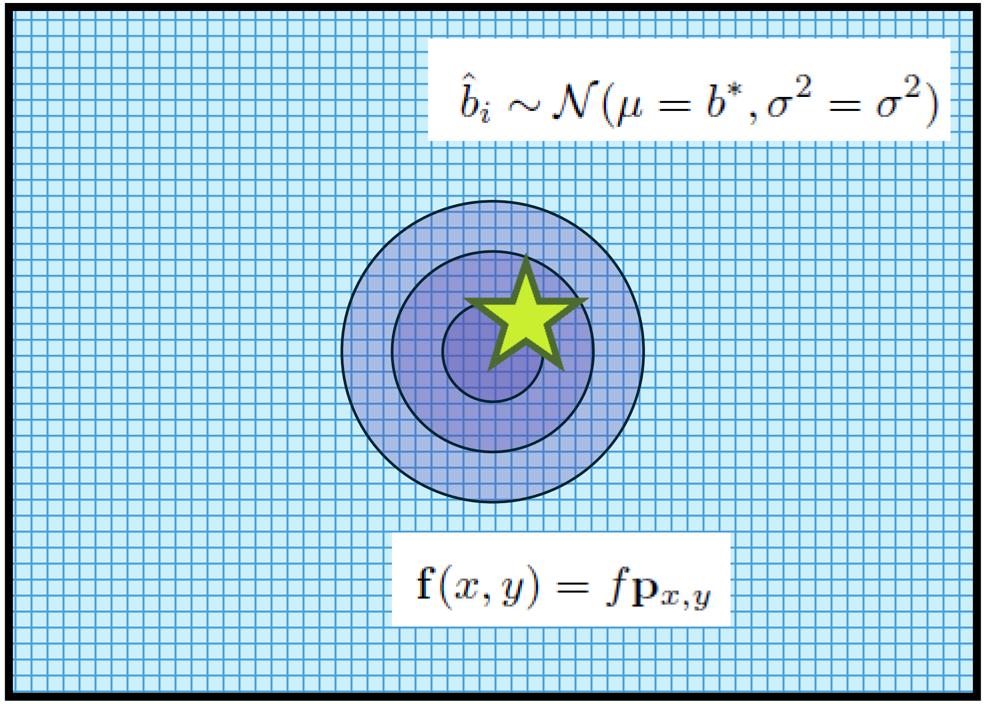
$\ln \mathcal{L}(x, y, f, b)$

Starting Point

PSF = "point spread function"

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

$n \times m$ footprint



Likelihood

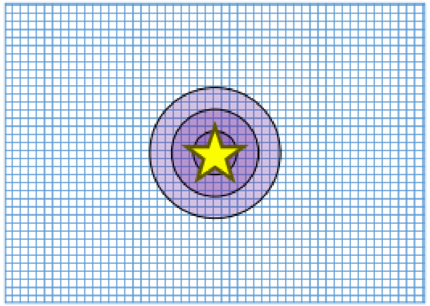
$$\ln \mathcal{L}(x, y, f, b)$$

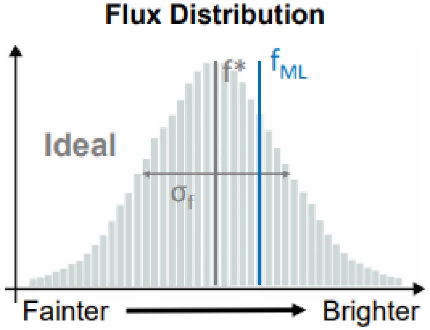
Maximum-Likelihood Solution

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

Position Known

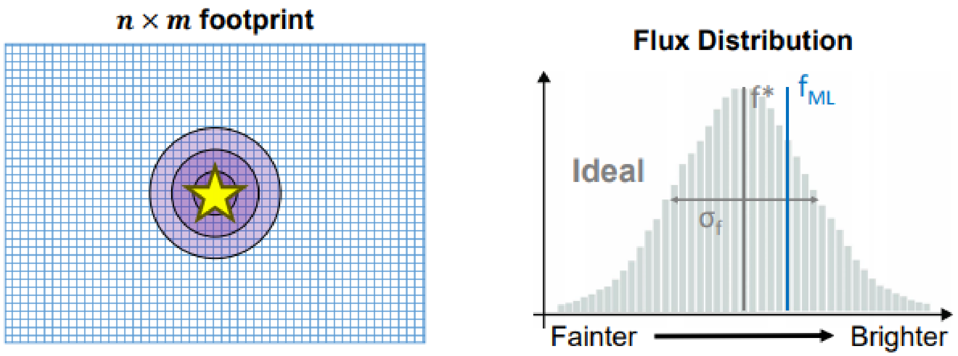
$n \times m$ footprint



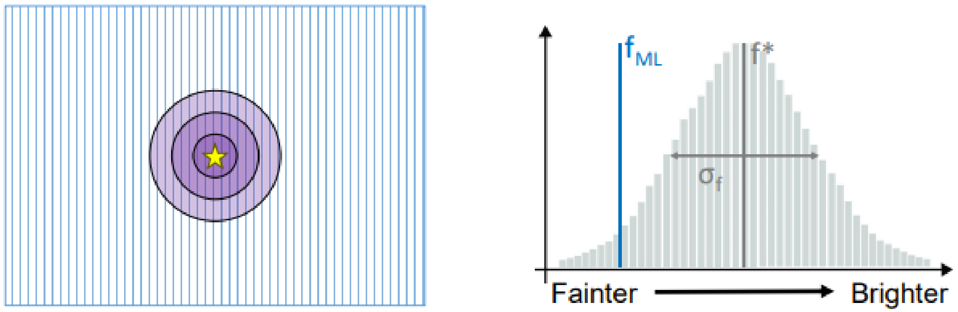


Random noise sometimes leads to high fluctuations.

Position Known

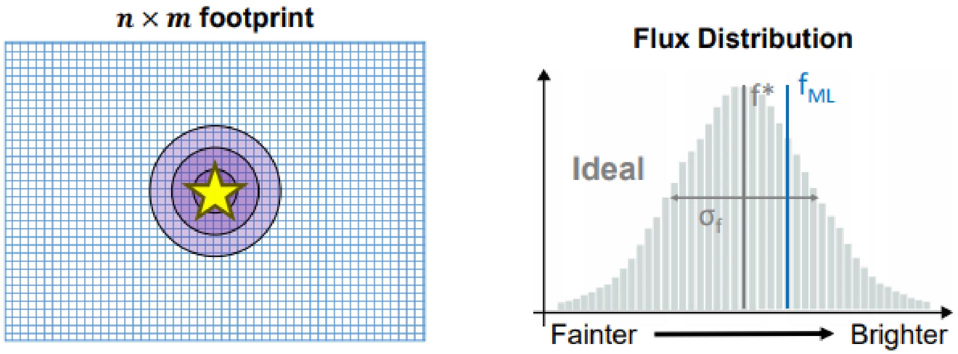


Random noise sometimes leads to high fluctuations.

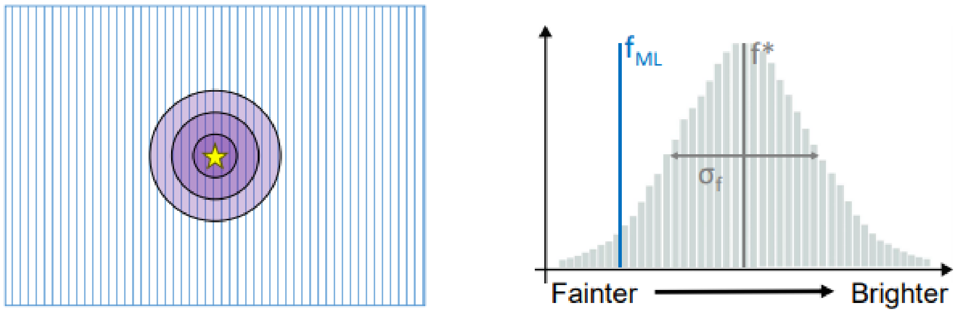


It also leads to low fluctuations with equal probability.

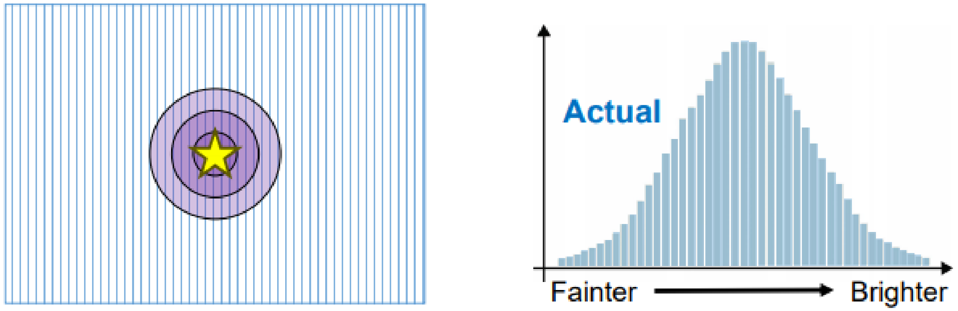
Position **Known**



Random noise sometimes leads to high fluctuations.

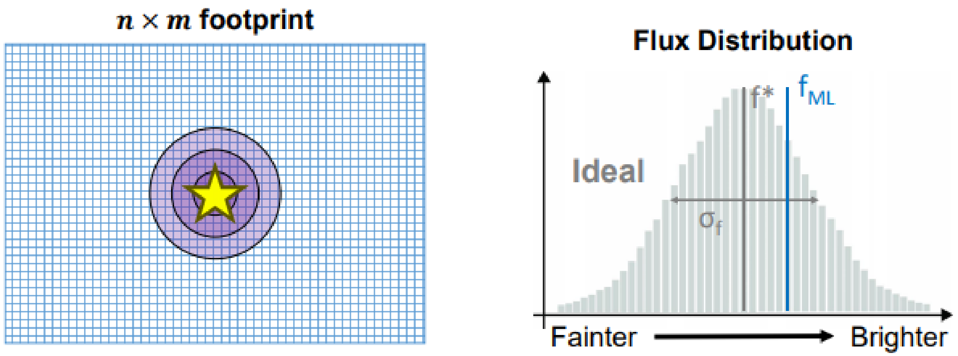


It also leads to low fluctuations with equal probability.

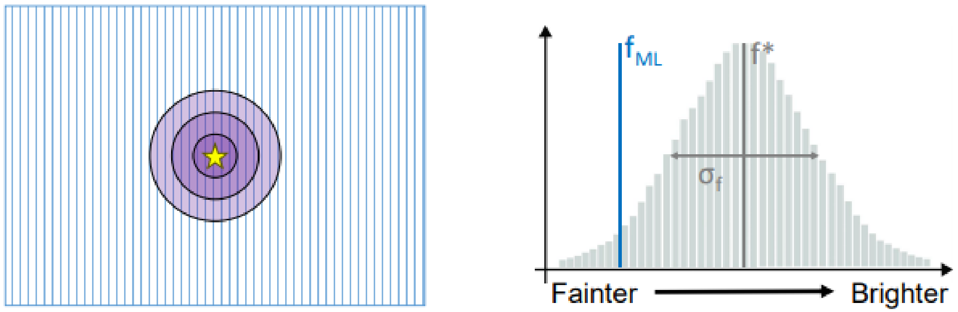


Because fluctuations are symmetric, the maximum-likelihood estimate is unbiased.

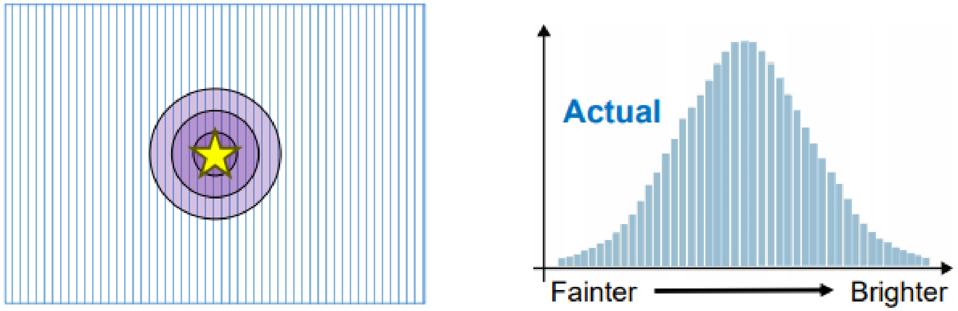
Position Known



Random noise sometimes leads to high fluctuations.

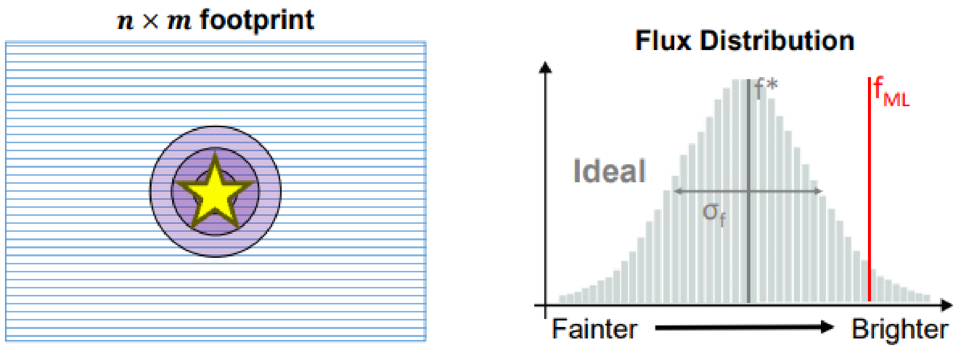


It also leads to low fluctuations with equal probability.

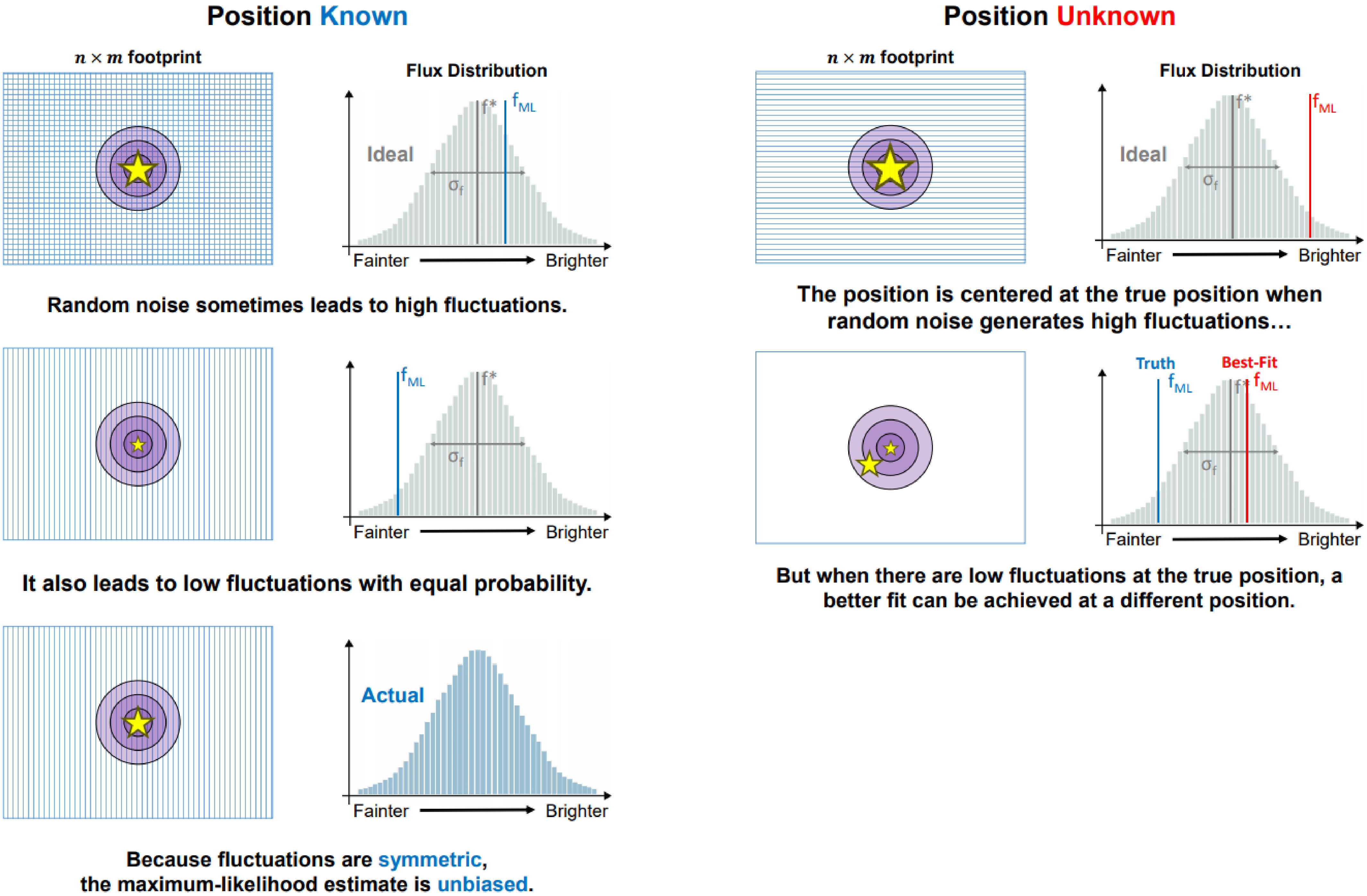


Because fluctuations are symmetric, the maximum-likelihood estimate is unbiased.

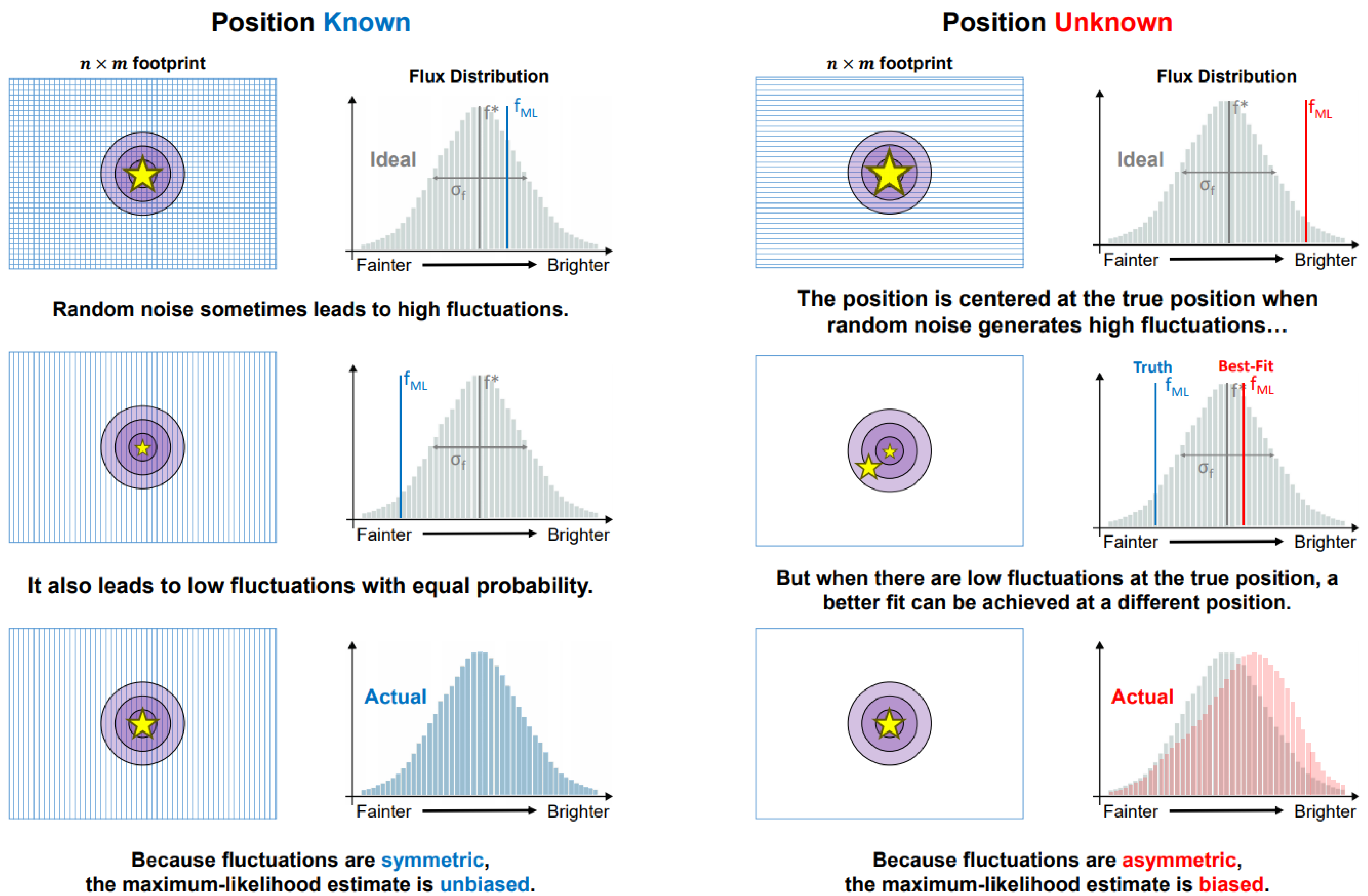
Position Unknown



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



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


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

Biases in PSF Photometry

- Under basic assumptions we can derive a 1st-order **bias**:

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

Biases in PSF Photometry

- Under basic assumptions we can derive a 1st-order **bias**:

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

SNR^{-2}

- Applying this correction increases the **variance**:

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

Biases in PSF Photometry

- Under basic assumptions we can derive a 1st-order **bias**:

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

SNR^{-2}

- Applying this correction increases the **variance**:

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

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

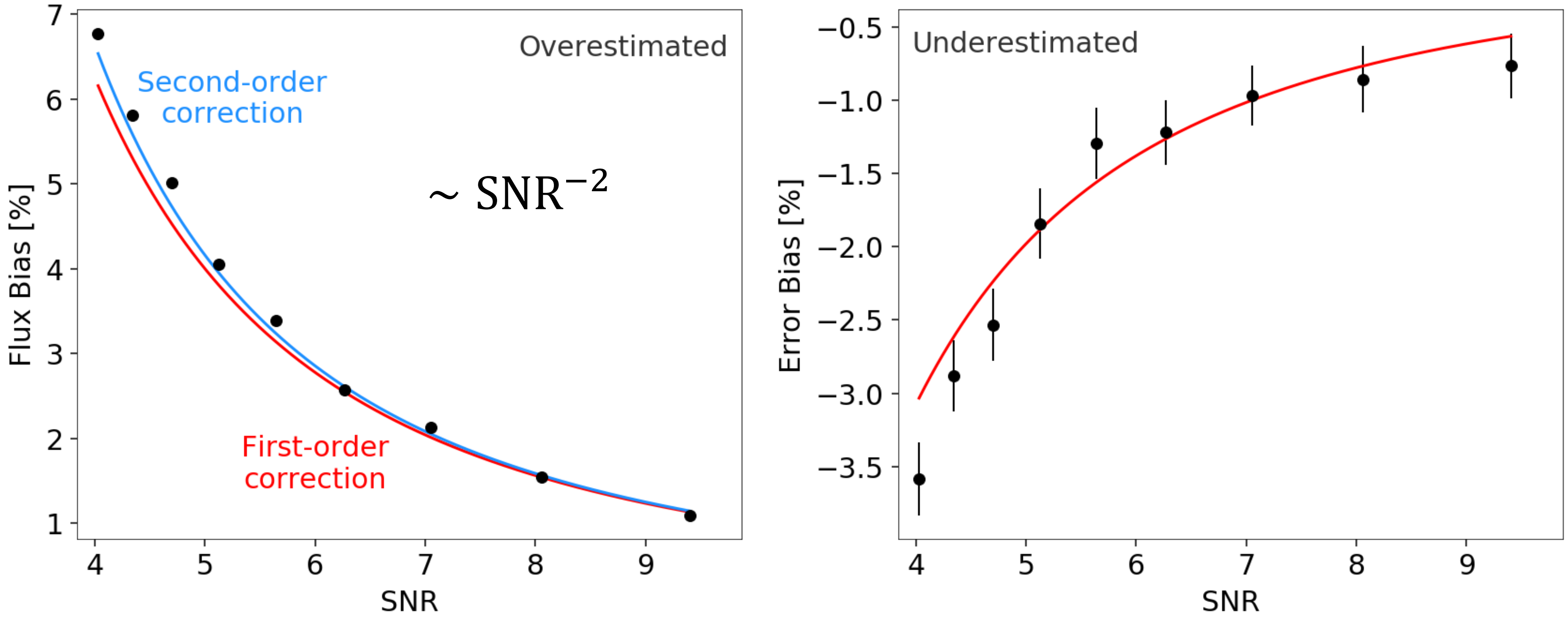
10-sigma source:

- Flux: +1% bias.

Biases in PSF Photometry

- Position **unknown**, background known

Background fixed



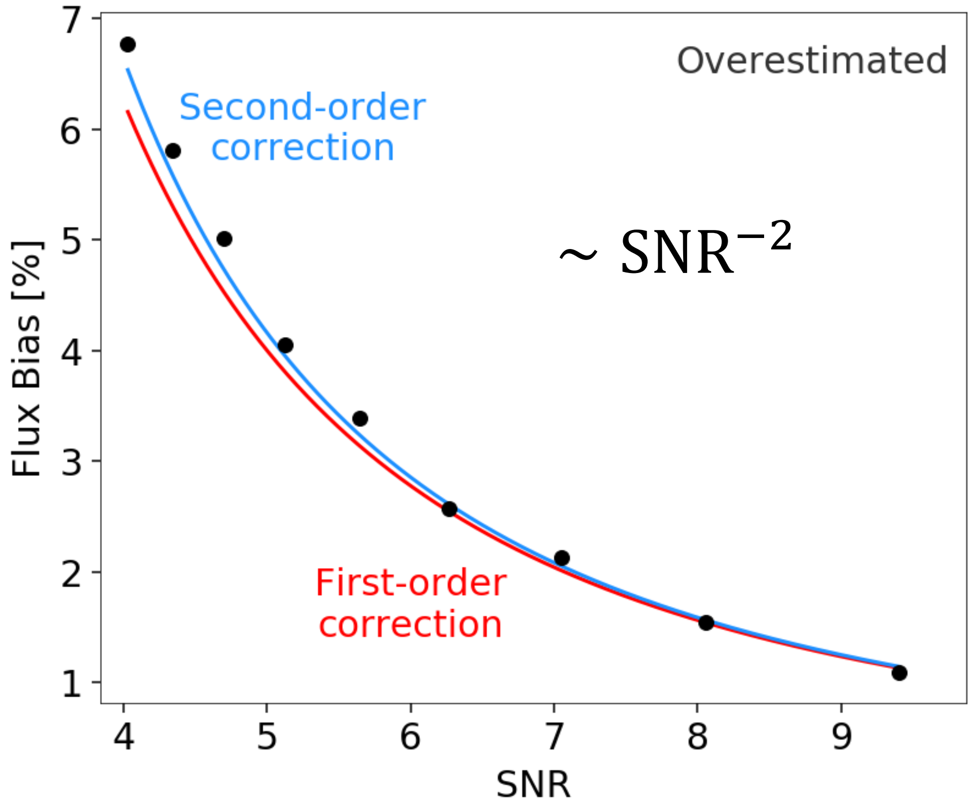
10-sigma source:

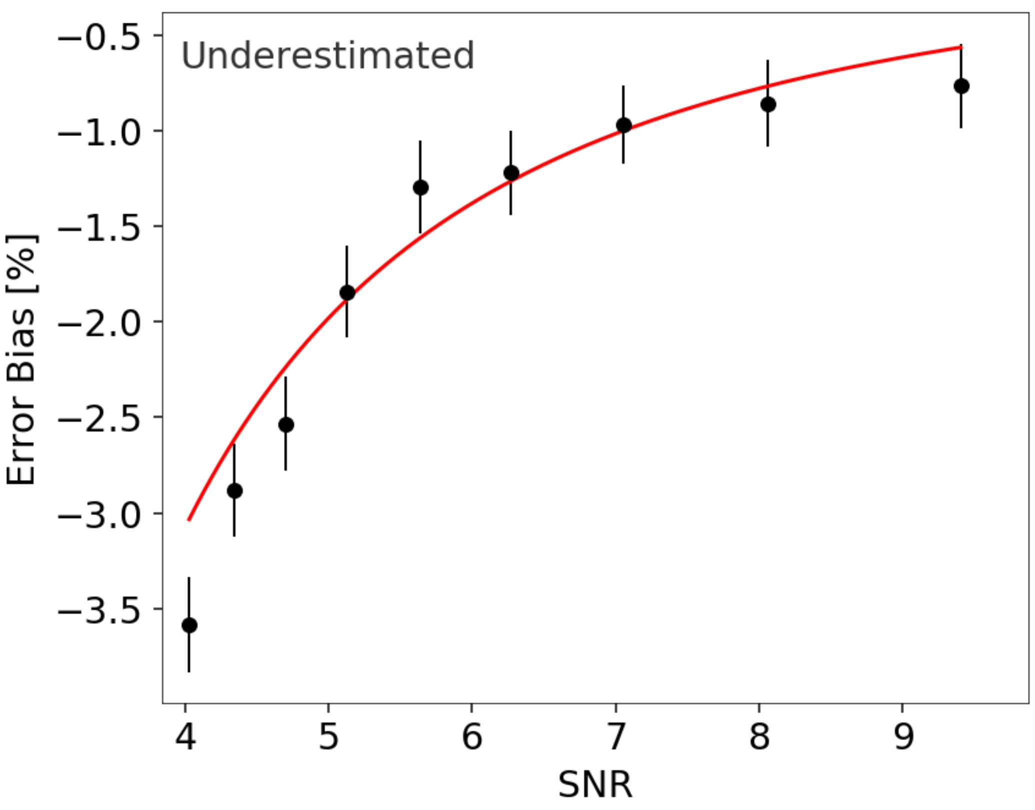
- Flux: +1% bias.

Biases in PSF Photometry

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

Background fixed



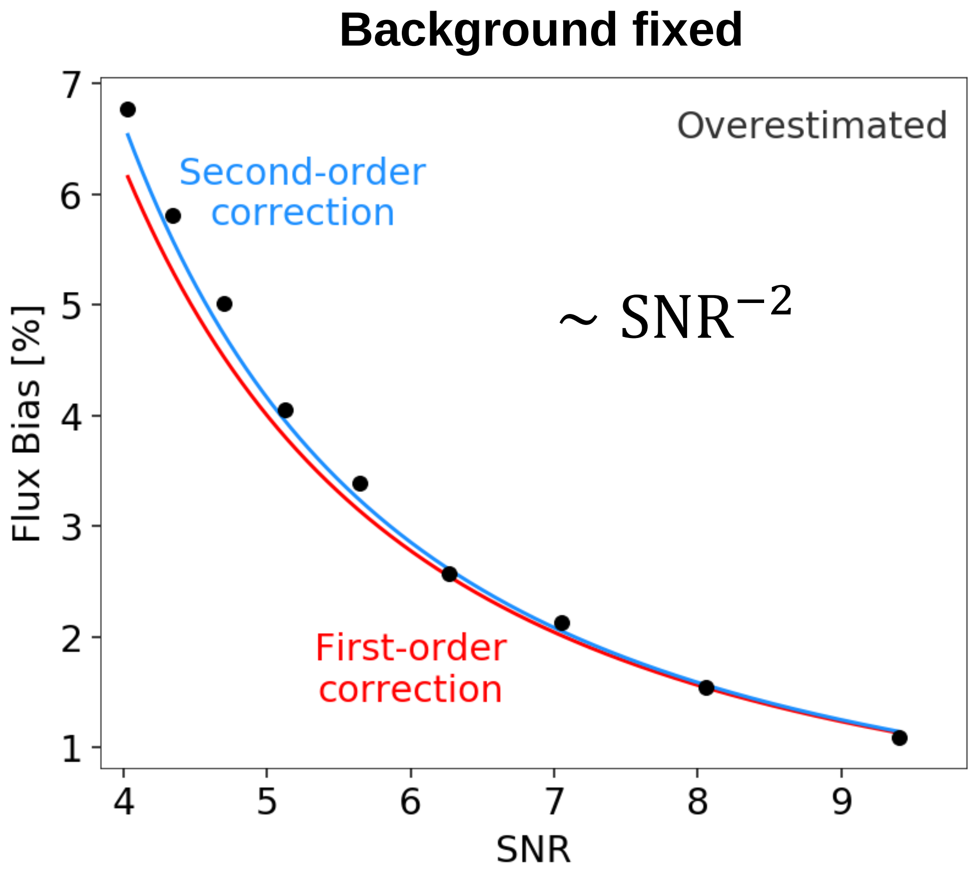


10-sigma source:

- Flux: +1% bias.

Biases in PSF Photometry

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



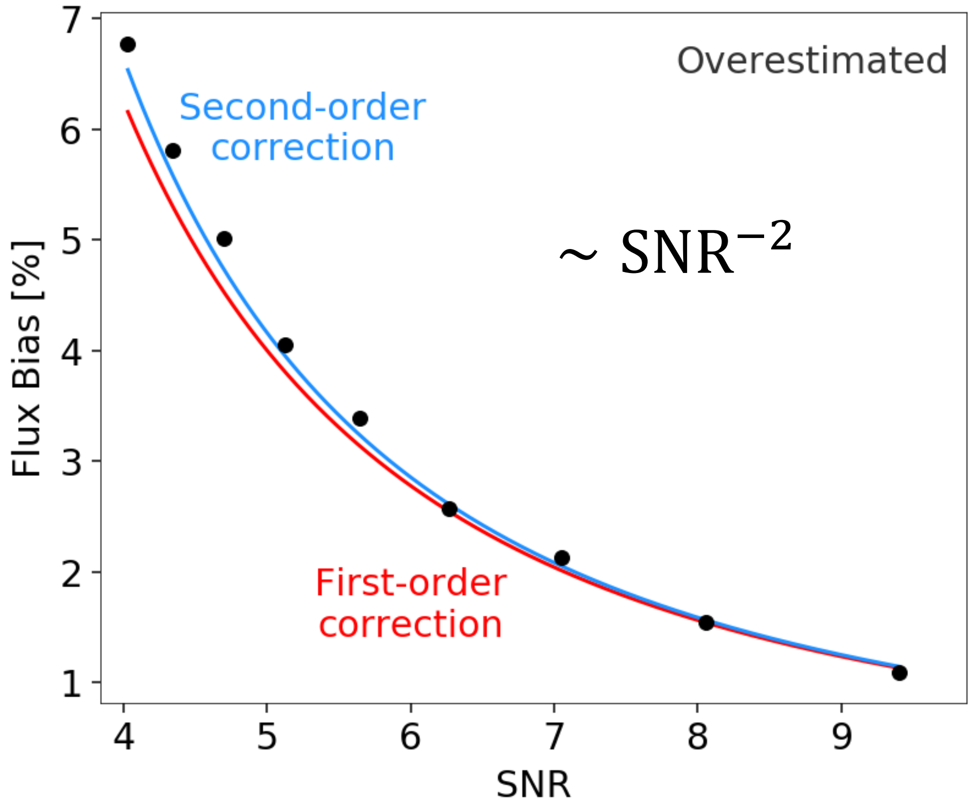
- Need to account for covariance between background and other parameters.

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

Biases in PSF Photometry

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

Background fixed



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

Area used to estimate background

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

Effective PSF area

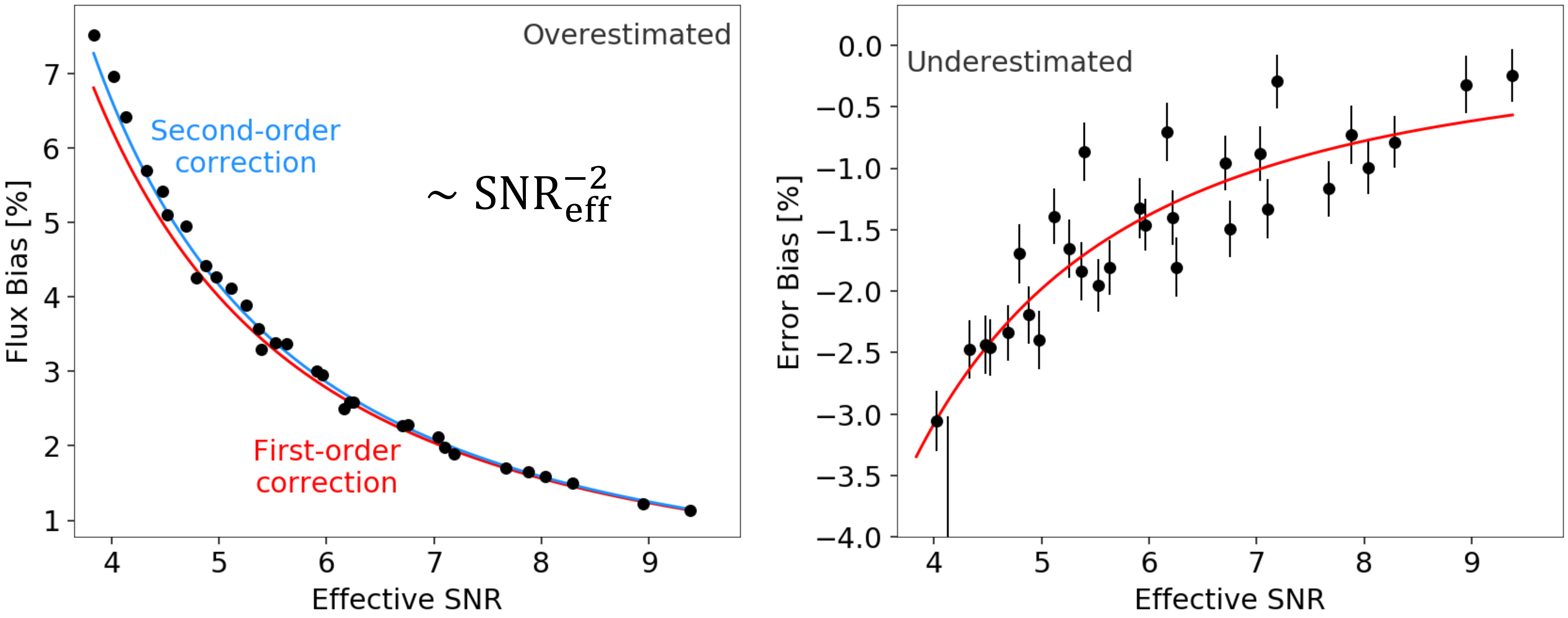
10-sigma source:

- **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.

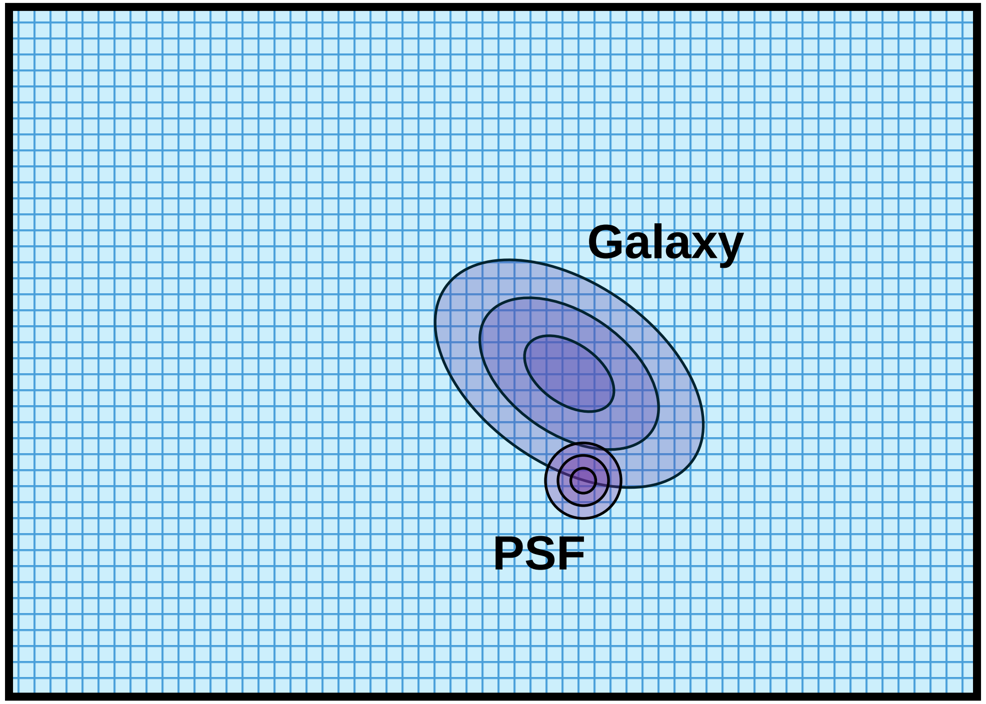
Error Bias

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

Ignoring covariances underestimates errors.

Biases in Extended Source Photometry

$n \times m$ footprint



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

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

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

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

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

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

Extra degrees of freedom allows fit to chase noise.

$p = 3 \rightarrow 6 - 10$

Shape parameters soak up noise.

Error Bias

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

Ignoring covariances underestimates errors.

Shape parameters add covariances.

Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area

Flux Bias

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

Extra degrees of freedom allows fit to chase noise.

$p = 3 \rightarrow 6 - 10$

Shape parameters soak up noise.

Error Bias

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

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

$A_{\rm gal+psf}$

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

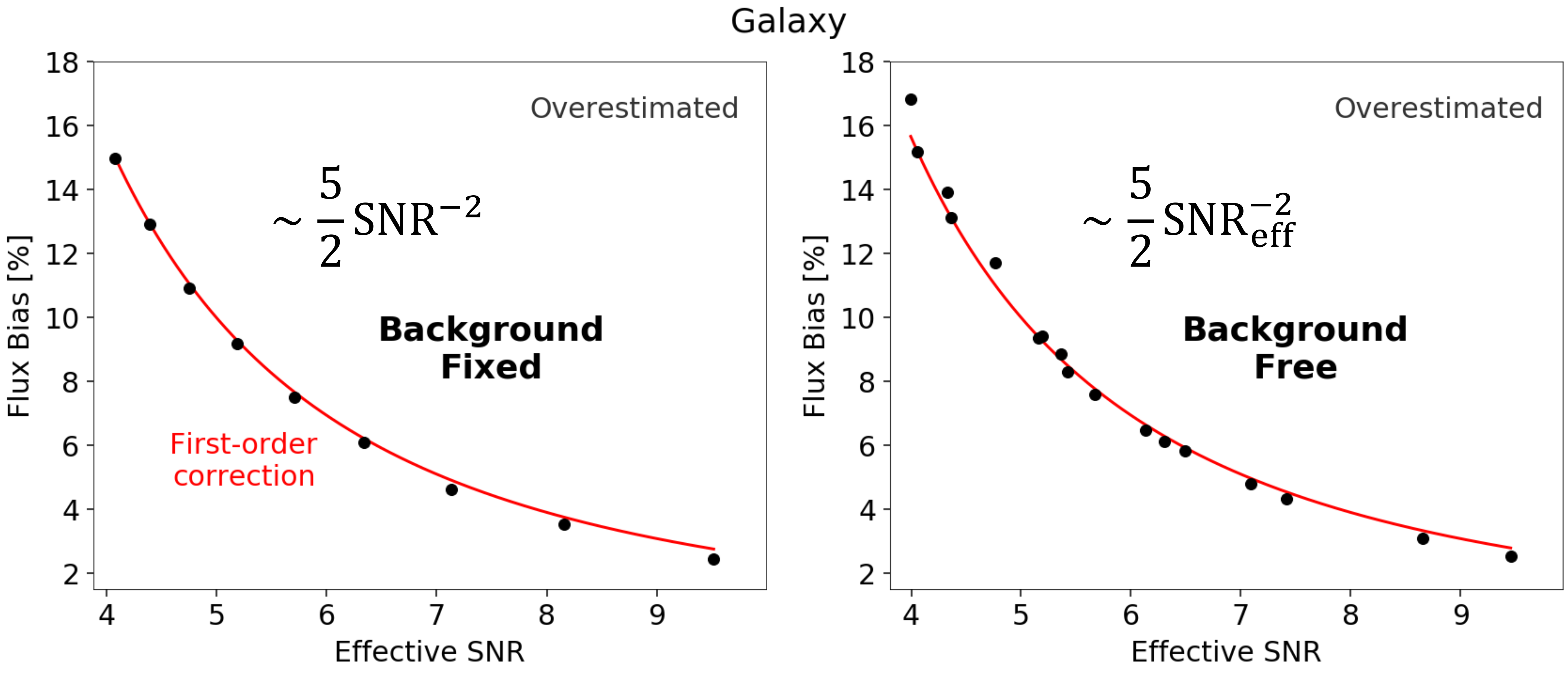
Ignoring covariances underestimates errors.

Shape parameters add covariances.

Extended shape impedes background estimation.

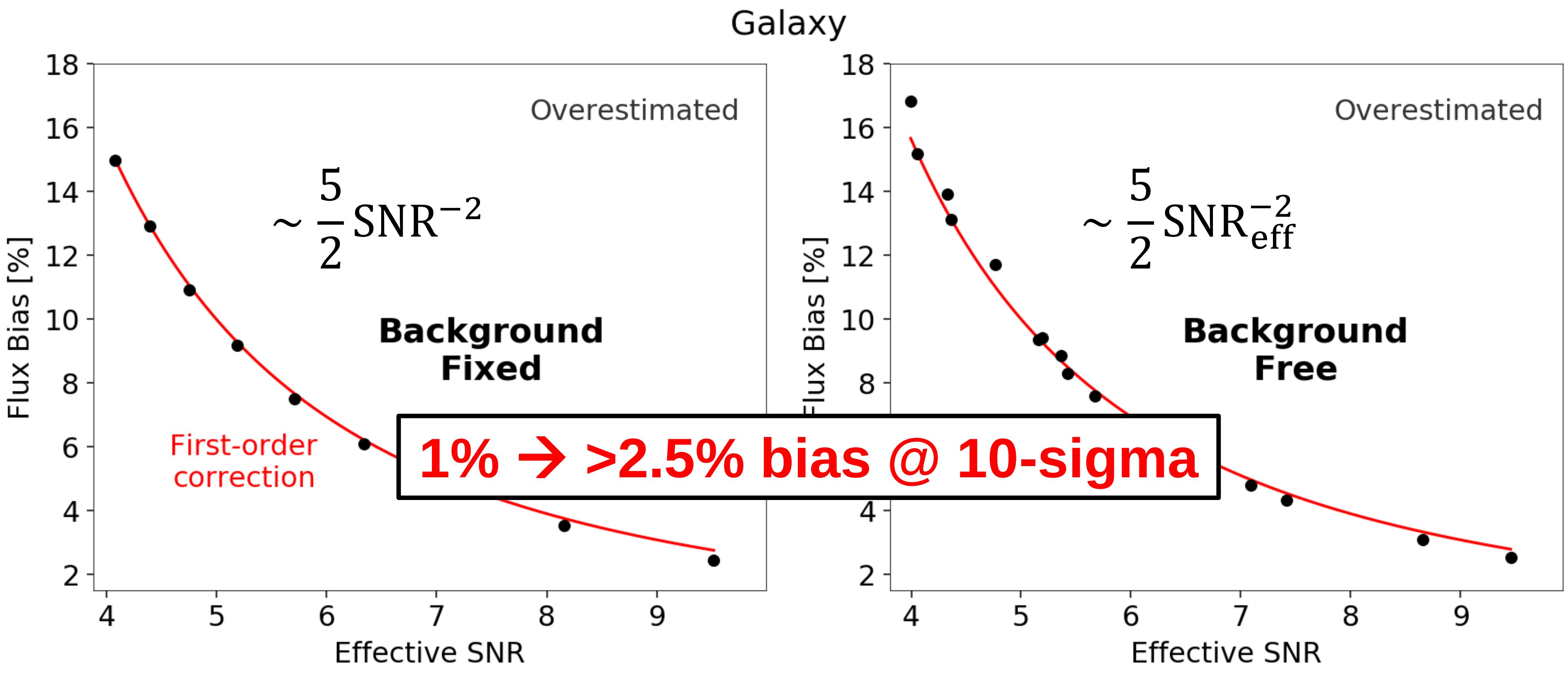
Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area

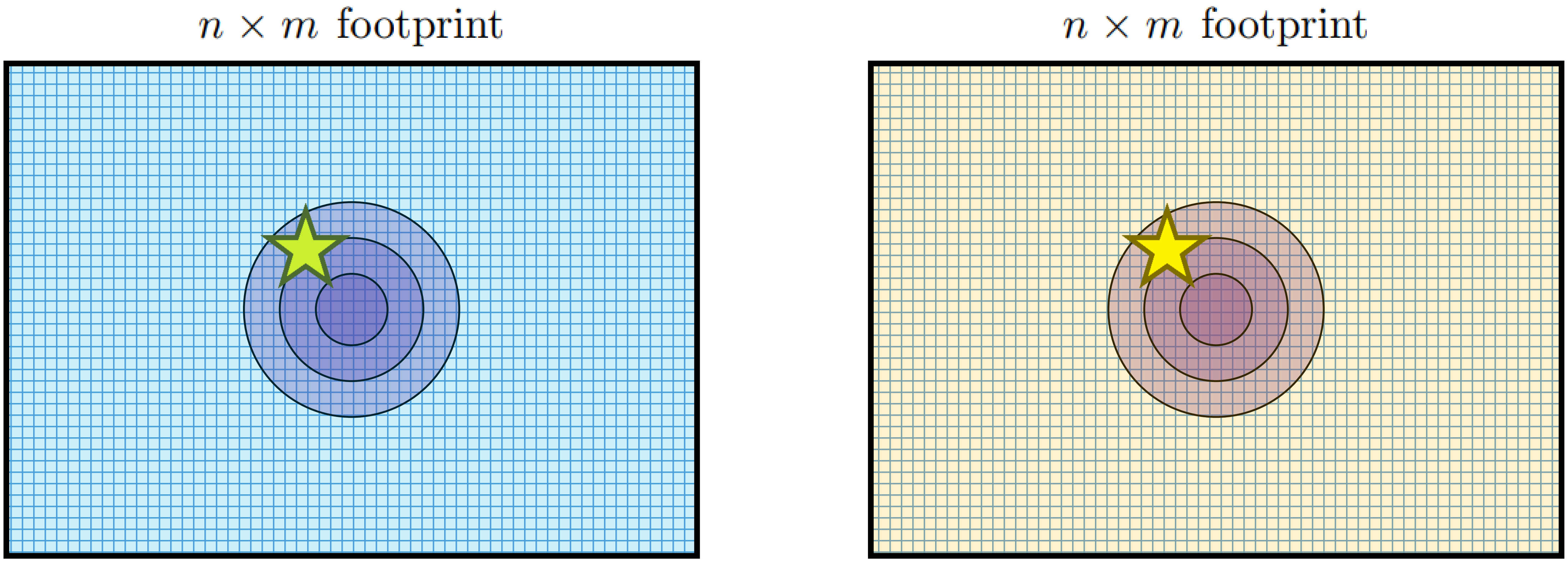


Biases in Extended Source Photometry

- More parameters, more covariances, larger effective area



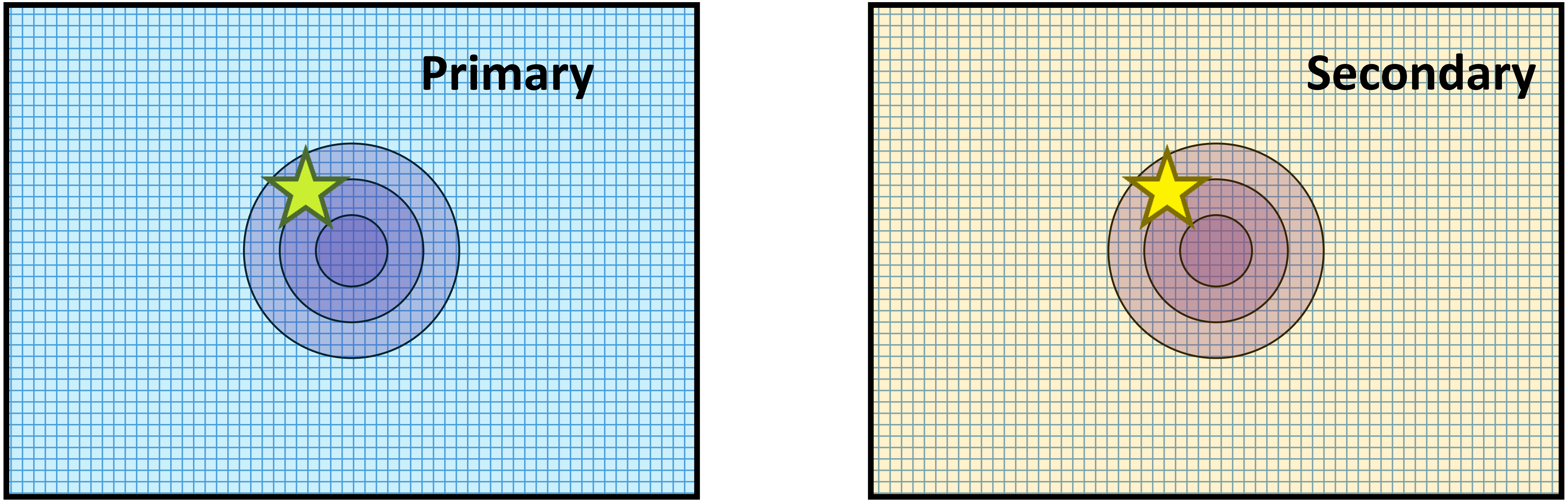
Biases in Multi-band Photometry



Biases in Forced Multi-band Photometry

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

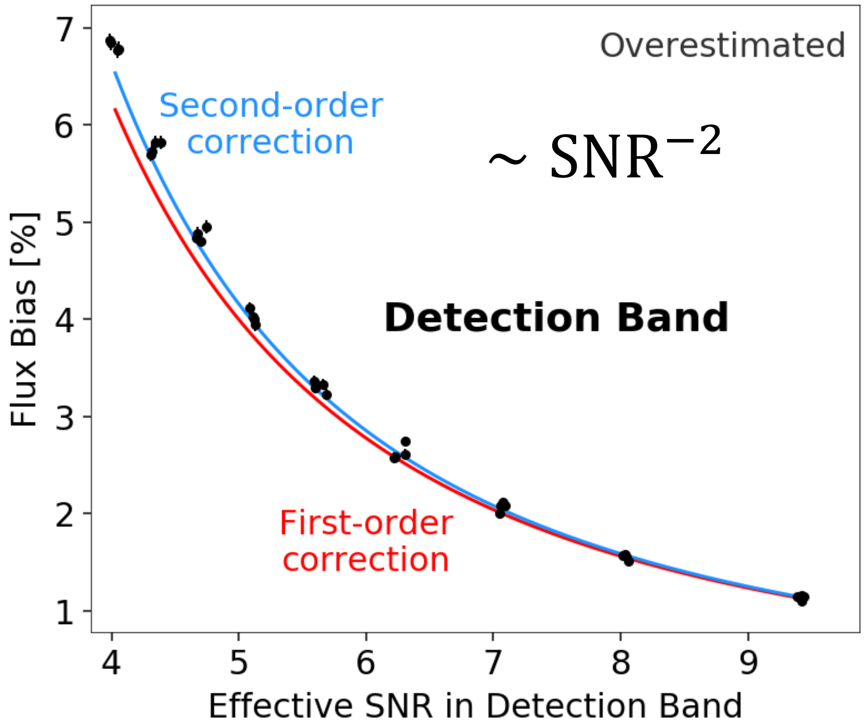
$n \times m$ footprint $\qquad$ $n \times m$ footprint



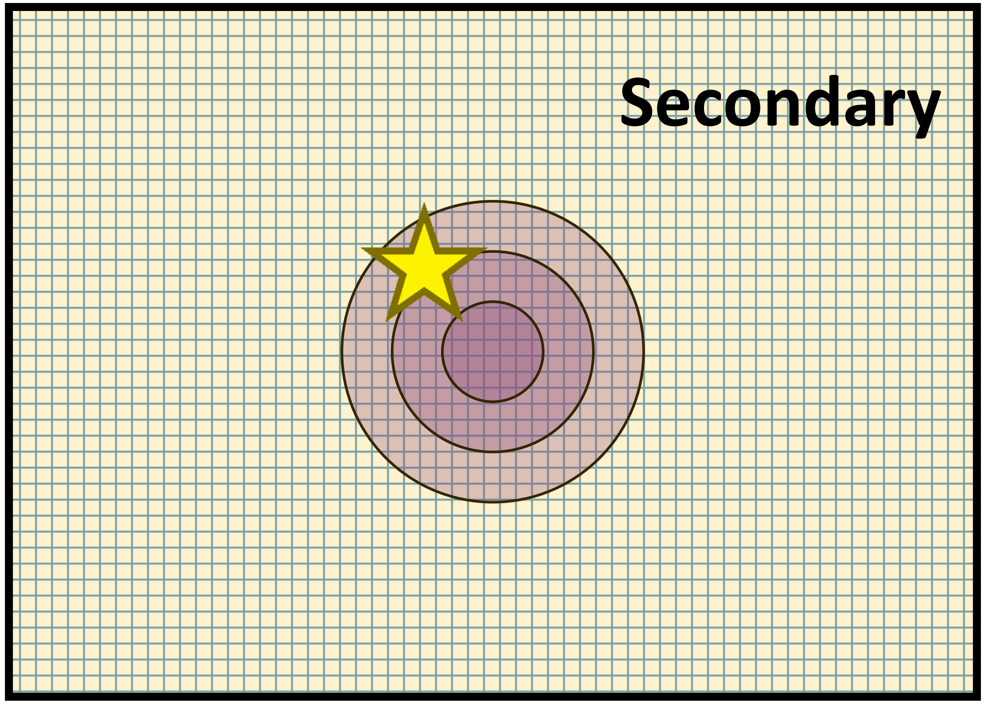
Biases in Forced Multi-band Photometry

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

$n \times m$ footprint



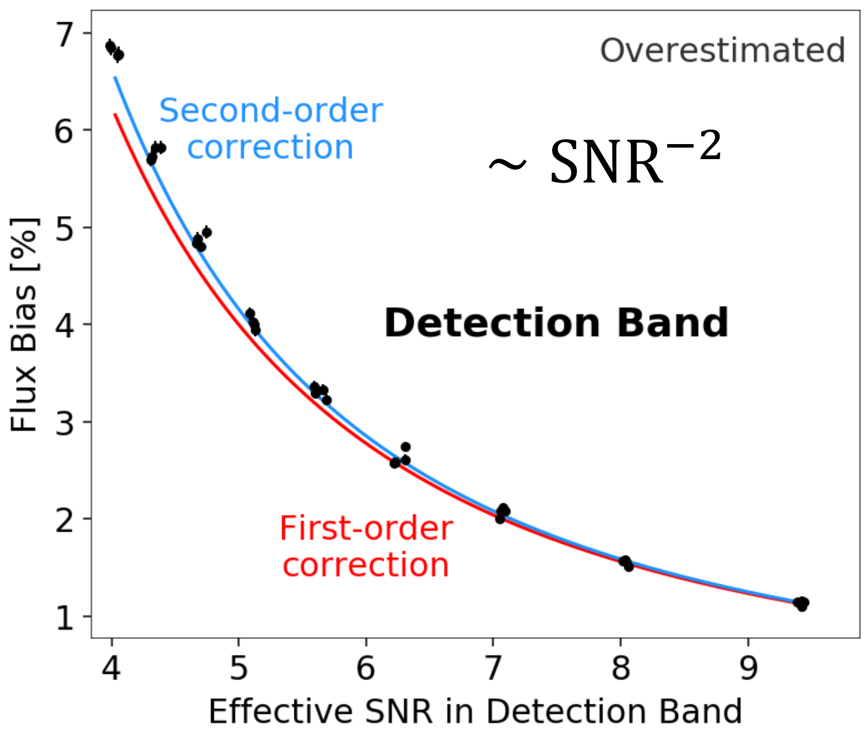
$n \times m$ footprint



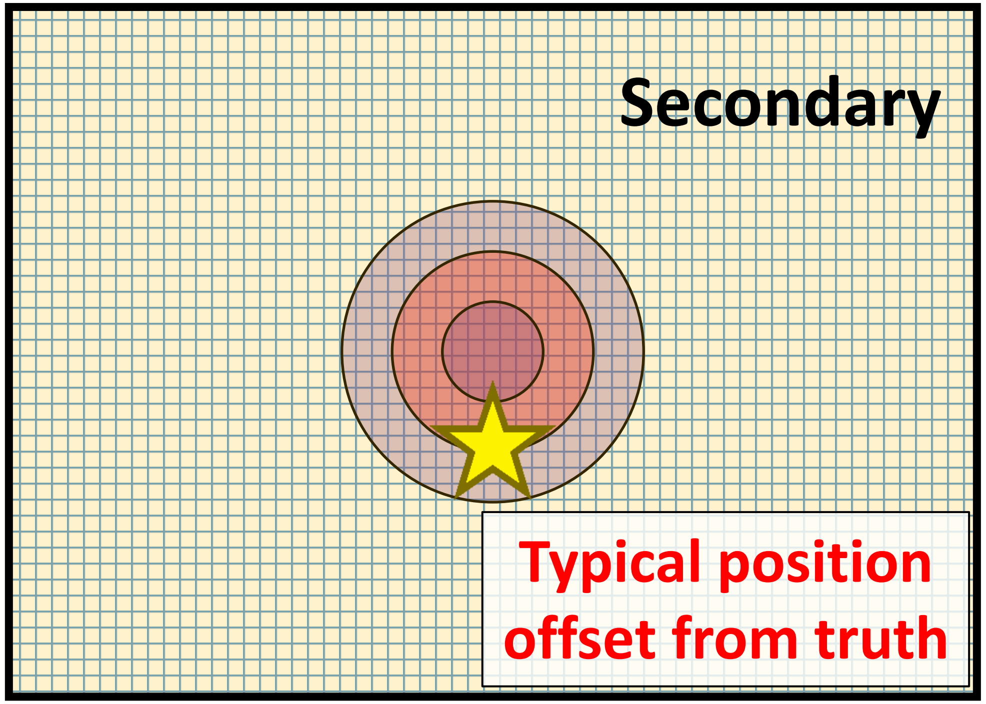
Biases in Forced Multi-band Photometry

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

$n \times m$ footprint



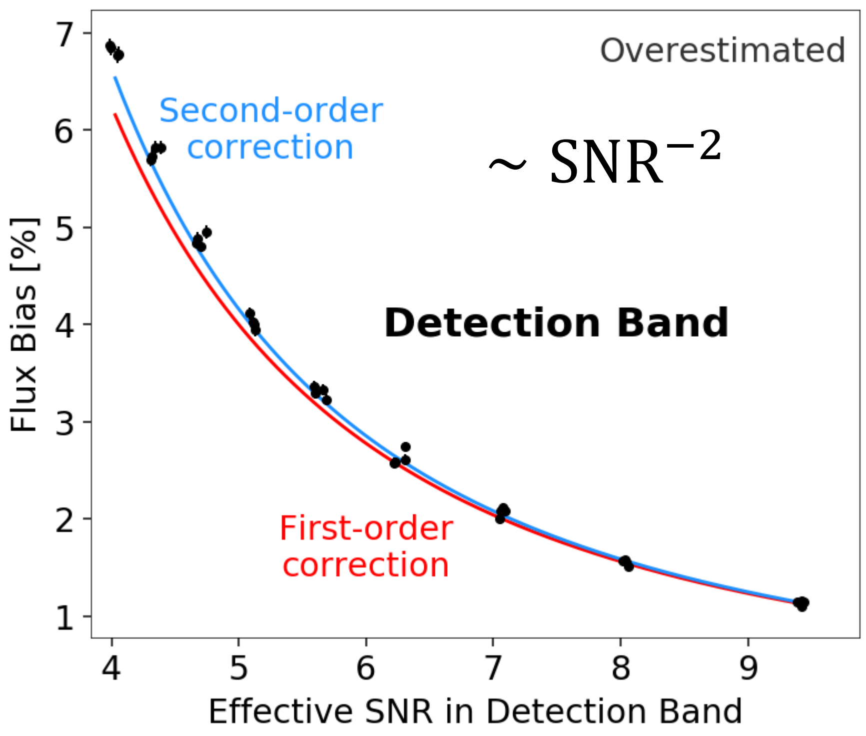
$n \times m$ footprint



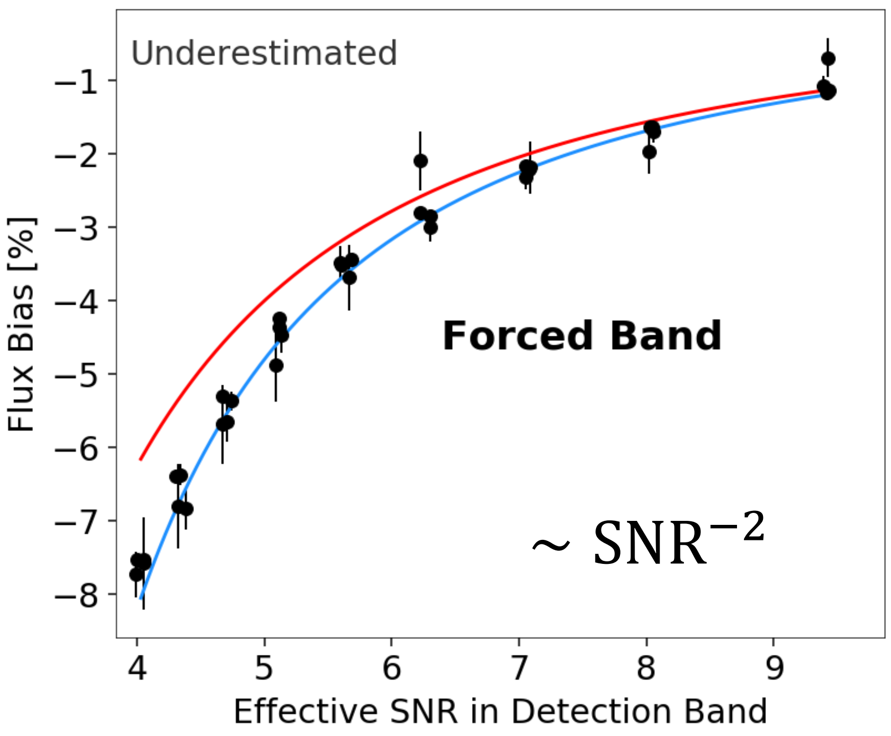
Biases in Forced Multi-band Photometry

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

$n \times m$ footprint

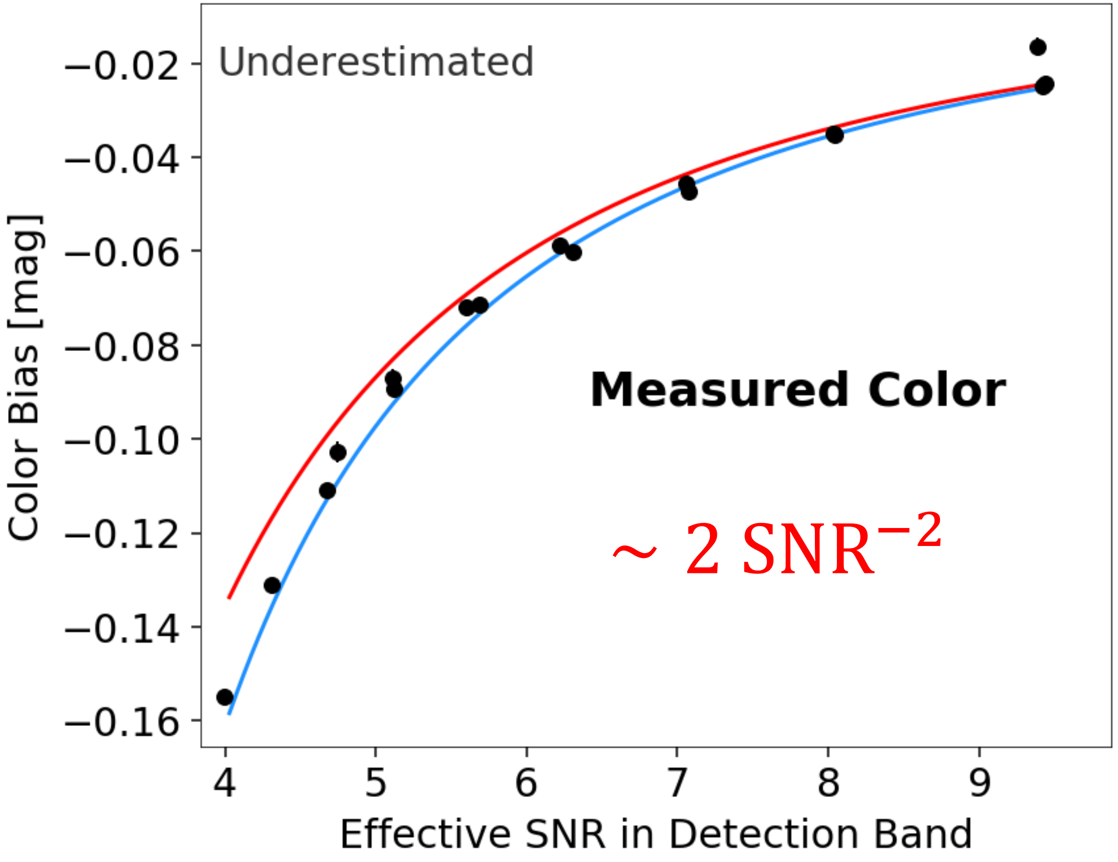


$n \times m$ footprint



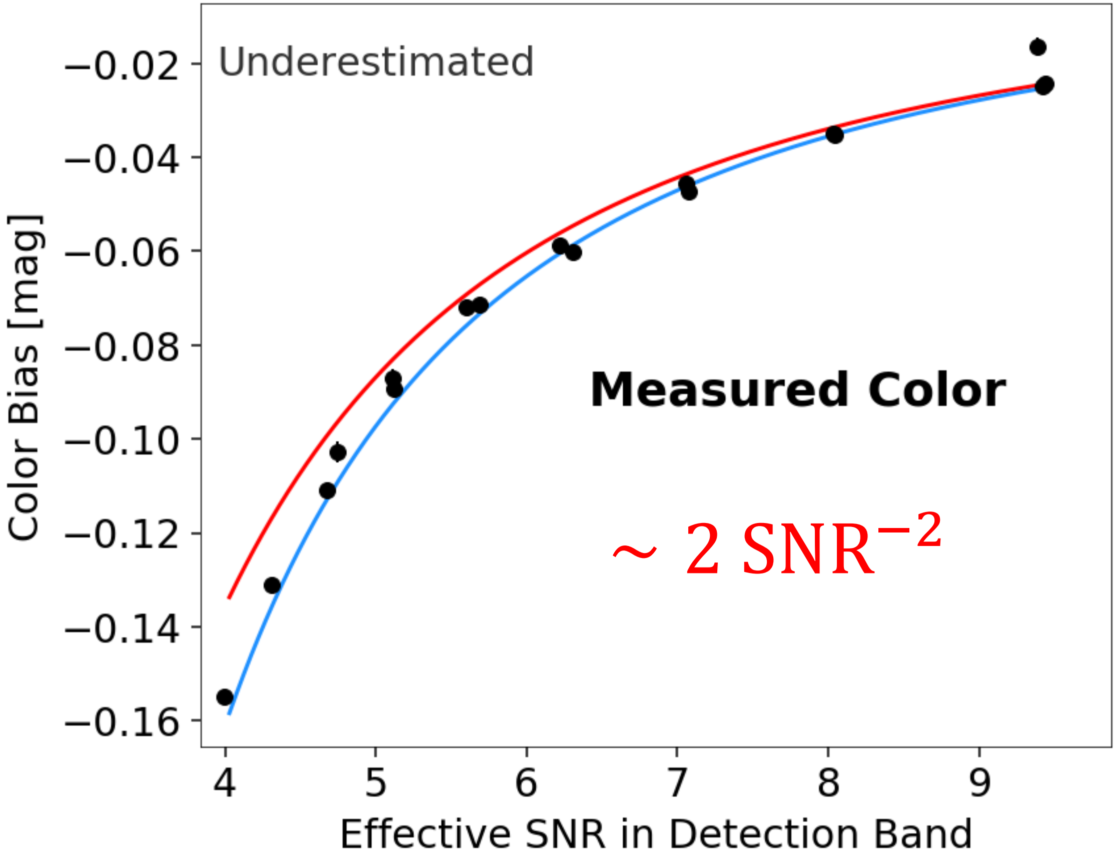
Biases in Forced Multi-band Photometry

- 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



Star @ 10-sigma:

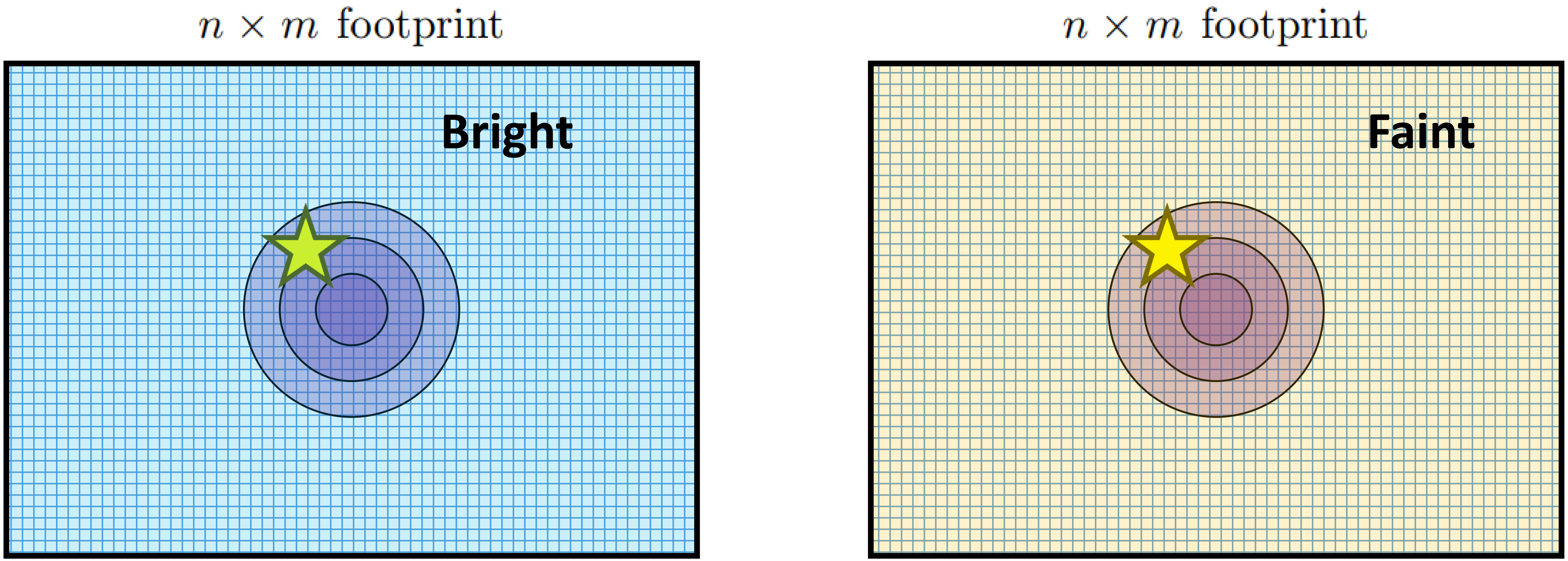
- 1% flux bias
- 0.02 mag color bias.

Galaxy @ 10-sigma:

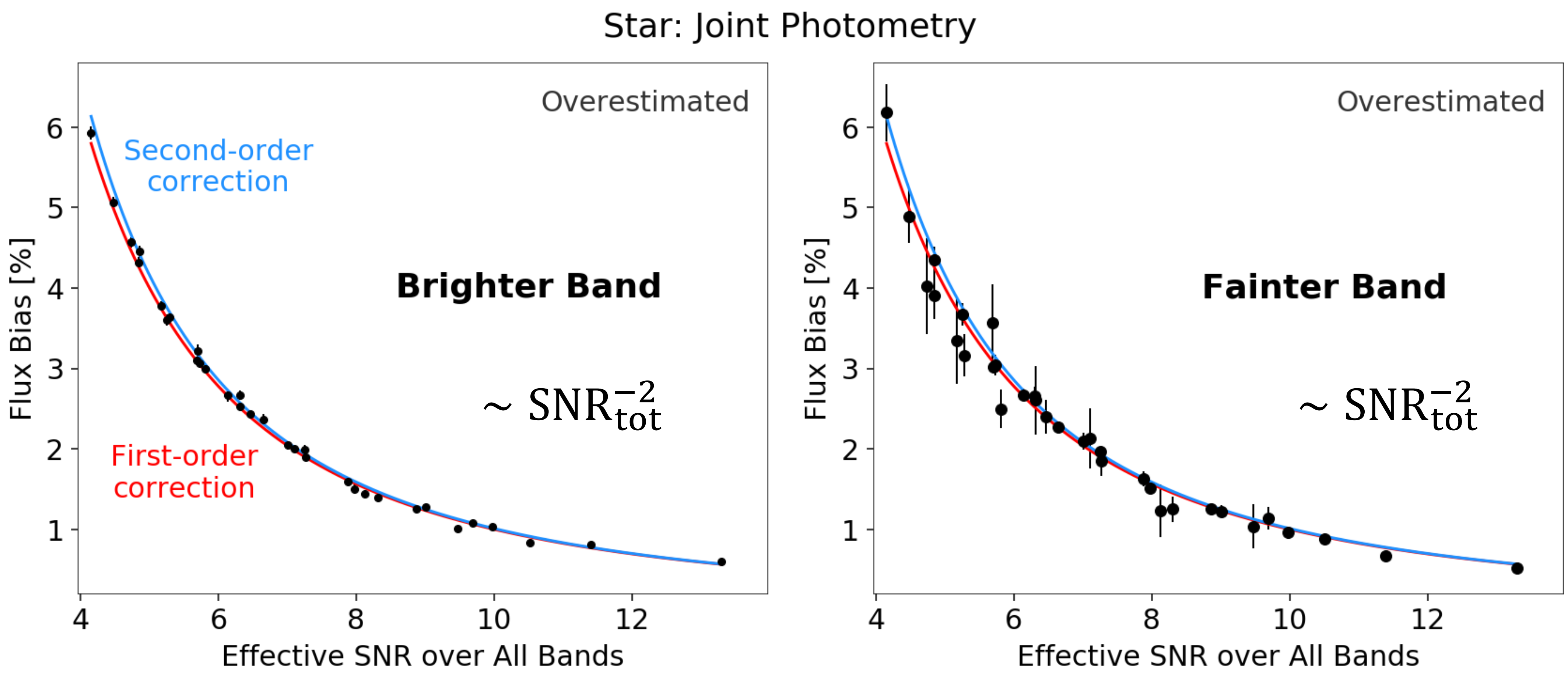
- 2.5% flux bias
- **0.05 mag color bias.**

Biases in Joint Multi-band Photometry

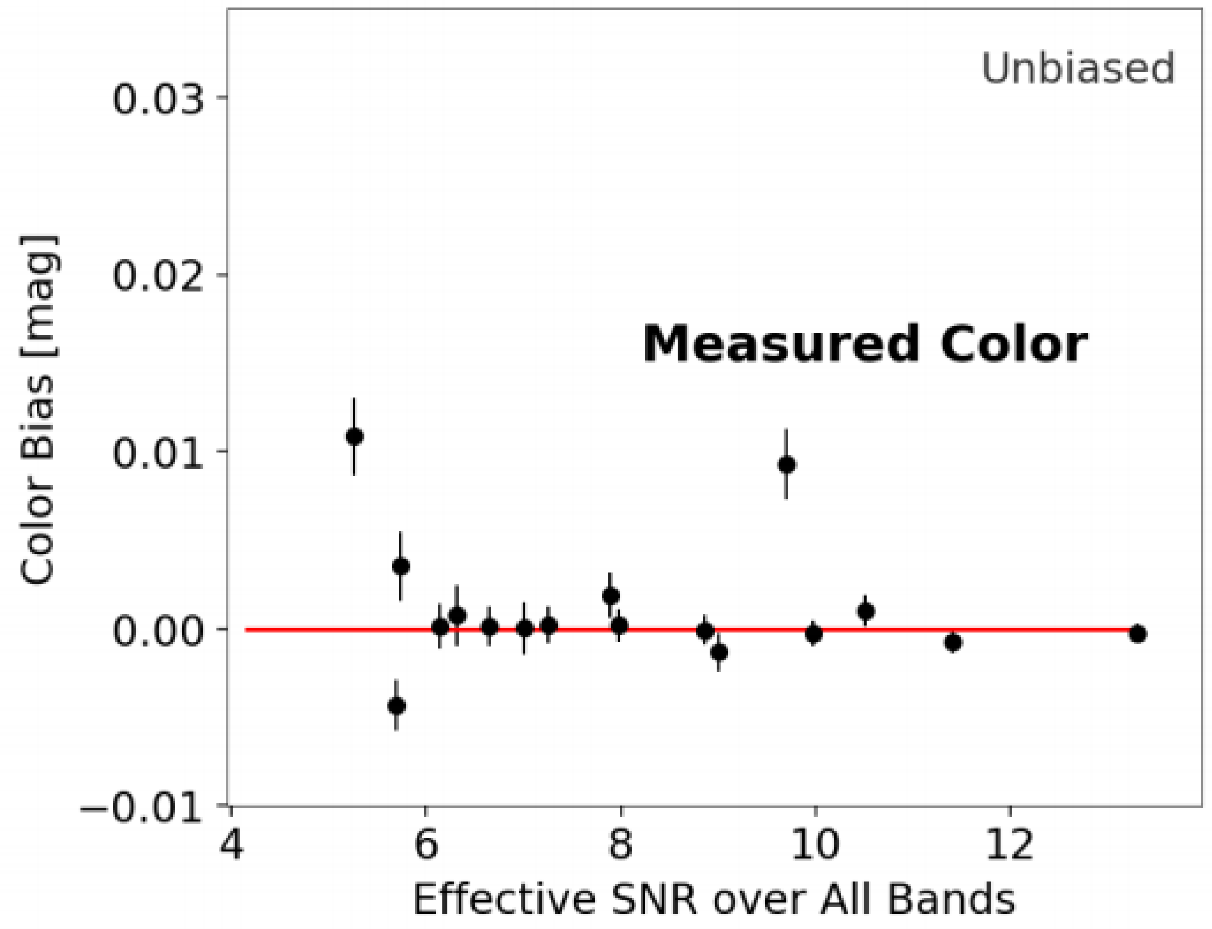
- Fit all bands simultaneously (~ detect on stack, force in bands)



Biases in Joint Multi-band Photometry



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?

SDSS Stripe 82

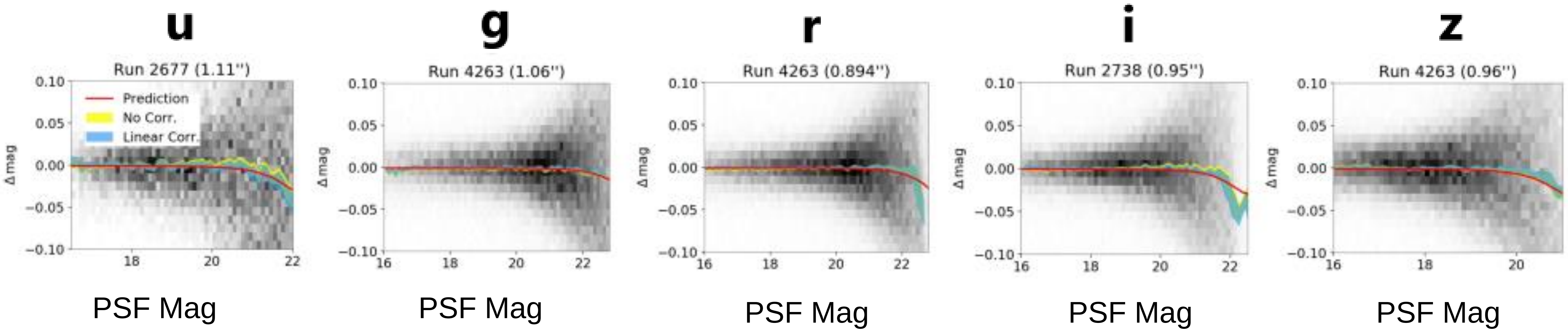
- **Repeated imaging**: compare catalog computed from a **“deep stack”** of all images (“truth”) vs **individual runs** (“realization”)

- **“Forced” photometry**: detect in r-band, force in others

- **PSF magnitudes**

Stripe 82

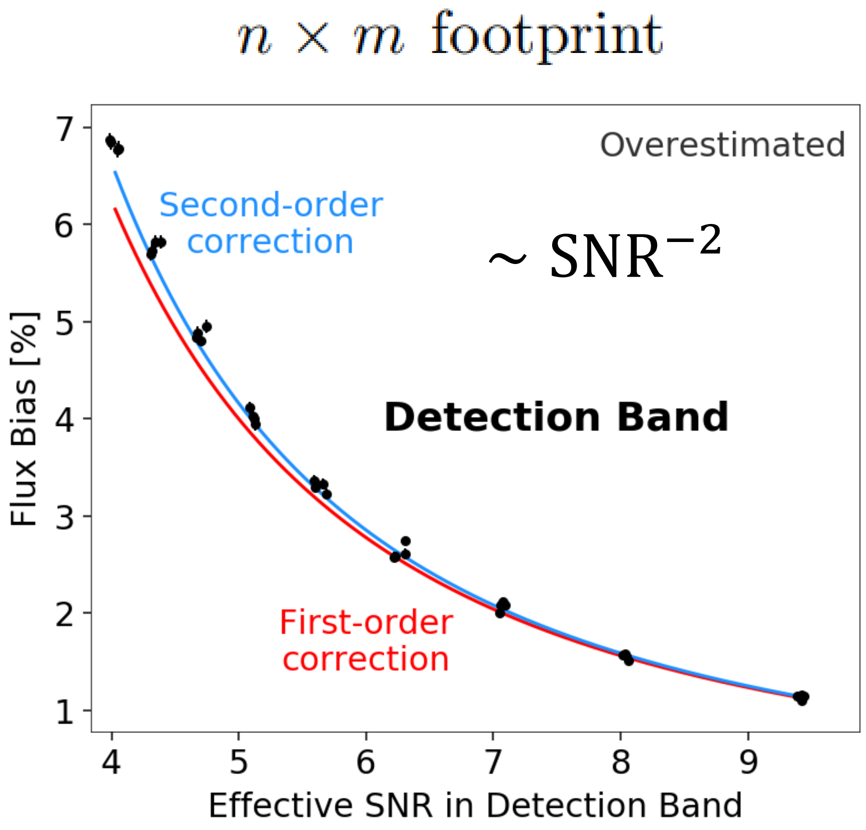
See all the data (and hundreds of similar plots) at https://github.com/joshspeagle/phot_bias

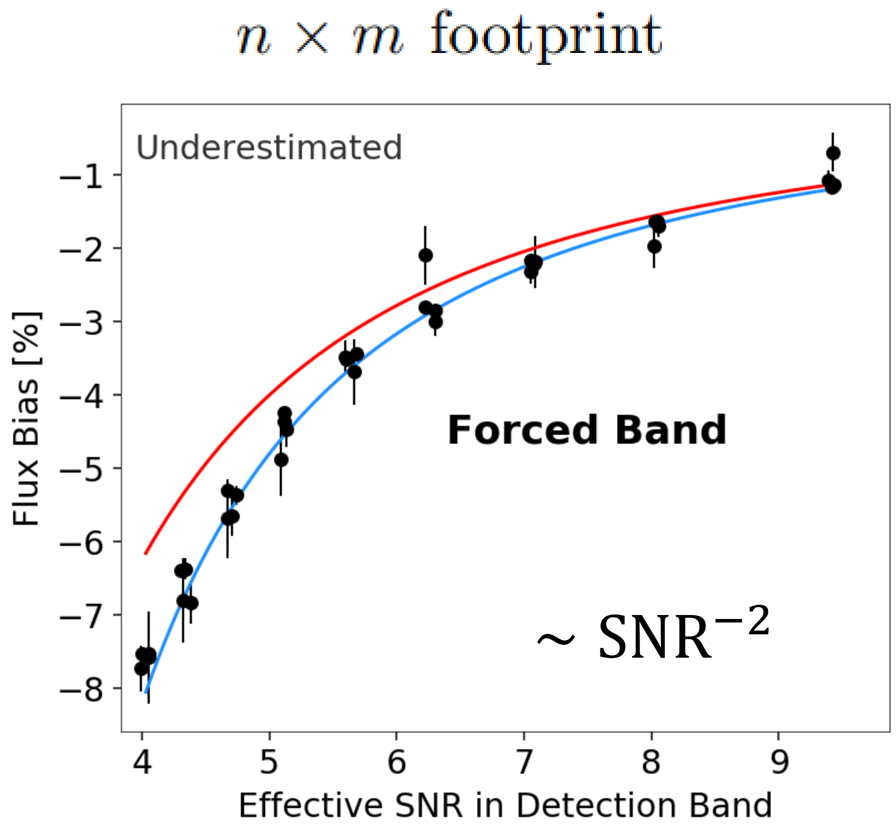


Identical bias across all bands!

Biases in Forced Multi-band Photometry

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





Astrostatistics Interest Group: Student Paper Award
Topic Contributed Papers Session (401) JSM 2020
August 5 2020

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

Stephen Portillo[1,2,*] and **Josh Speagle**[1,3,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington), [3]U. of Toronto

[*]Equal contribution



Portillo, Speagle, & Finkbeiner (**PSF**) 2020, [1902.02374](1902.02374)

Astrostatistics Interest Group: Student Paper Award
Topic Contributed Papers Session (401) JSM 2020
August 5 2020

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

Stephen Portillo[1,2,*] and **Josh Speagle**[1,3,*] and Doug Finkbeiner[1]

[1]Harvard U., [2]DIRAC (U. of Washington), [3]U. of Toronto

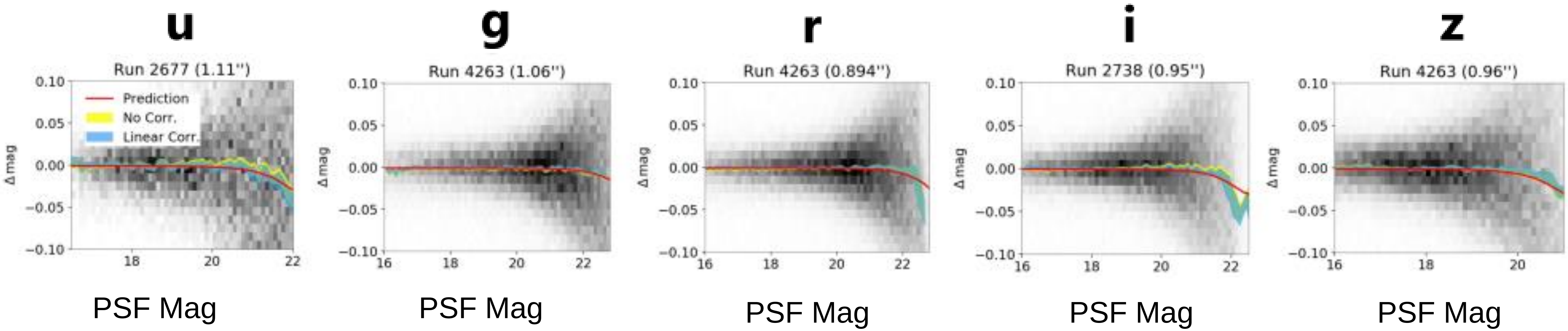
[*]Equal contribution



Portillo, Speagle, & Finkbeiner (**PSF**) 2020, [1902.02374](1902.02374)

See all the data (and hundreds of similar plots) at **https://github.com/joshspeagle/phot_bias**

Stripe 82



Implementation-specific effect:

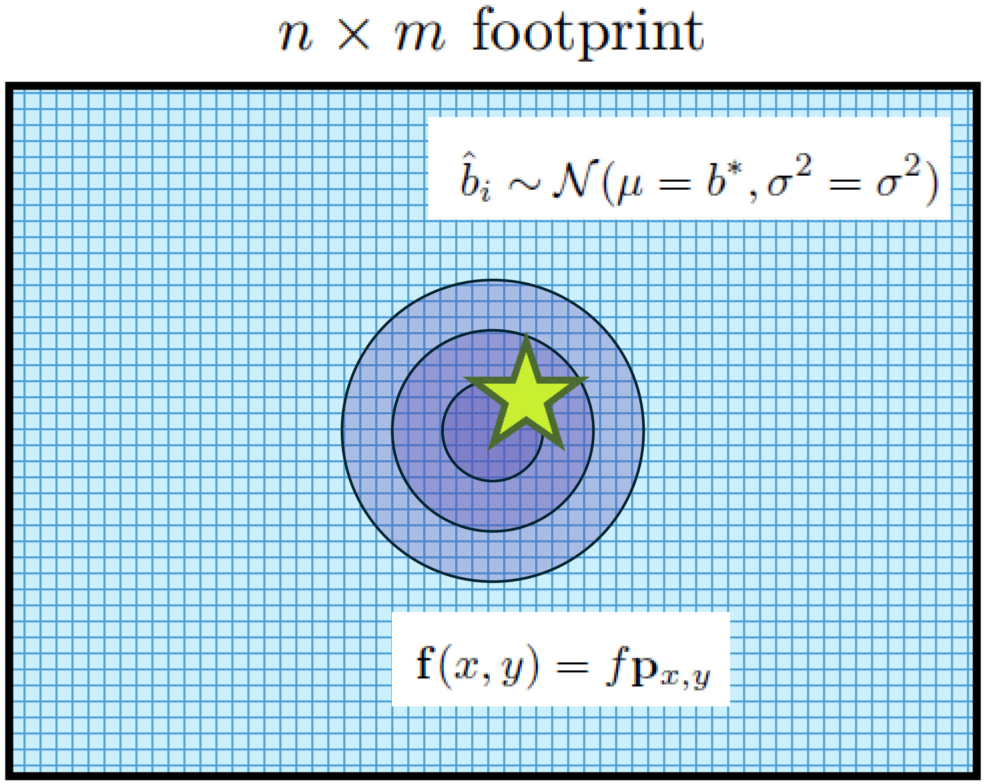
- SDSS pipeline allows for **"local re-centering"** of forced position.
- Enough to undo forcing effect!

Why use models at all?

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

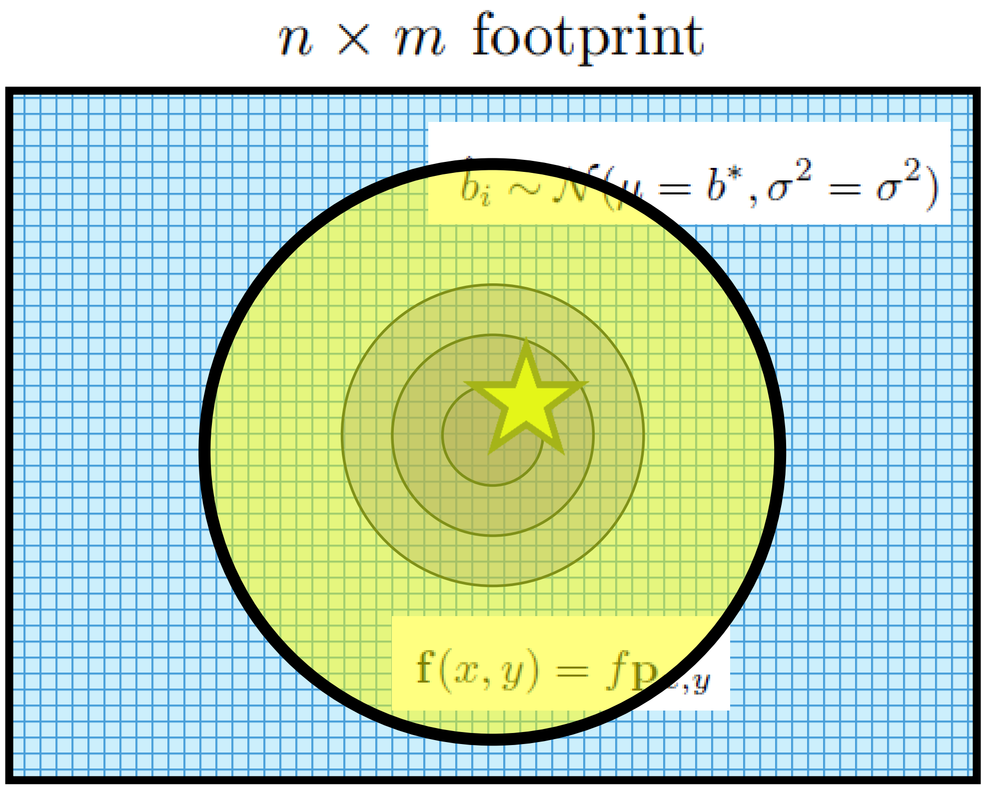
Why use models at all?

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



Why use models at all?

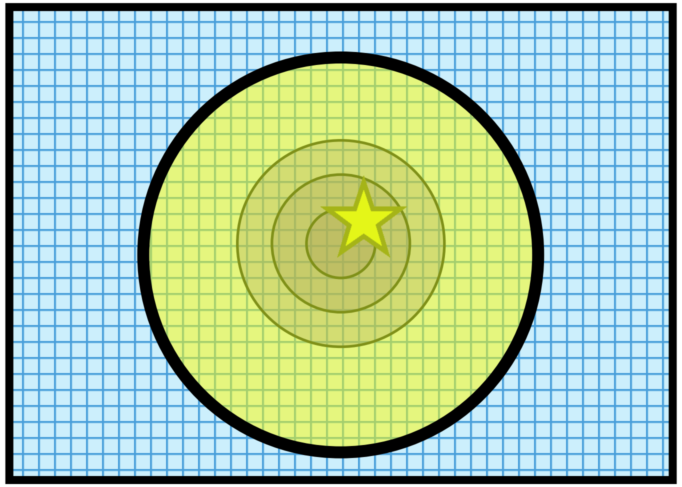
- If we know that the MLE is biased and our models may be wrong, why not do something simple like **aperture photometry**?



Why use models at all?

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

- Apertures are:
  - generally a worse-performing model
  - with larger statistical and systematic errors
  - that are harder to characterize
  - (we confirm this with SDSS S82 data)

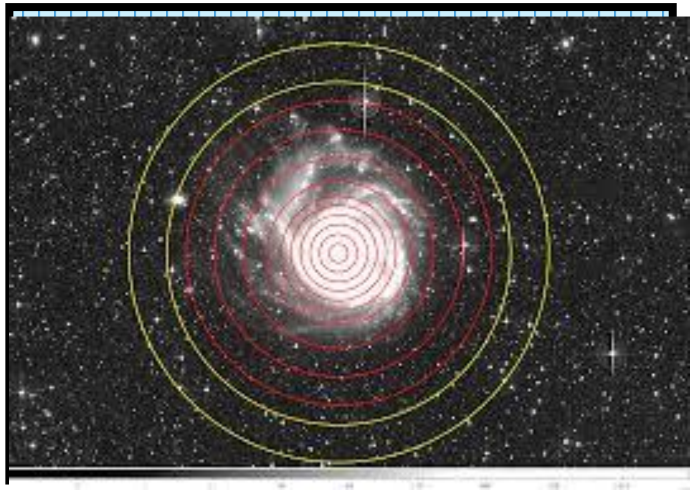


Why use models at all?

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

- Apertures are:
  - generally a worse-performing model
  - with larger statistical and systematic errors
  - that are harder to characterize
  - (we confirm this with SDSS S82 data)

burro.case.edu



Has a purpose, but should be used judiciously!

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, **more severe for galaxies** (>2.5x worse).
- **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, **more severe for galaxies** (>2.5x worse).
- **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.