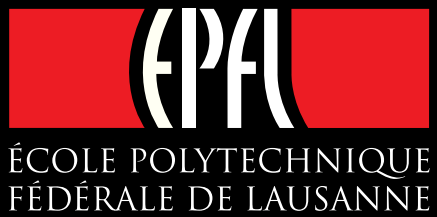
Machine learning for PSF characterisation and interpolation



Thibault Kuntzer

Frédéric Courbin

CEA

17 October 2016

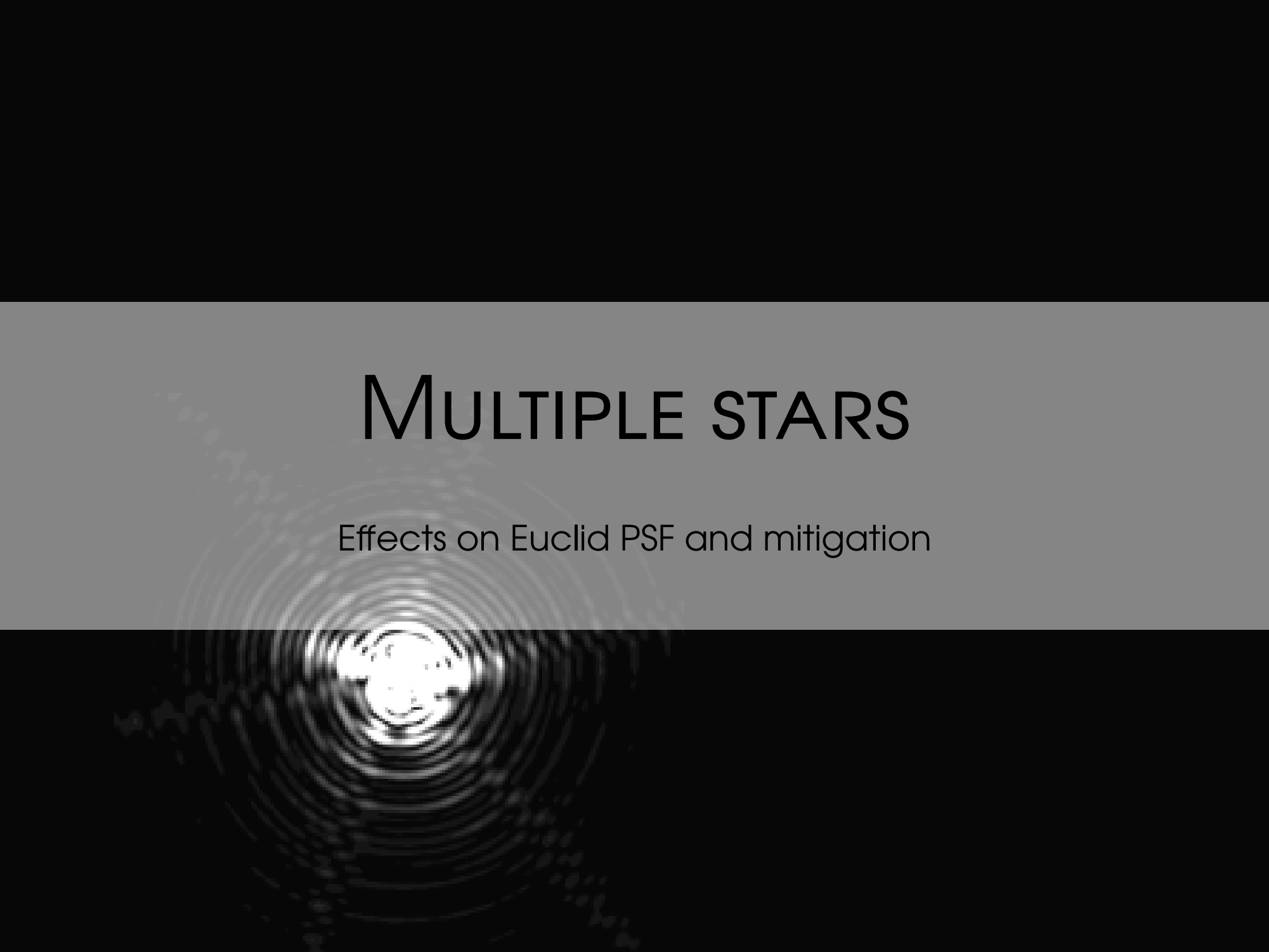
PDF available at **http://obswww.unige.ch/∼kuntzer/talks/cea_psf_kuntzer.pdf**

Outline

- Binaries issue, PSF systematics & mitigation
- Building a colour-magnitude diagram with one filter
- PSF interpolation revisited

Multiple stars

Effects on Euclid PSF and mitigation



PSF requirements and measurement

- Measured shear power spectrum

$$\hat{C} \approx (1 + \mathcal{M})C + \mathcal{A}, \qquad \mathcal{M}, \mathcal{A} \sim \langle |\delta e_{\mathrm{PSF}}|^2 \rangle, \langle \delta R^2_{\mathrm{PSF}} \rangle$$

- Science requirement for the PSF shape:

size rms (relative)	$\frac{\sigma(R^2)}{\langle R^2 \rangle} \leq 10^{-3}$
ellipticity rms	$\sigma(e_i) \leq 10^{-4}$
ellipticity	$e_i < 0.15$

From Paulin-Henriksson+2008

- Performance needed for measurement

size	$\sim 5\%$
ellipticity	$\sim 1\%$

PSF requirements and measurement

- Measured shear power spectrum

$$\hat{C} \approx (1 + \mathcal{M})C + \mathcal{A}, \qquad \mathcal{M}, \mathcal{A} \sim \langle |\delta e_{\mathrm{PSF}}|^2 \rangle, \langle \delta R^2_{\mathrm{PSF}} \rangle$$

- Science requirement for the PSF shape:

size rms (relative)	$\frac{\sigma(R^2)}{\langle R^2 \rangle} \leq 10^{-3}$
ellipticity rms	$\sigma(e_i) \leq 10^{-4}$
ellipticity	$e_i < 0.15$

From Paulin-Henriksson+2008

- Performance needed for measurement

size	$\sim 5\%$
ellipticity	$\sim 1\%$

PSF requirements and measurement

- Measured shear power spectrum

$$\hat{C} \approx (1 + \mathcal{M})C + \mathcal{A}, \qquad \mathcal{M}, \mathcal{A} \sim \langle |\delta e_{\mathrm{PSF}}|^2 \rangle, \langle \delta R^2_{\mathrm{PSF}} \rangle$$

- Science requirement for the PSF shape:

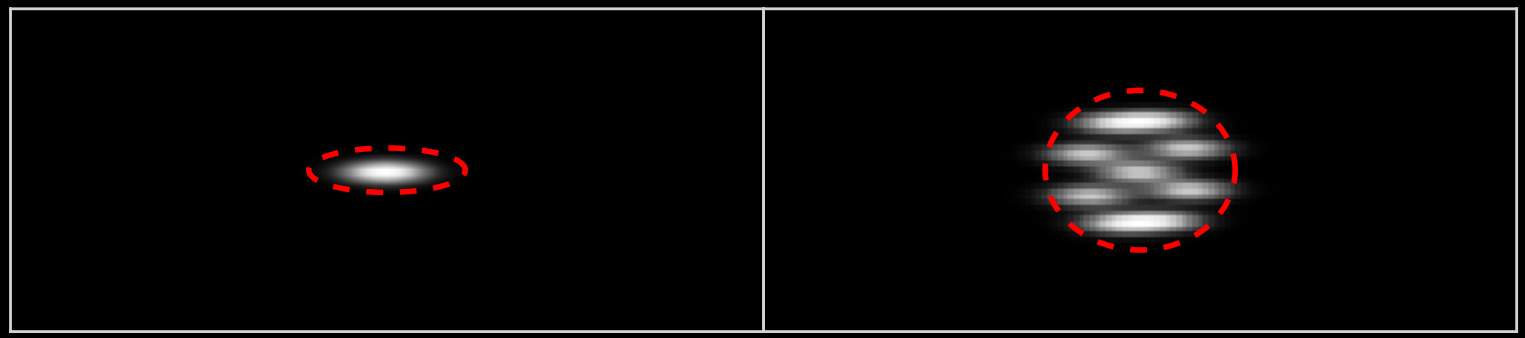
size rms (relative)	$\frac{\sigma(R^2)}{\langle R^2 \rangle} \leq 10^{-3}$
ellipticity rms	$\sigma(e_i) \leq 10^{-4}$
ellipticity	$e_i < 0.15$

From Paulin-Henriksson+2008

- Performance needed for measurement

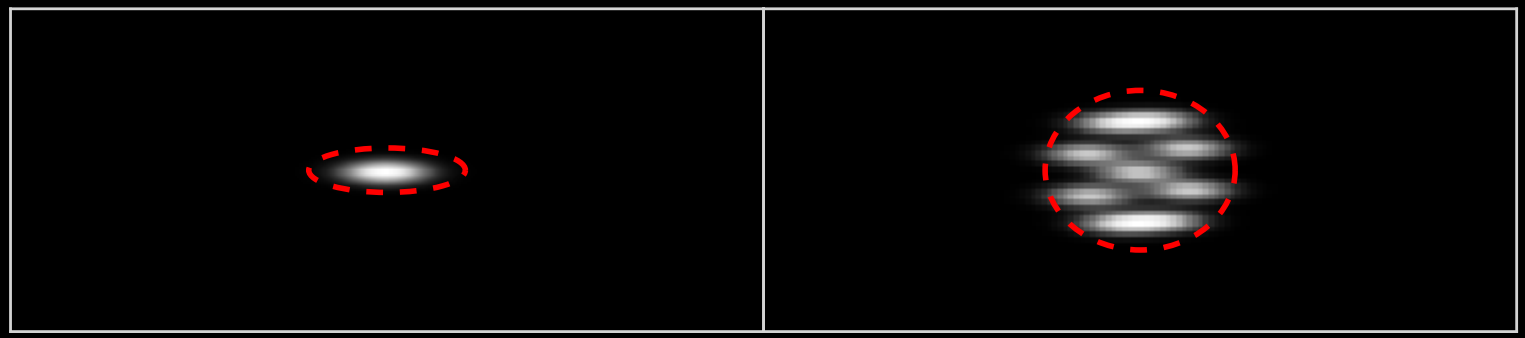
size	$\sim 5\%$
ellipticity	$\sim 1\%$

The curse of binaires



- Unresolved objects affect the convolution kernel
- Multiple stars are ubiquitous ($\gtrsim$ 35%)

The curse of binaires

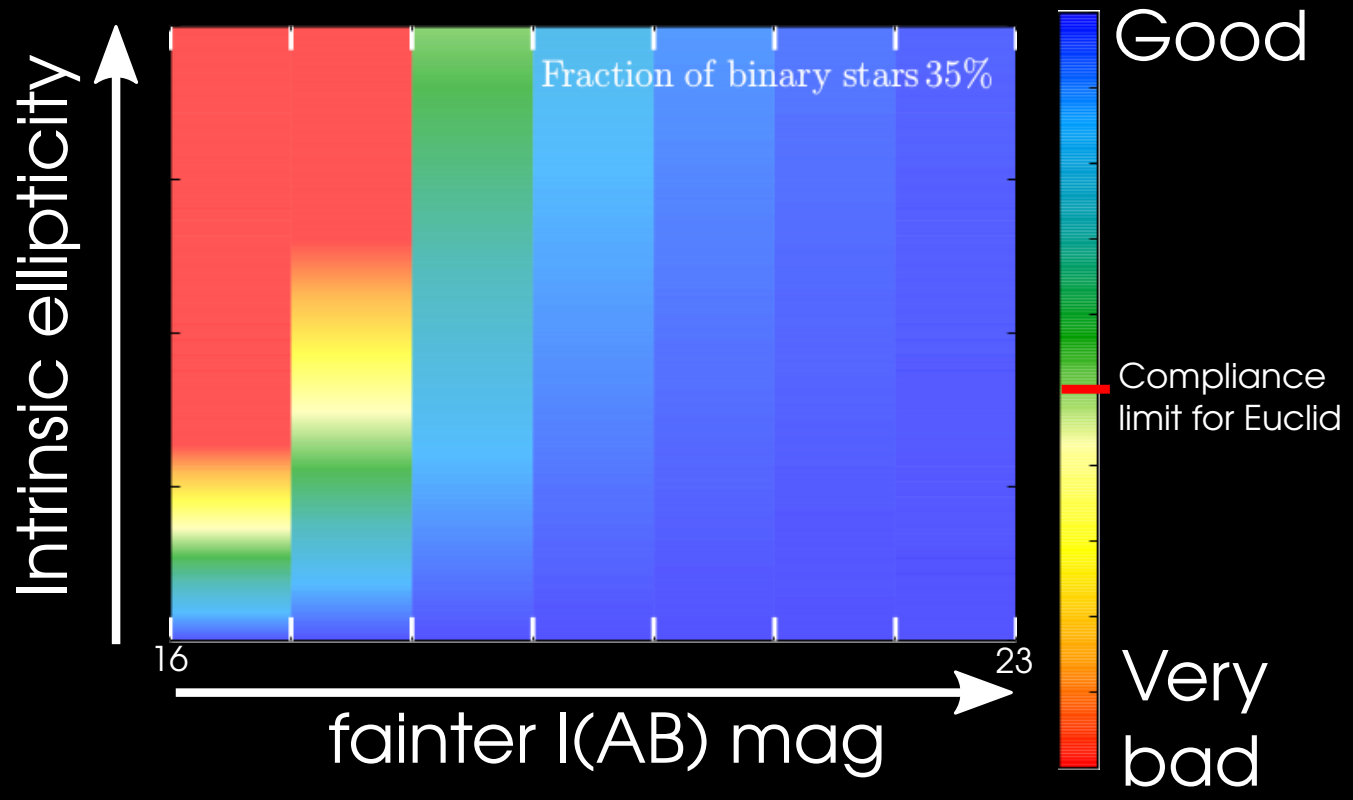


- Unresolved objects affect the convolution kernel
- Multiple stars are ubiquitous ($\gtrsim$ 35%)

The curse of binaires

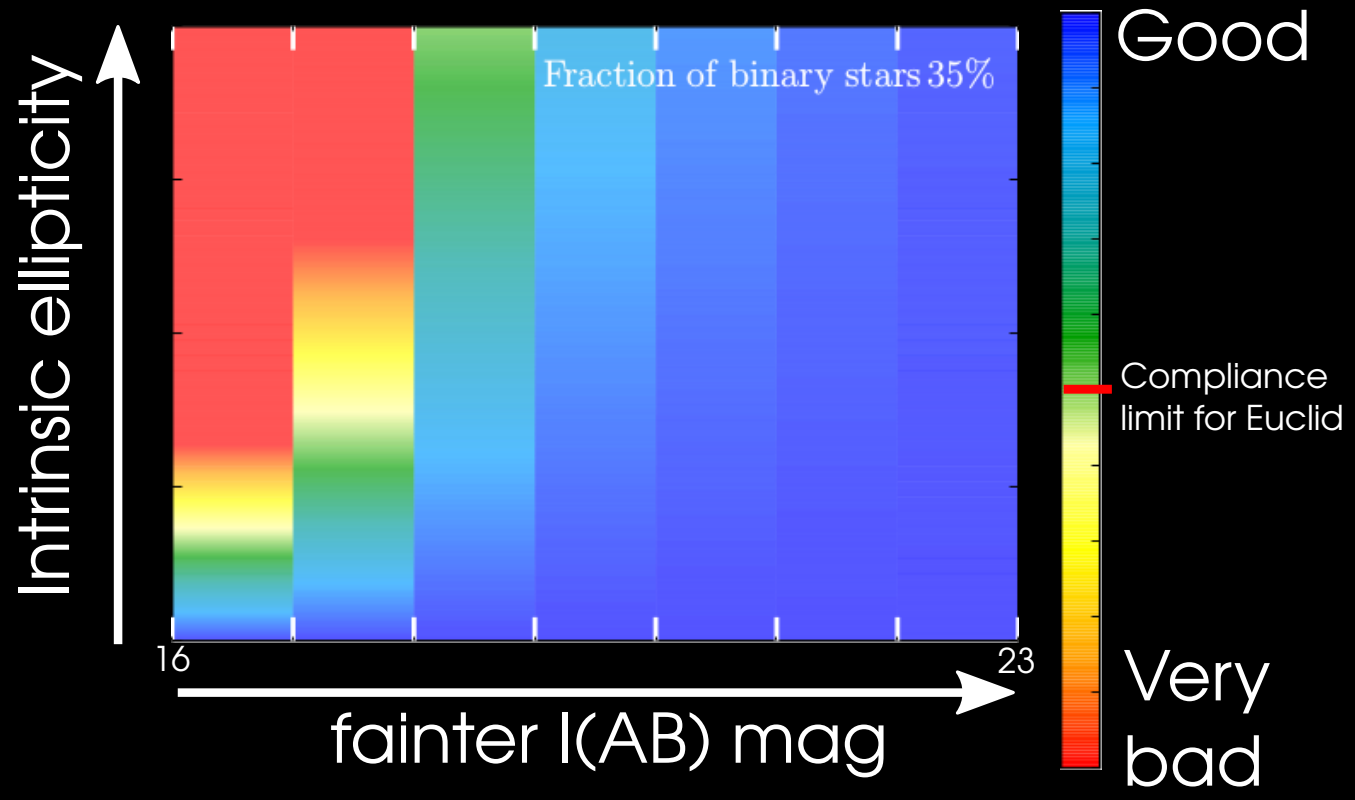
- Unresolved objects affect the convolution kernel
- Multiple stars are ubiquitous ($\gtrsim$ 35%)

Predictions in an ideal world



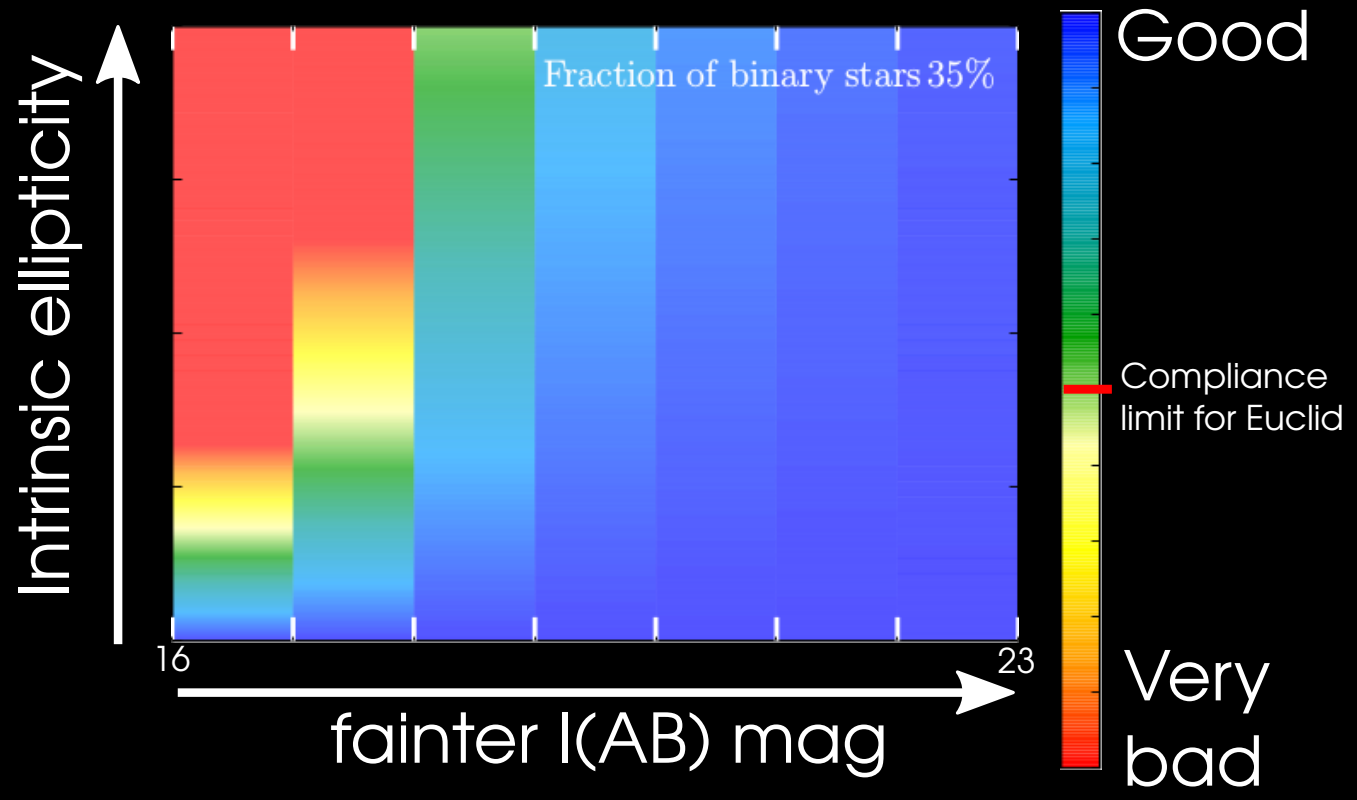
- Stars $i(AB) \lesssim 18$: Non compliant
- Stars $i(AB) \lesssim 20$: Significant part of budget

Predictions in an ideal world



- Stars $i(AB) \lesssim 18$: Non compliant
- Stars $i(AB) \lesssim 20$: Significant part of budget

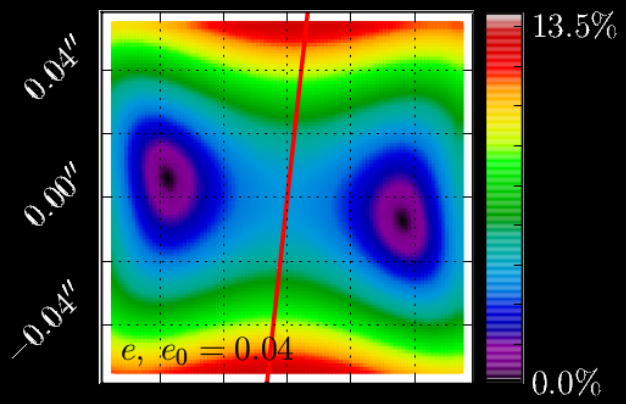
Predictions in an ideal world



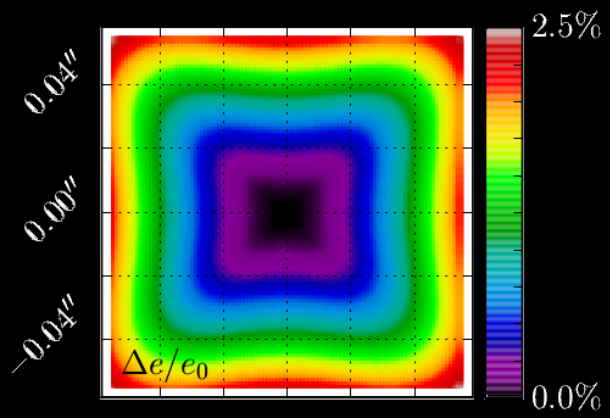
- Stars $i(AB) \lesssim 18$: Non compliant
- Stars $i(AB) \lesssim 20$: Significant part of budget

Alterations of the PSF due to a single companion

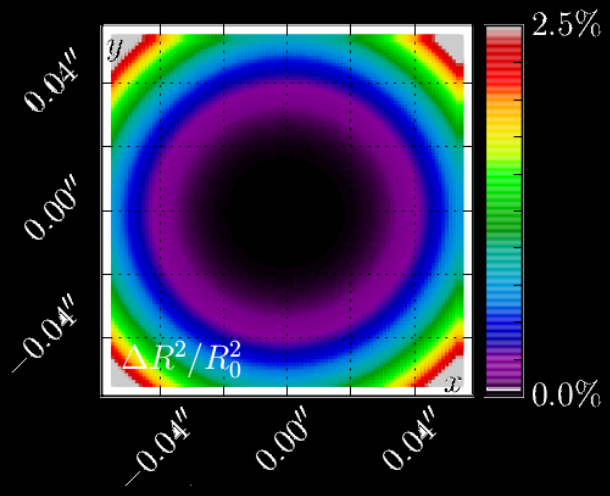
e



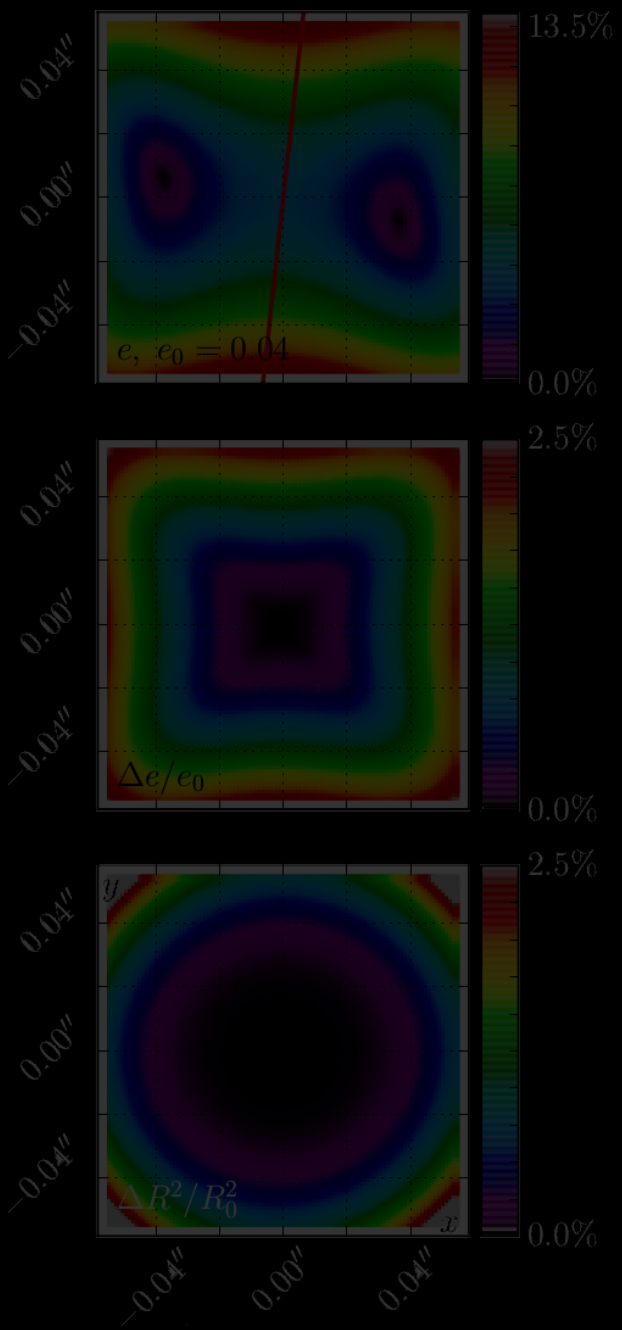
$\frac{\Delta e}{e_0}$



$\frac{\Delta R^2}{R_0^2}$

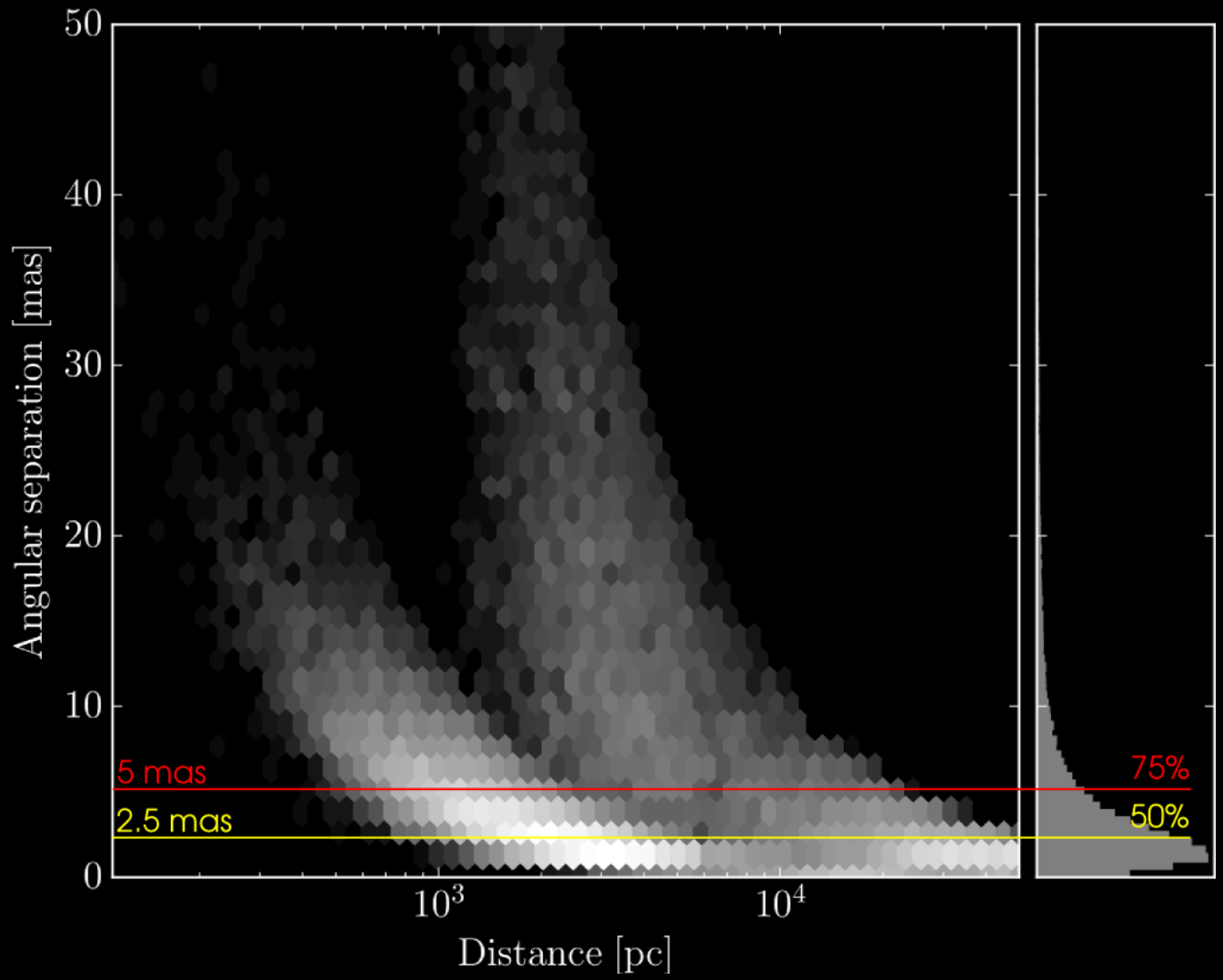


Alterations of the PSF due to a single companion

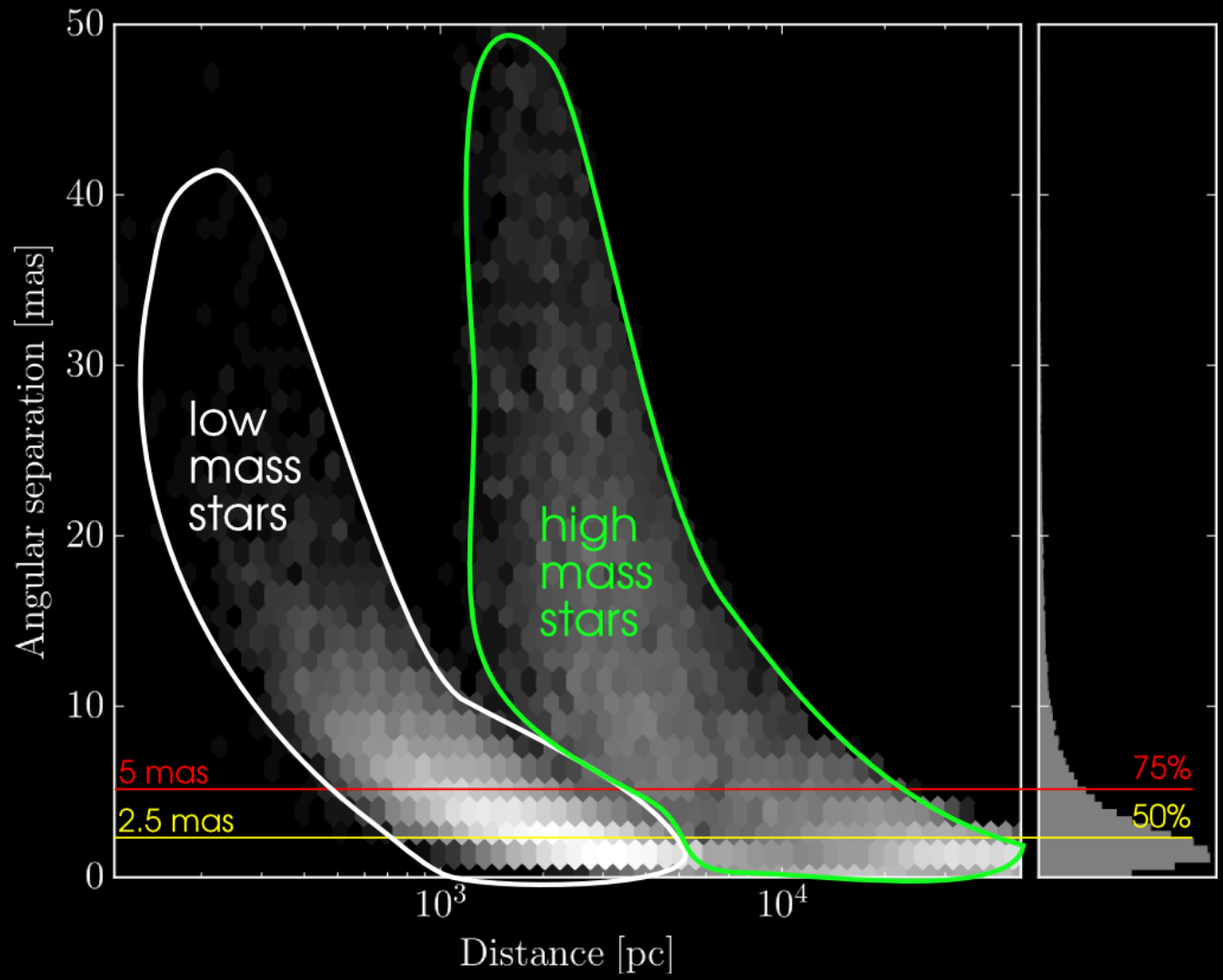


size	∼ 5%
ellipticity	∼ 1%

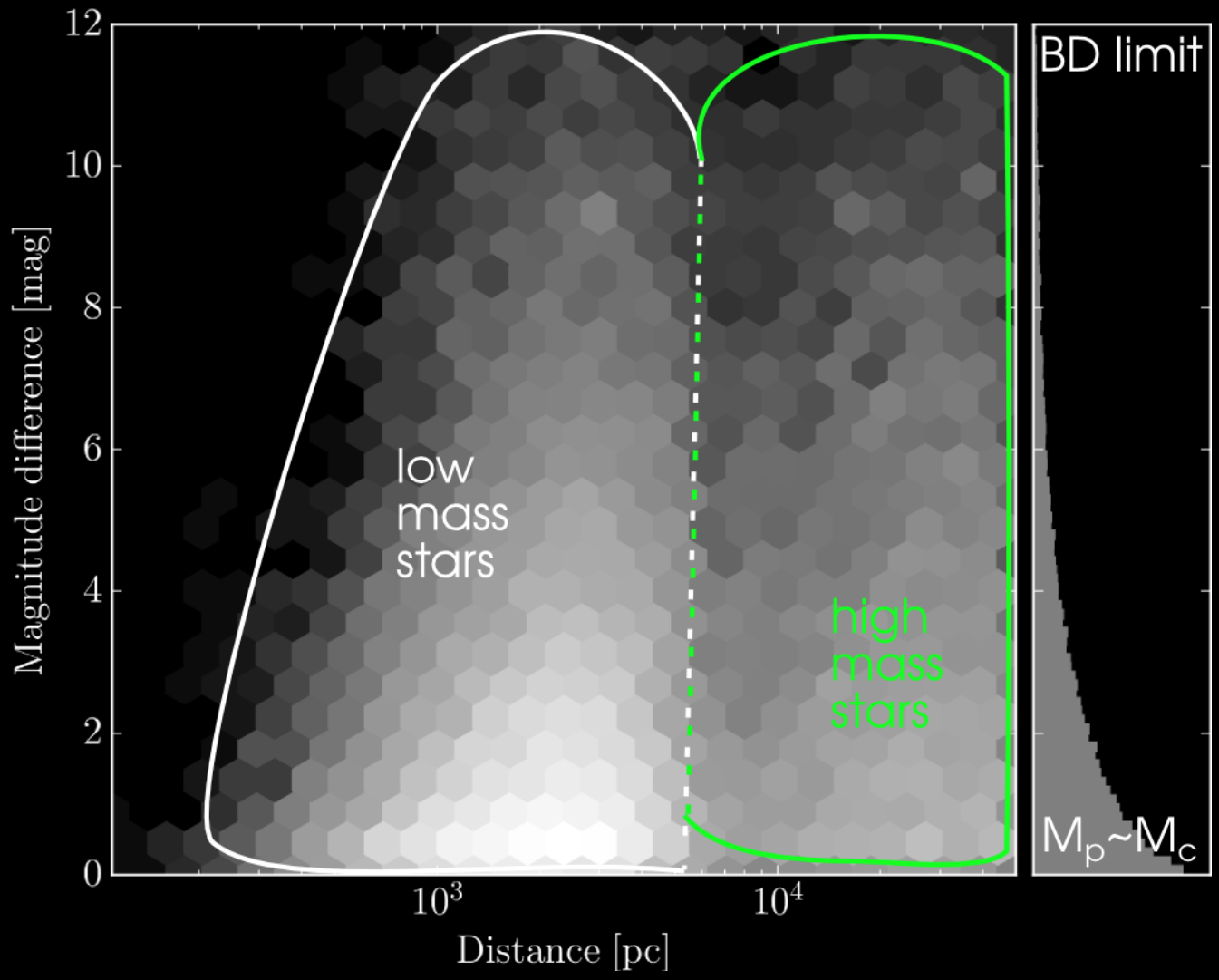
The binaries in the sky of Euclid



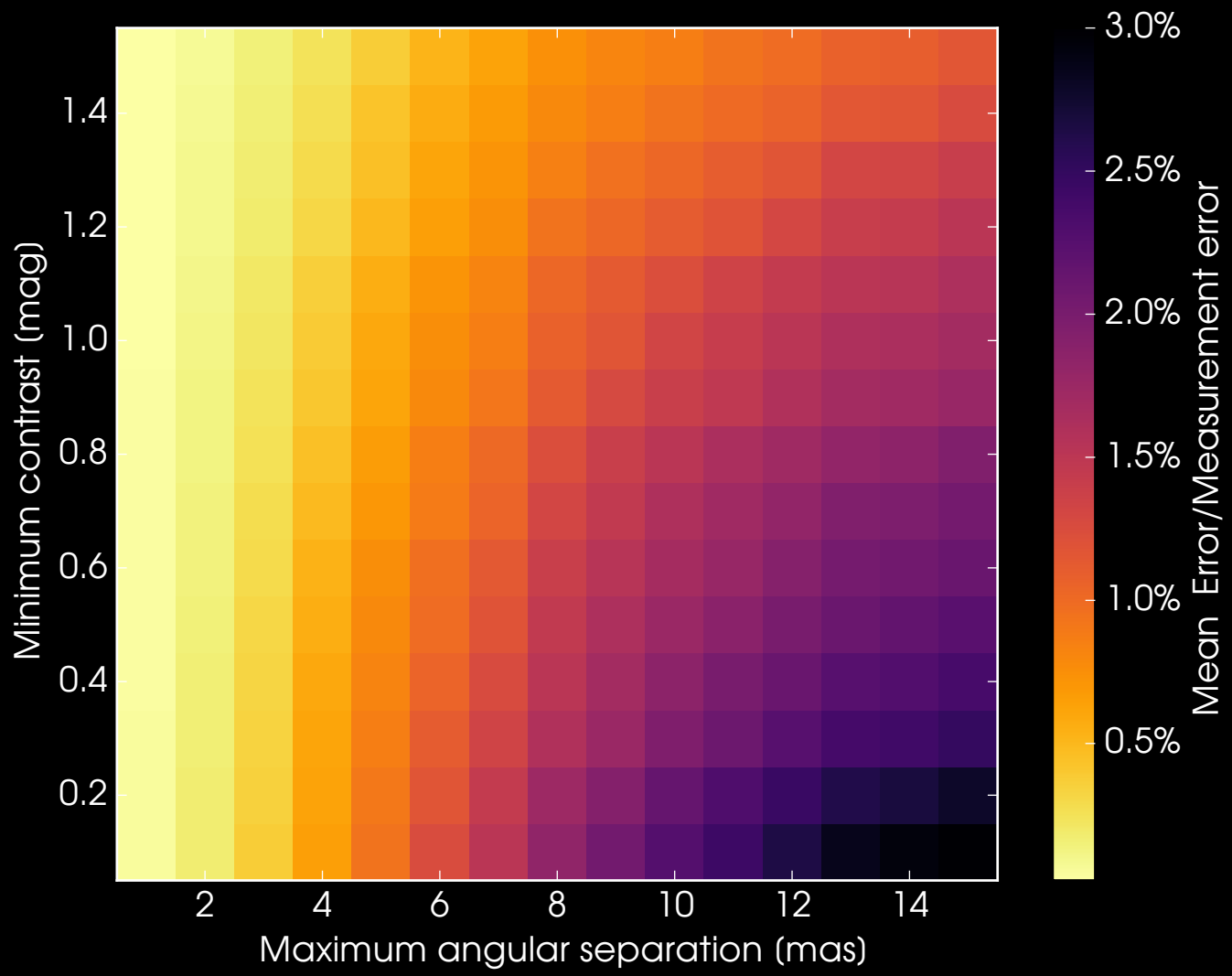
The binaries in the sky of Euclid



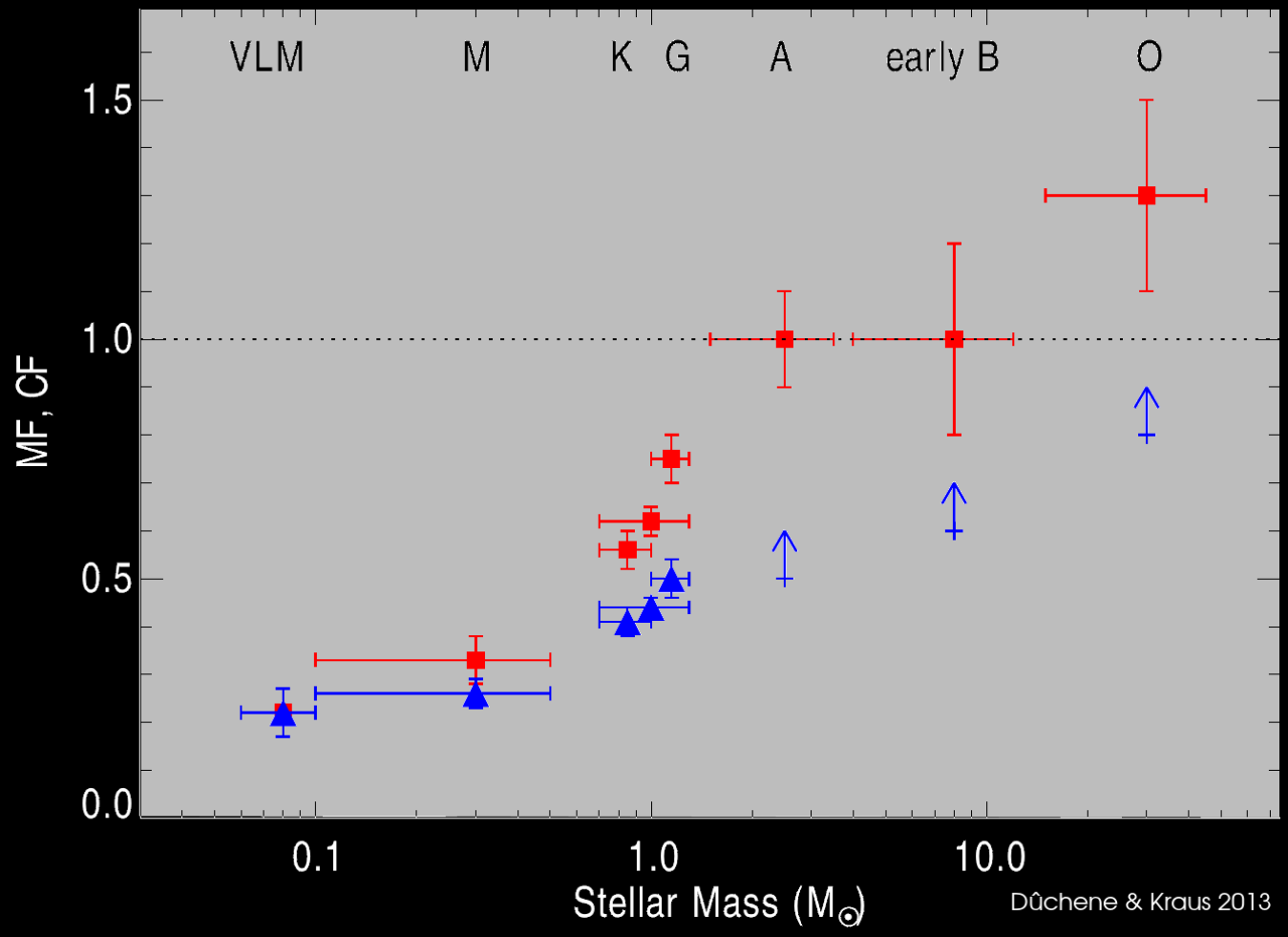
The binaries in the sky of Euclid

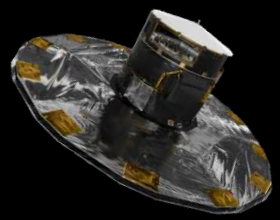


The binaries in the sky of Euclid



Identifying multiple stars?



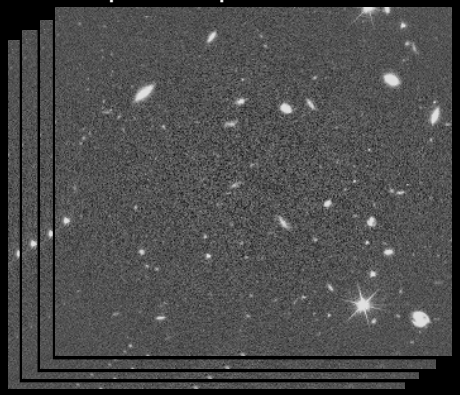


Star w/in specs => use!

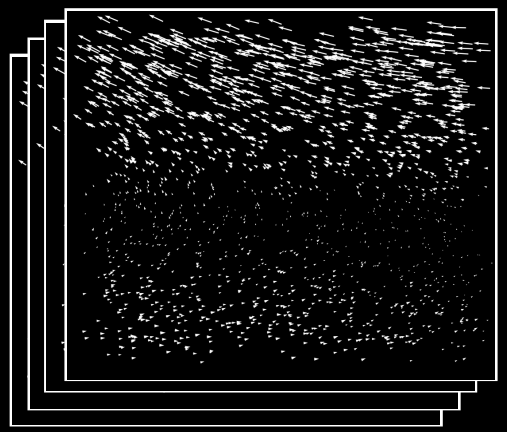
multiple star => reject!

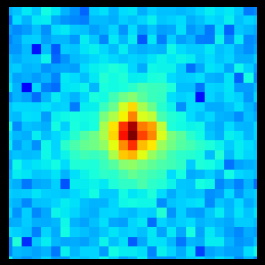
Identifying multiple stars?

Multiple exposures



PSF measurements





Spectral class

Cross exposure analysis

Correlated noise?

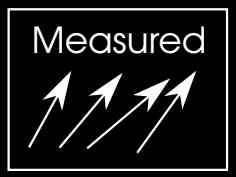
Not measurable: single star w/in specs => use!

Yes: might be a multiple star => reject!

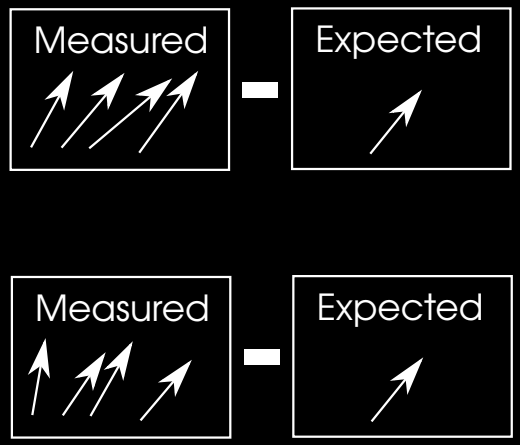
P(binary) low => use!

P(binary) high => reject!

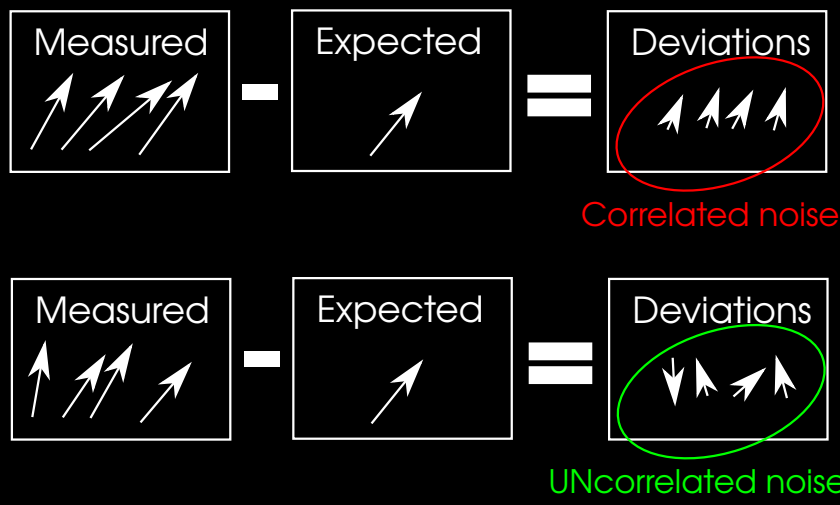
Exploiting correlated noise



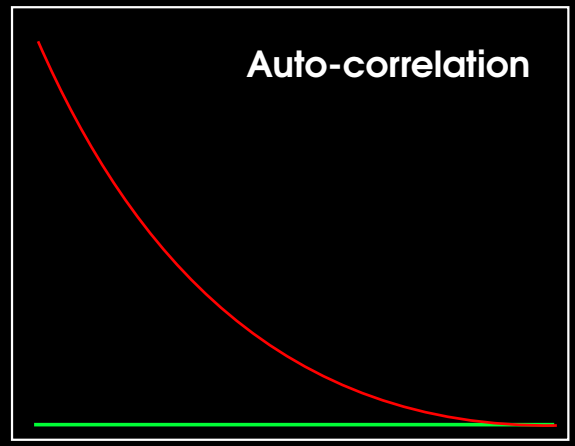
Exploiting correlated noise



Exploiting correlated noise



Data-driven



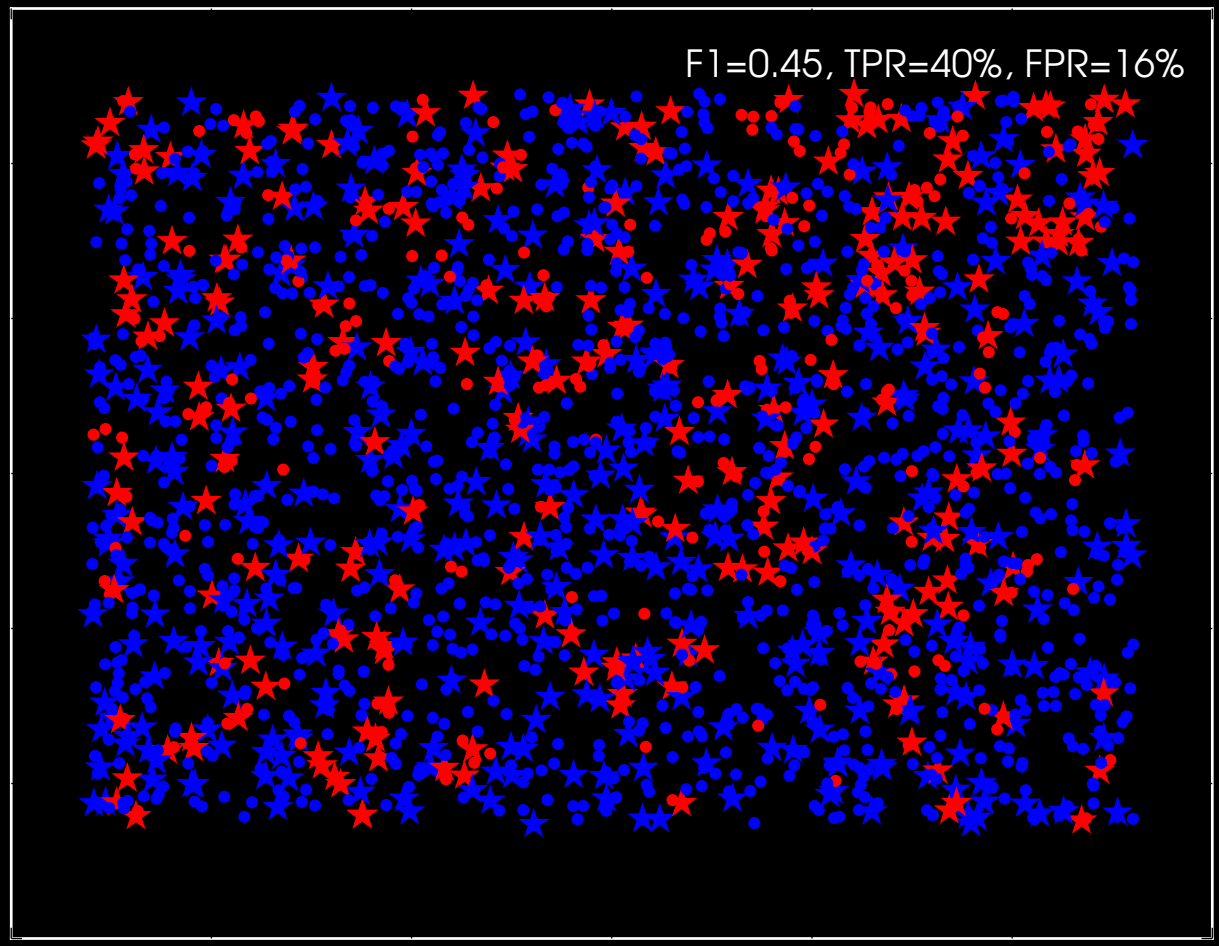
Model-dependent

Reconstructing the PSF field

For every star:

- Interpolate PSF parameter at position (Out-of-bag prediction)
- Measured PSF parameter − interpolated
- Infer binary presence from systematic bias

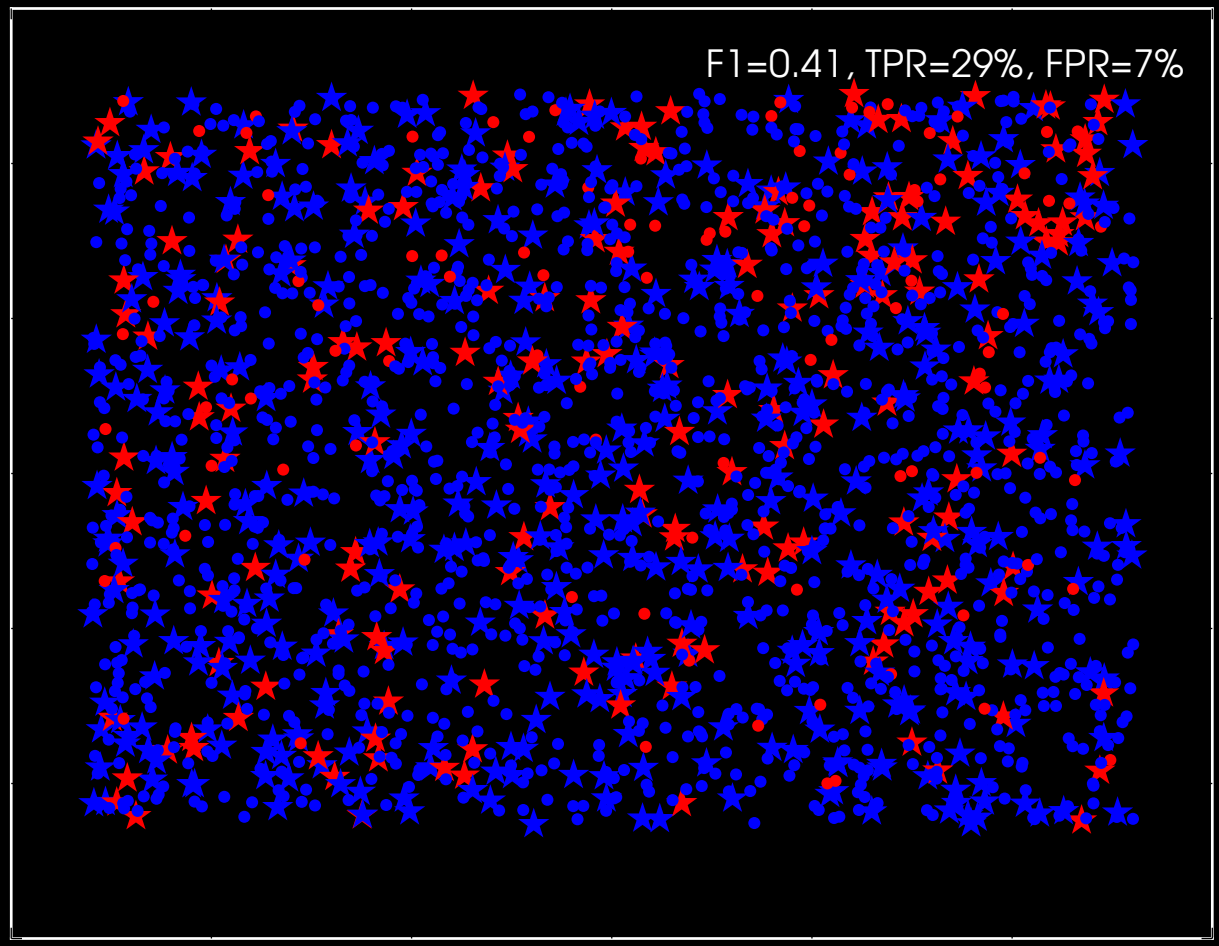
Reconstructing the PSF field



For every star:

- Interpolate PSF parameter at position (Out-of-bag prediction)
- Measured PSF parameter − interpolated
- Infer binary presence from systematic bias

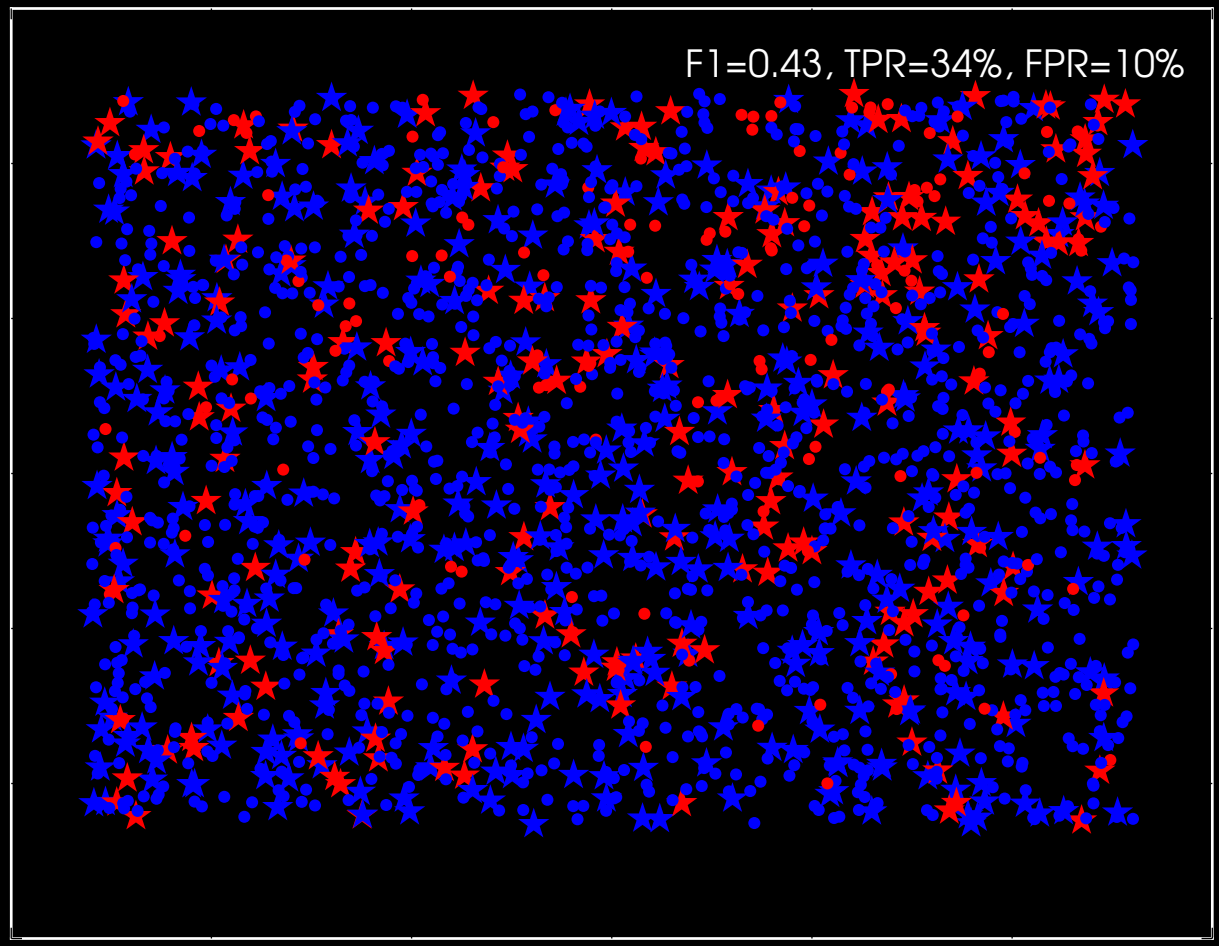
Reconstructing the PSF field



For every star:

- Interpolate PSF parameter at position (Out-of-bag prediction)
- Measured PSF parameter − interpolated
- Infer binary presence from systematic bias

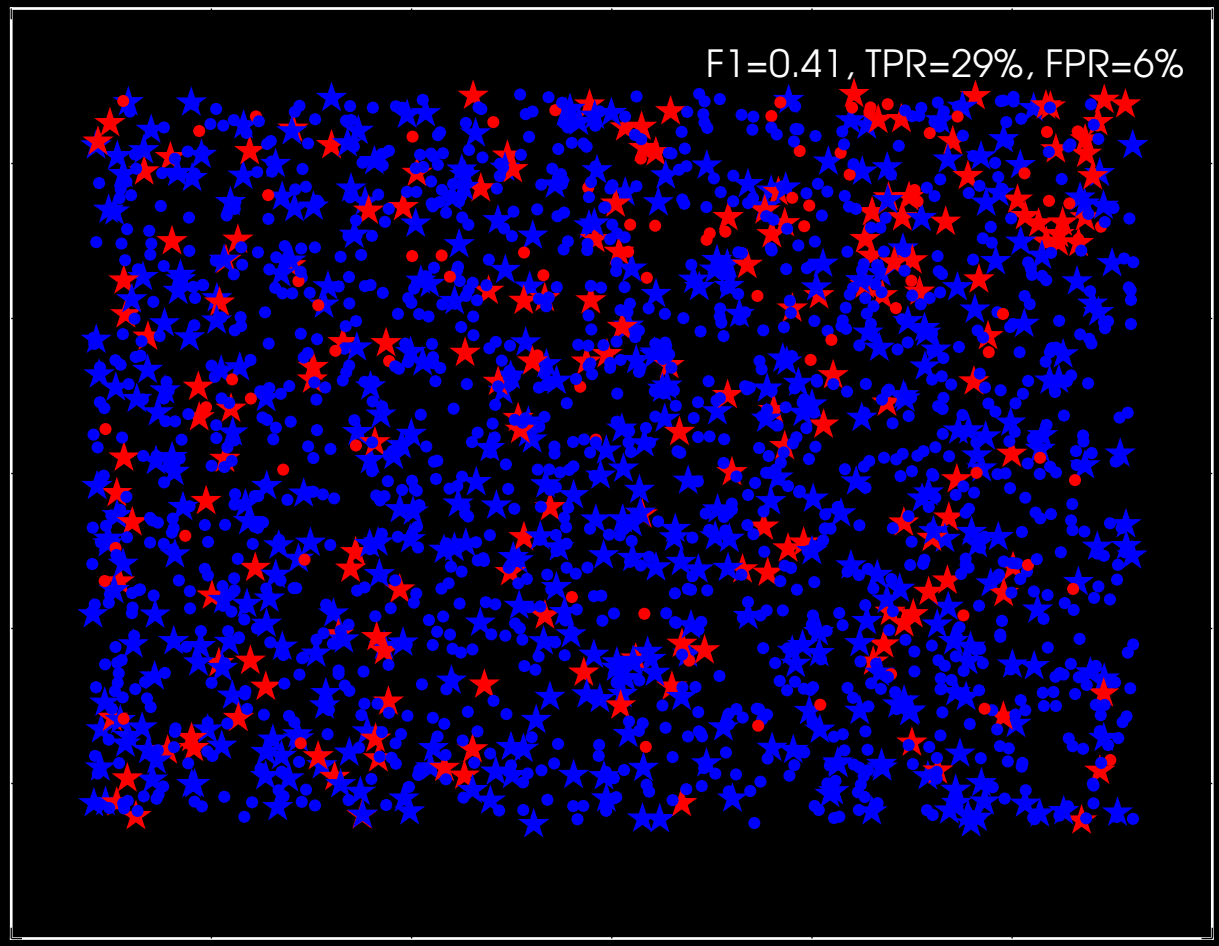
Reconstructing the PSF field



For every star:

- Interpolate PSF parameter at position (Out-of-bag prediction)
- Measured PSF parameter − interpolated
- Infer binary presence from systematic bias

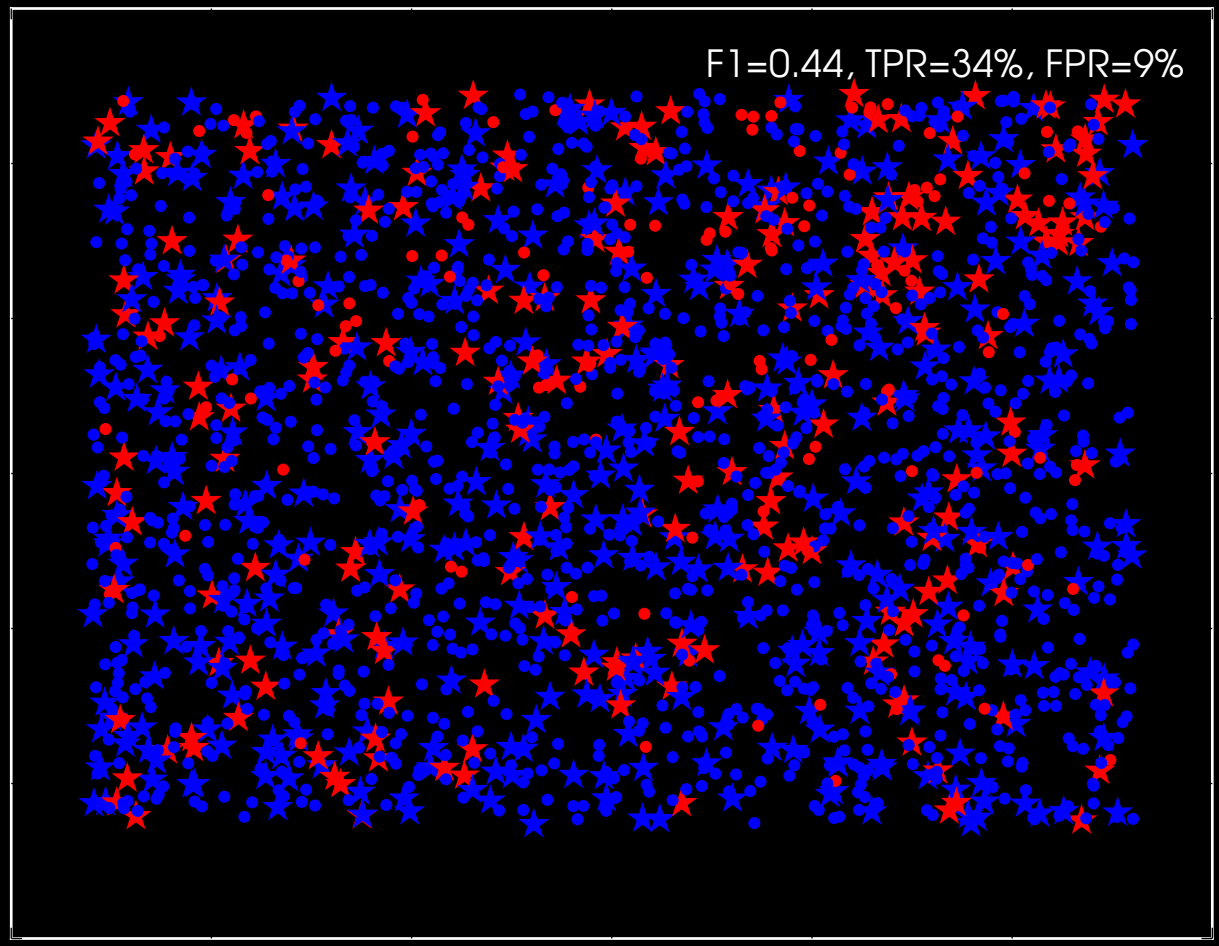
Reconstructing the PSF field



For every star:

- Interpolate PSF parameter at position (Out-of-bag prediction)
- Measured PSF parameter − interpolated
- Infer binary presence from systematic bias

Reconstructing the PSF field



For every star:

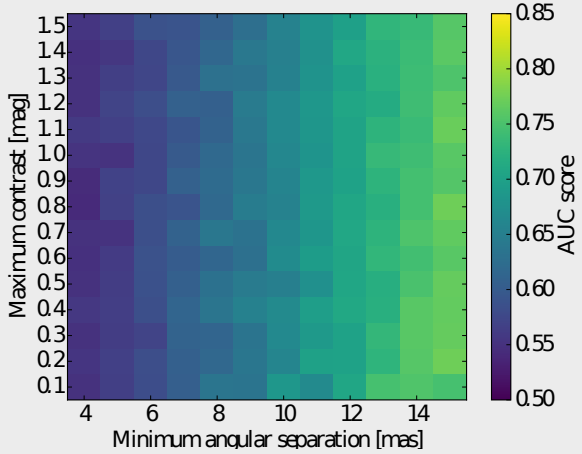
- Interpolate PSF parameter at position (Out-of-bag prediction)
- Measured PSF parameter − interpolated
- Infer binary presence from systematic bias

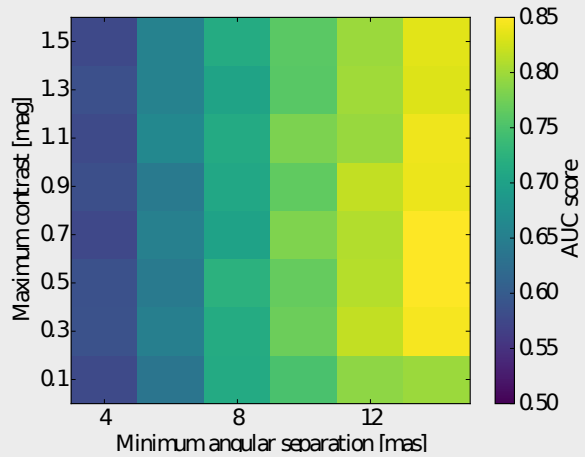
AUC metrics

ACF/Naive

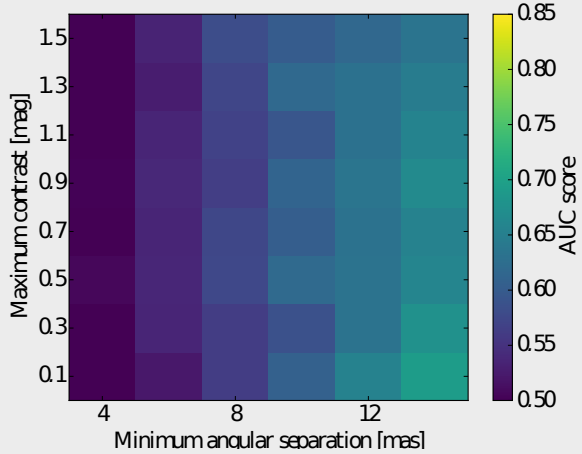
Random Forest

Known PSF

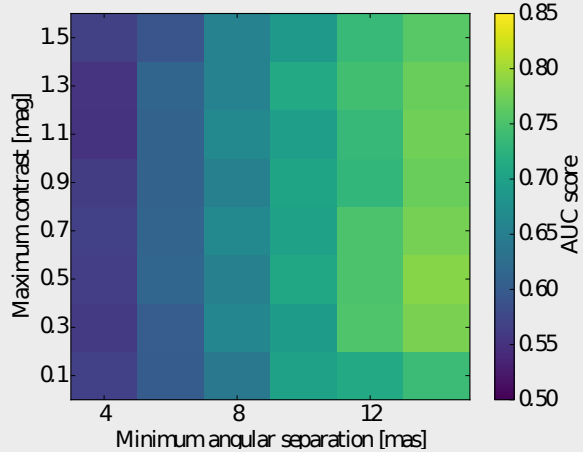




Unknown PSF



False positive: ~20%



False positive: ~5-10%

Summary: finding binaries

- Most of Euclid's binaries have small angular separations
- Effects on PSF measurement is $\mathcal{O}(10^{-4})$
- Identify harmful binaries (check Kuntzer+16a) using only VIS 4 exposures
- Looking for systematic biases in measured PSF parameters
- ✓ "Naive:" auto-correlation analysis, works, but high FPR (data-driven)
- ✓ "Machine learning:" promising, with low FPR (depends on model)
- Work in progress:
  - More simulations for healthier metrics evaluations
  - Bayesian approach
  - Inclusion of realistic issues ? (Stars not present in all exposures, dithering, ...)

Summary: finding binaries

- Most of Euclid's binaries have small angular separations
- Effects on PSF measurement is $\mathcal{O}(10^{-4})$
- Identify harmful binaries (check Kuntzer+16a) using only VIS 4 exposures
- Looking for systematic biases in measured PSF parameters

✓ "Naive:" auto-correlation analysis, works, but high FPR (data-driven)

✓ "Machine learning:" promising, with low FPR (depends on model)

- Work in progress:
  - More simulations for healthier metrics evaluations
  - Bayesian approach
  - Inclusion of realistic issues ? (Stars not present in all exposures, dithering, ...)

Summary: finding binaries

- Most of Euclid's binaries have small angular separations
- Effects on PSF measurement is $\mathcal{O}(10^{-4})$
- Identify harmful binaries (check Kuntzer+16a) using only VIS 4 exposures
- Looking for systematic biases in measured PSF parameters

✓ "Naive:" auto-correlation analysis, works, but high FPR (data-driven)

✓ "Machine learning:" promising, with low FPR (depends on model)

- Work in progress:
  - More simulations for healthier metrics evaluations
  - Bayesian approach
  - Inclusion of realistic issues ? (Stars not present in all exposures, dithering, ...)

Summary: finding binaries

- Most of Euclid's binaries have small angular separations
- Effects on PSF measurement is $\mathcal{O}(10^{-4})$
- Identify harmful binaries (check Kuntzer+16a) using only VIS 4 exposures
- Looking for systematic biases in measured PSF parameters
- ✓ "Naive:" auto-correlation analysis, works, but high FPR (data-driven)
- ✓ "Machine learning:" promising, with low FPR (depends on model)
- Work in progress:
  - More simulations for healthier metrics evaluations
  - Bayesian approach
  - Inclusion of realistic issues ? (Stars not present in all exposures, dithering, ...)

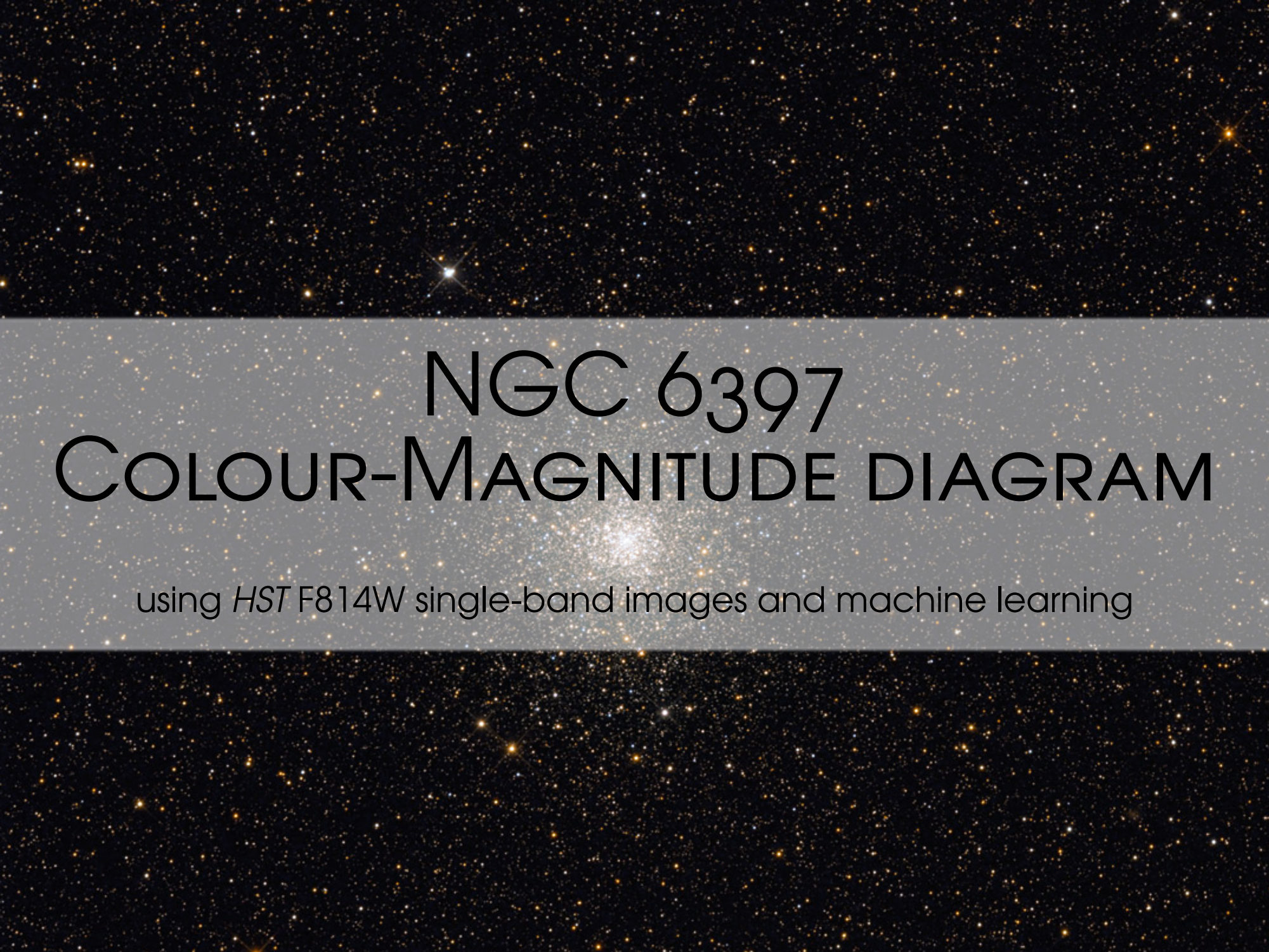
Summary: finding binaries

- Most of Euclid's binaries have small angular separations
- Effects on PSF measurement is $\mathcal{O}(10^{-4})$
- Identify harmful binaries (check Kuntzer+16a) using only VIS 4 exposures
- Looking for systematic biases in measured PSF parameters

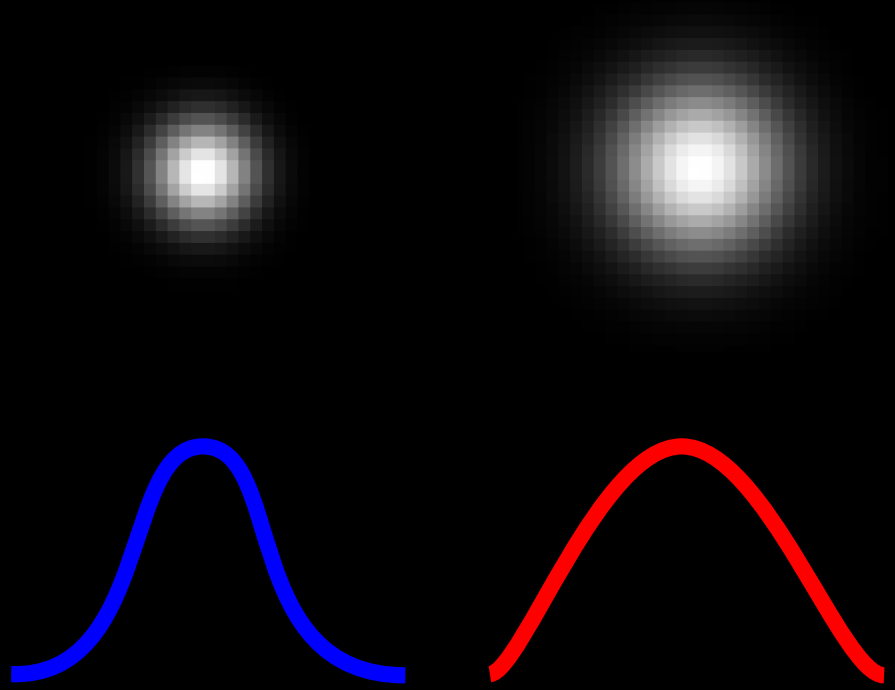
✓ "Naive:" auto-correlation analysis, works, but high FPR (data-driven)

✓ "Machine learning:" promising, with low FPR (depends on model)

- Work in progress:
  - More simulations for healthier metrics evaluations
  - Bayesian approach
  - Inclusion of realistic issues ? (Stars not present in all exposures, dithering, ...)


NGC 6397
Colour-Magnitude diagram
using *HST* F814W single-band images and machine learning

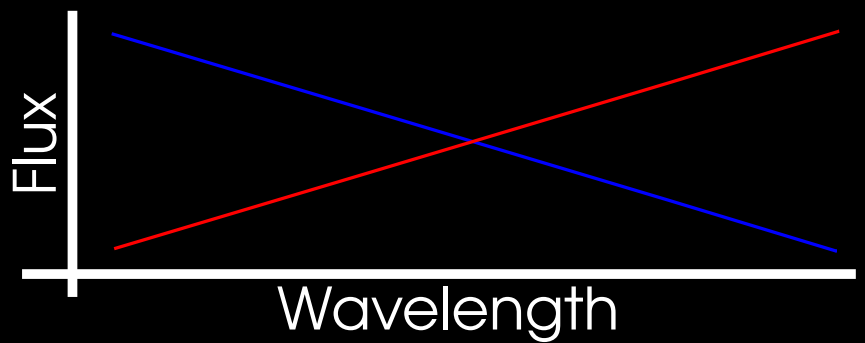
Stars in a wide band



- Stellar mass correlate with effective surface temperature
  $\Longrightarrow$ spectra $\sim T_{\text{eff}}$
- Different slope $\Longrightarrow$ different PSF (chromaticity!)

Angular resolution: $\theta \sim \frac{\lambda}{D}$

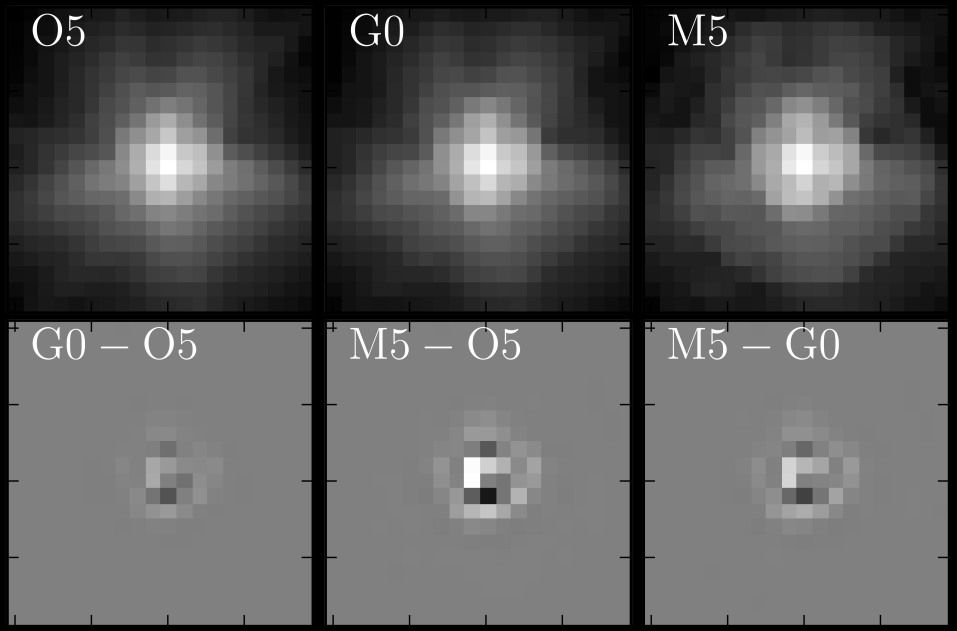
Stars in a wide band



- Stellar mass correlate with effective surface temperature
  $\Longrightarrow$ spectra $\sim T_{\text{eff}}$
- Different slope $\Longrightarrow$ different PSF (chromaticity!)

Angular resolution: $\theta \sim \frac{\lambda}{D}$

Stars in a wide band

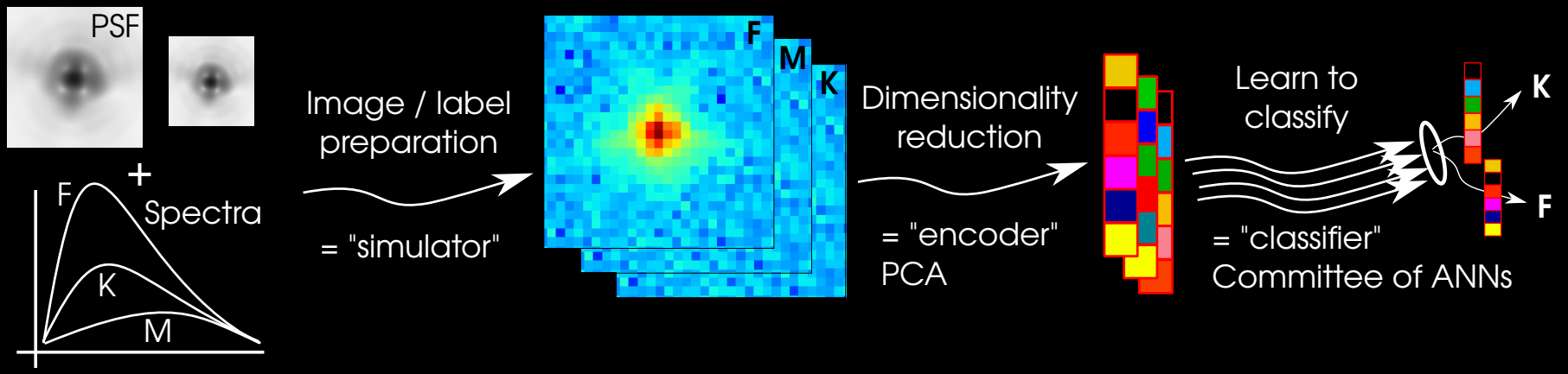


- Stellar mass correlate with effective surface temperature
  $\Longrightarrow$ spectra $\sim T_{\text{eff}}$
- Different slope $\Longrightarrow$ different PSF (chromaticity!)

Angular resolution: $\theta \sim \frac{\lambda}{D}$

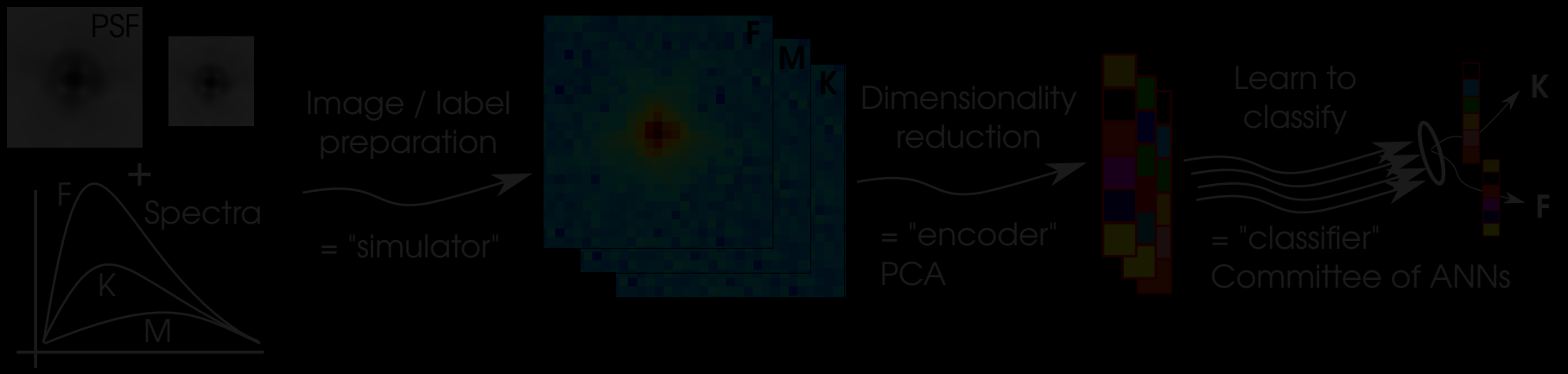
Classifying stars

Training Phase

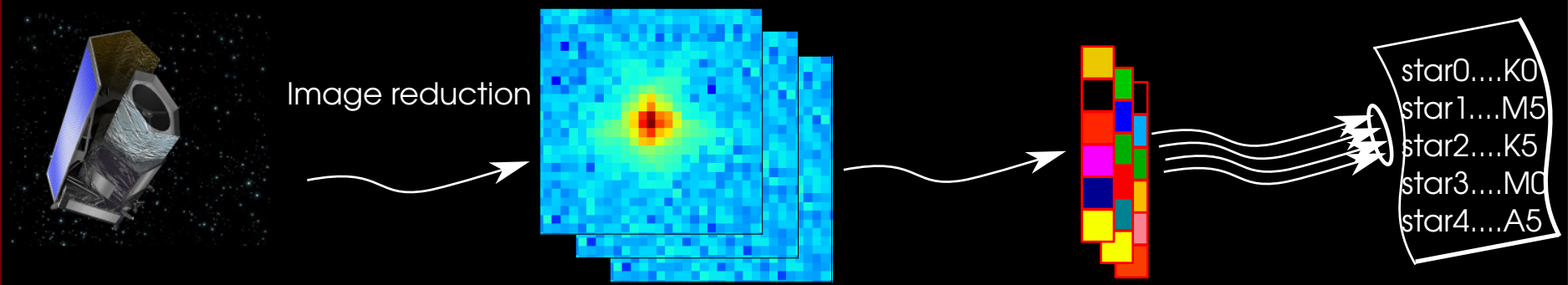


Classifying stars

Training Phase



Predicting / Testing phase



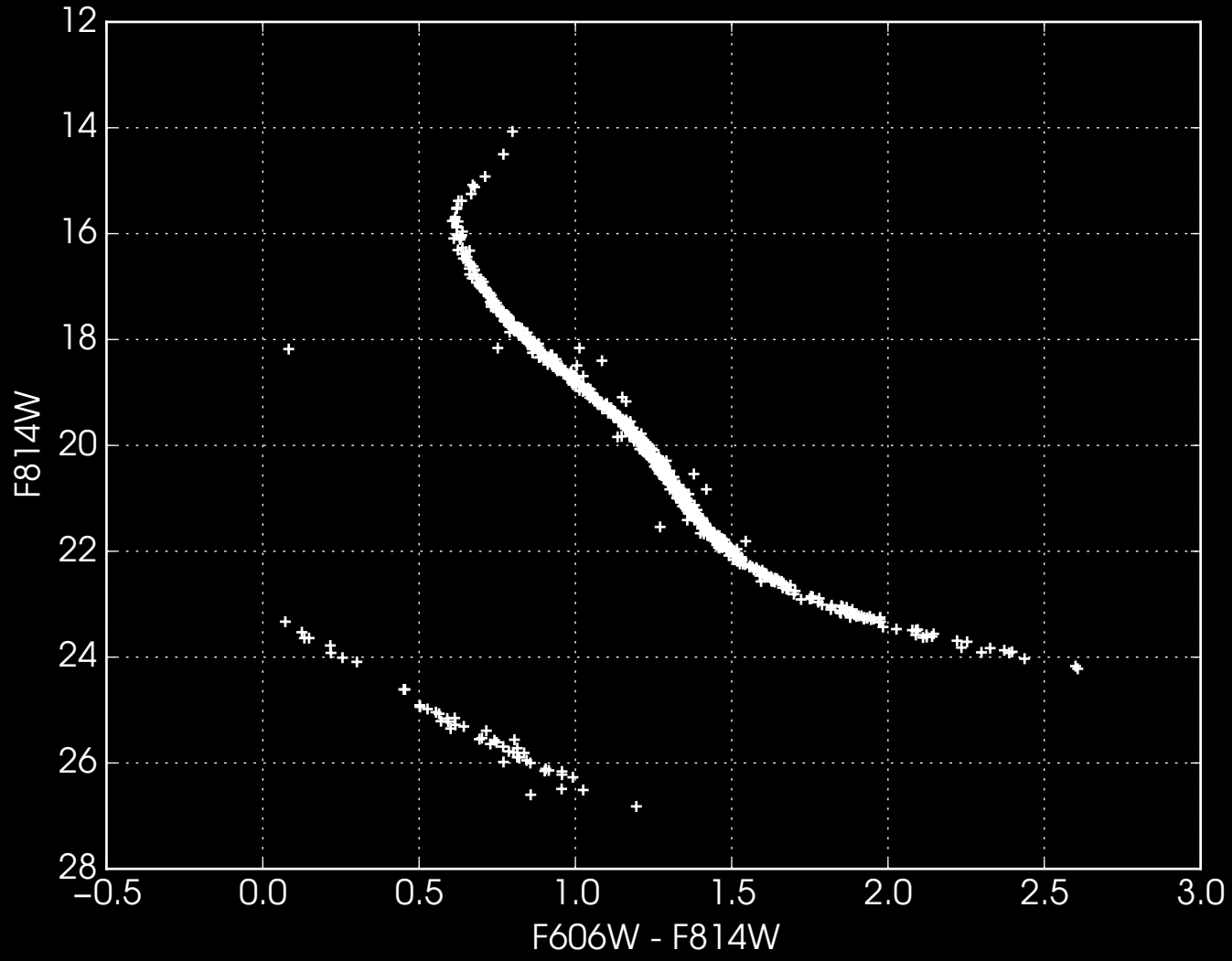
Data acquisition / preparation

Observations / Simulations

Compressed data

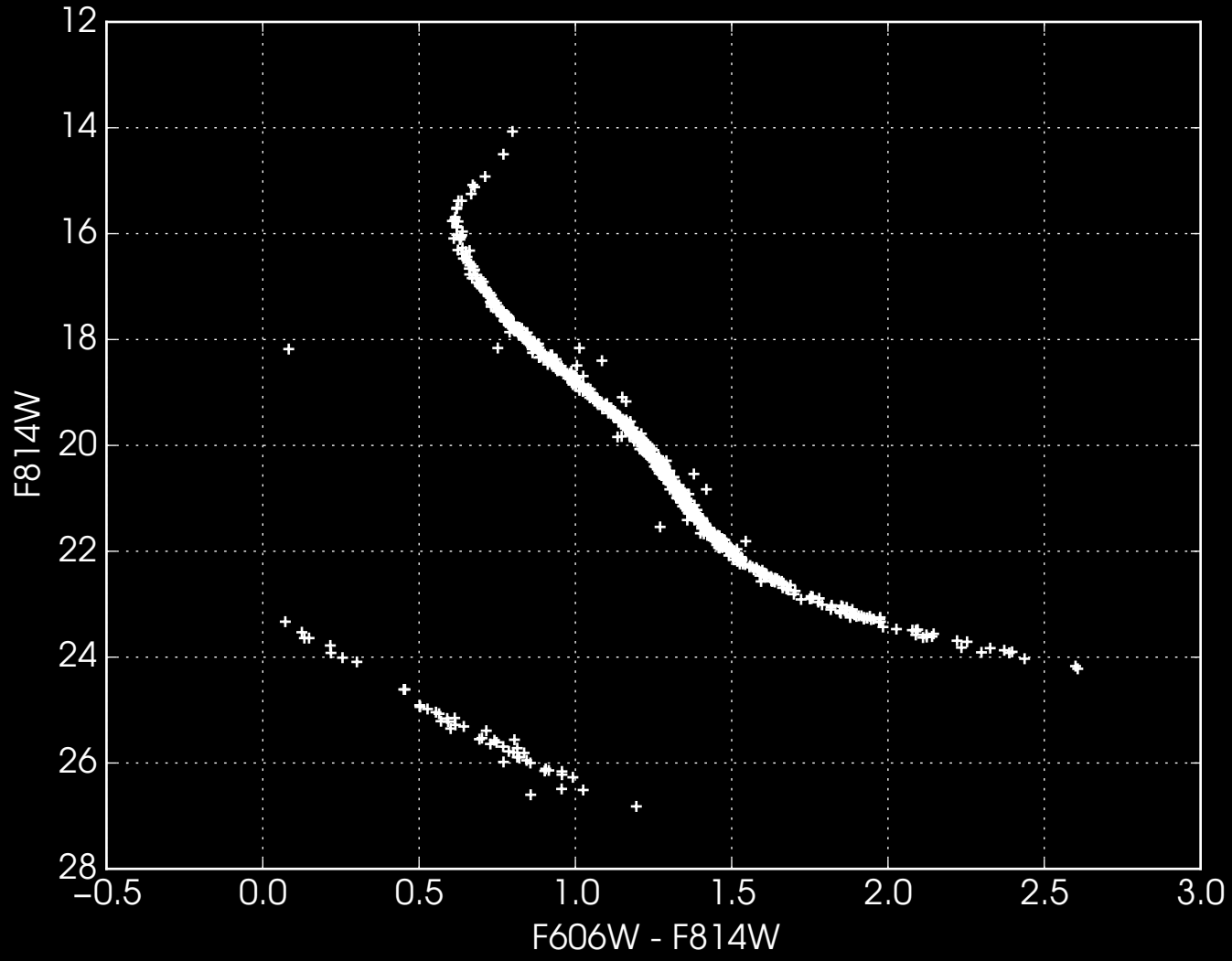
Classification

Colour-Magnitude Diagram



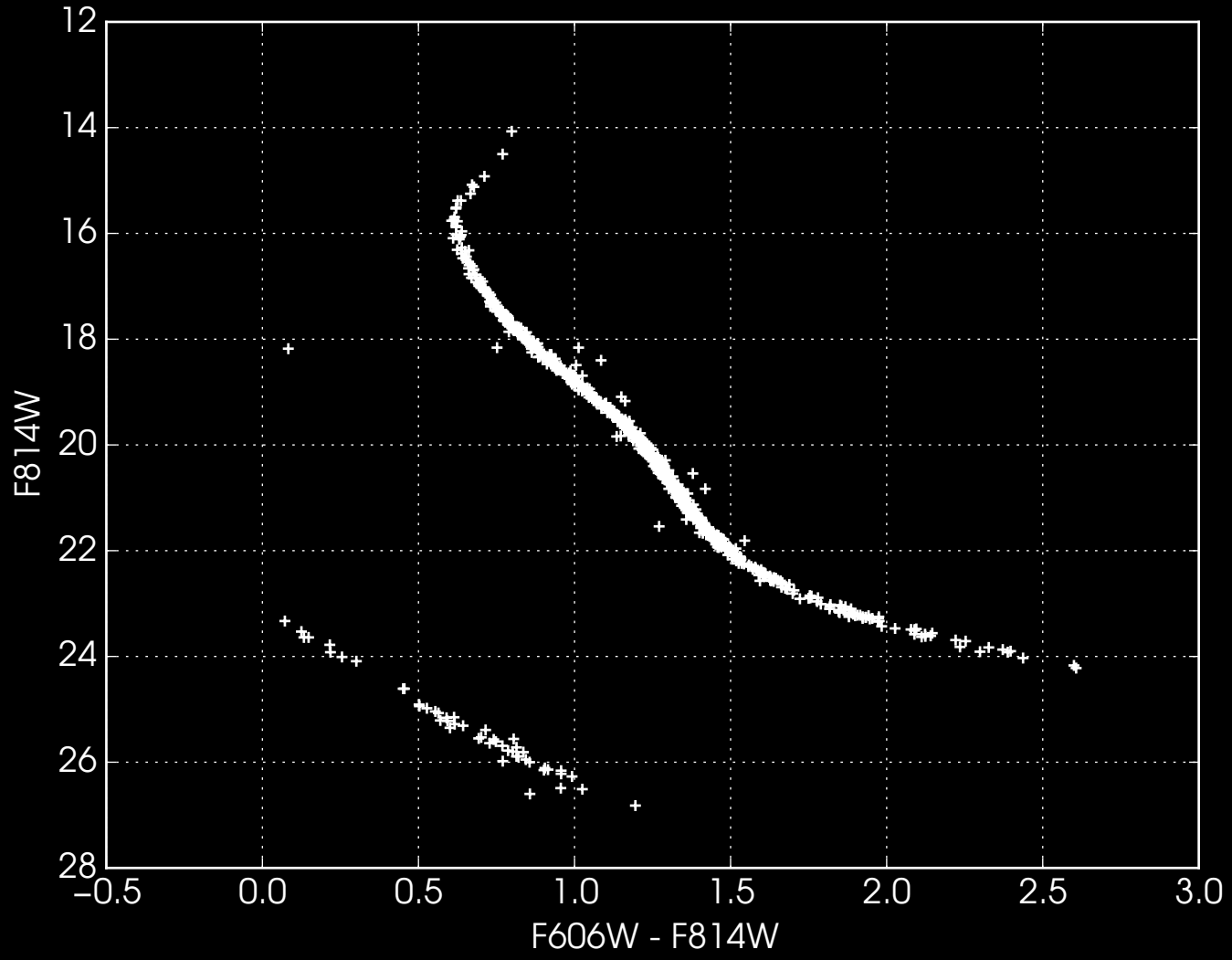
- Globular cluster, age 12.0 ± 0.8 Gyr, 2nd closest to $\odot$
- Richer+08: *HST* ACS for 126 orbits, F814W, F606W
- Goal: Reproduce colour from {F814W image + magnitude} or {image}

Colour-Magnitude Diagram



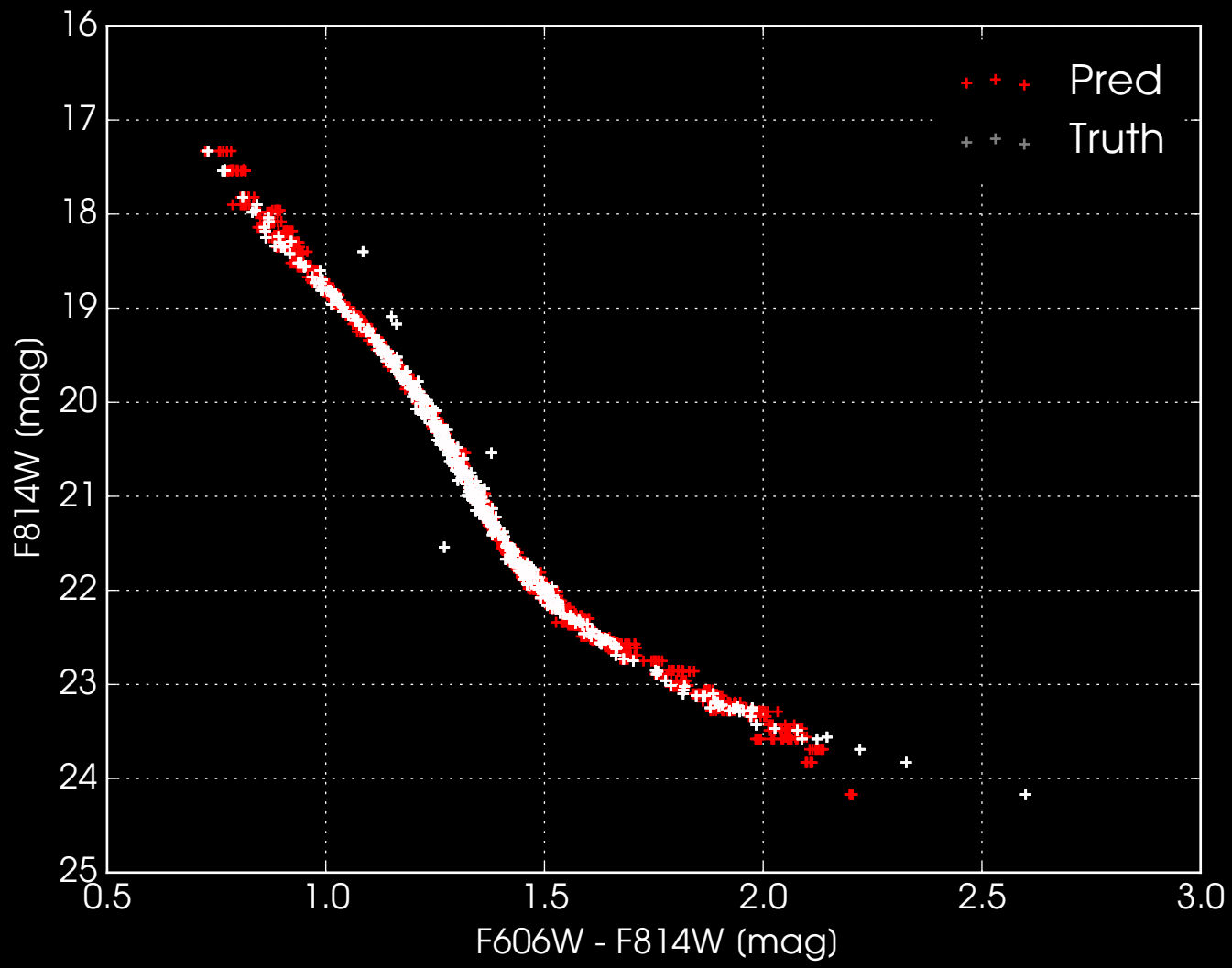
- Globular cluster, age 12.0 ± 0.8 Gyr, 2nd closest to $\odot$
- Richer+08: *HST* ACS for 126 orbits, F814W, F606W
- Goal: Reproduce colour from {F814W image + magnitude} or {image}

Colour-Magnitude Diagram



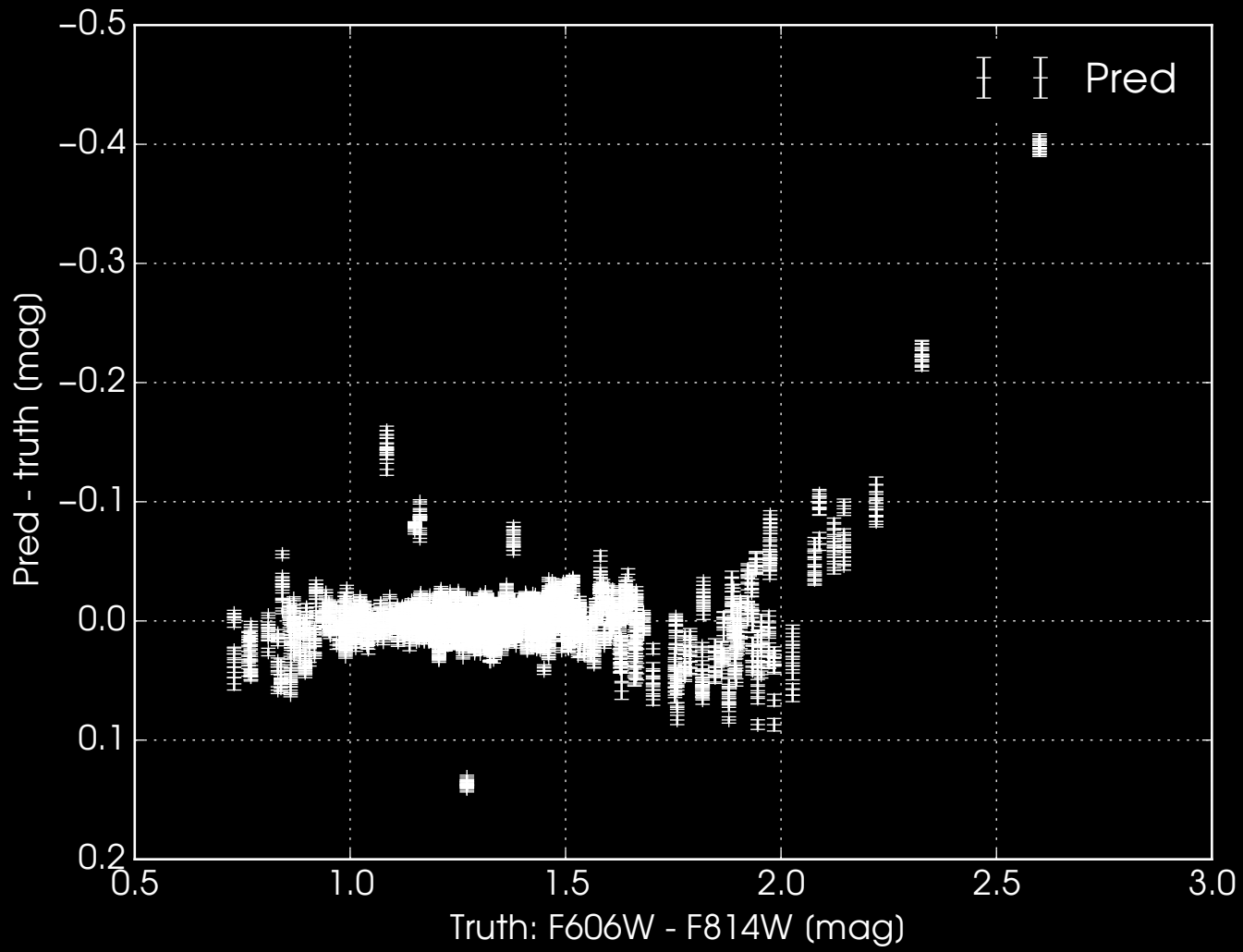
- Globular cluster, age 12.0 ± 0.8 Gyr, 2nd closest to $\odot$
- Richer+08: *HST* ACS for 126 orbits, F814W, F606W
- Goal: Reproduce colour from {F814W image + magnitude} or {image}

Colour-Magnitude Diagram



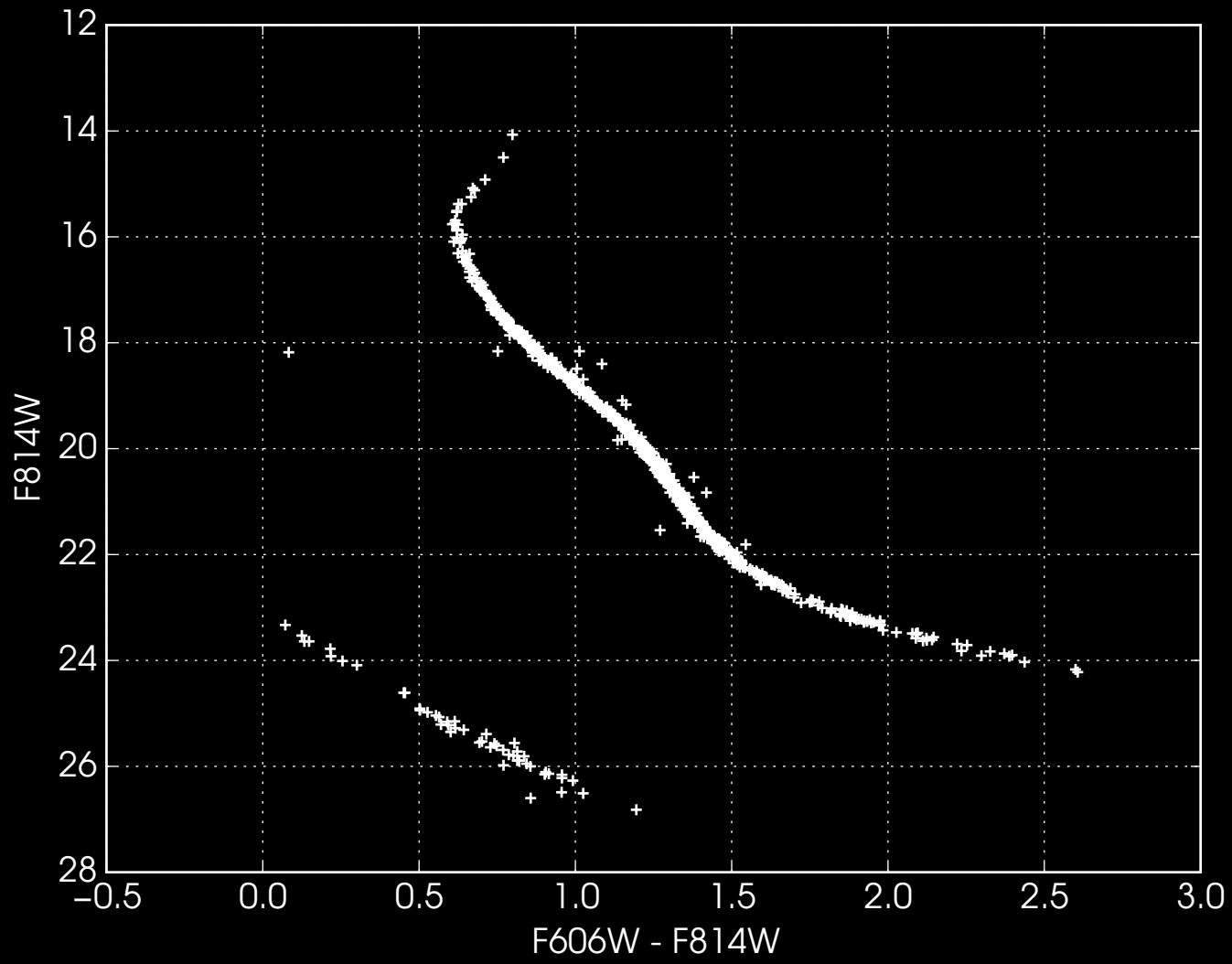
- Globular cluster, age 12.0 ± 0.8 Gyr, 2nd closest to $\odot$
- Richer+08: *HST* ACS for 126 orbits, F814W, F606W
- Goal: Reproduce colour from {F814W image + magnitude} or {image}

Colour-Magnitude Diagram



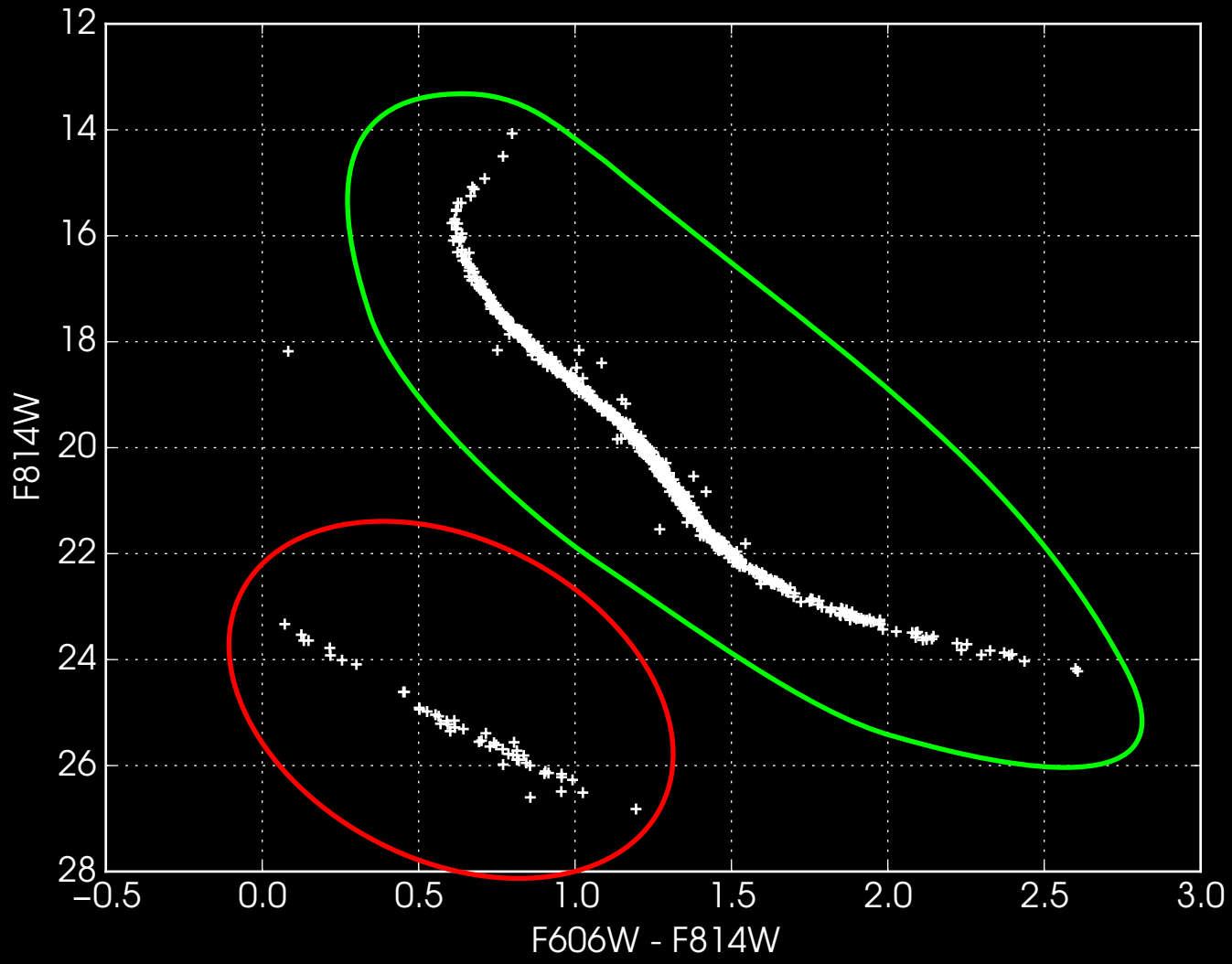
- Globular cluster, age 12.0 ± 0.8 Gyr, 2nd closest to $\odot$
- Richer+08: *HST* ACS for 126 orbits, F814W, F606W
- Goal: Reproduce colour from {F814W image + magnitude} or {image}

Dealing with non-MS stars



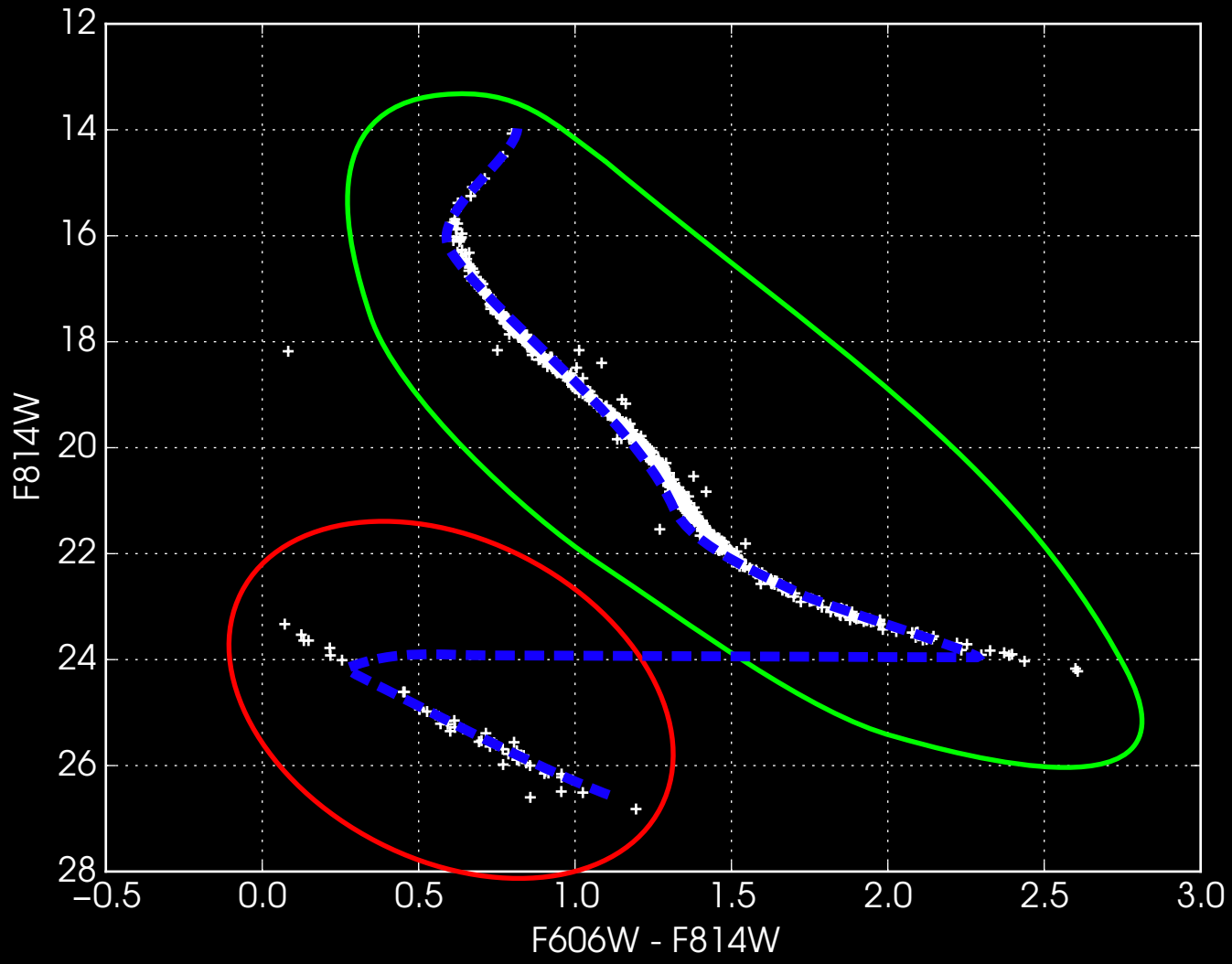
- Generate two regressions
- Identify non-MS for removal

Dealing with non-MS stars



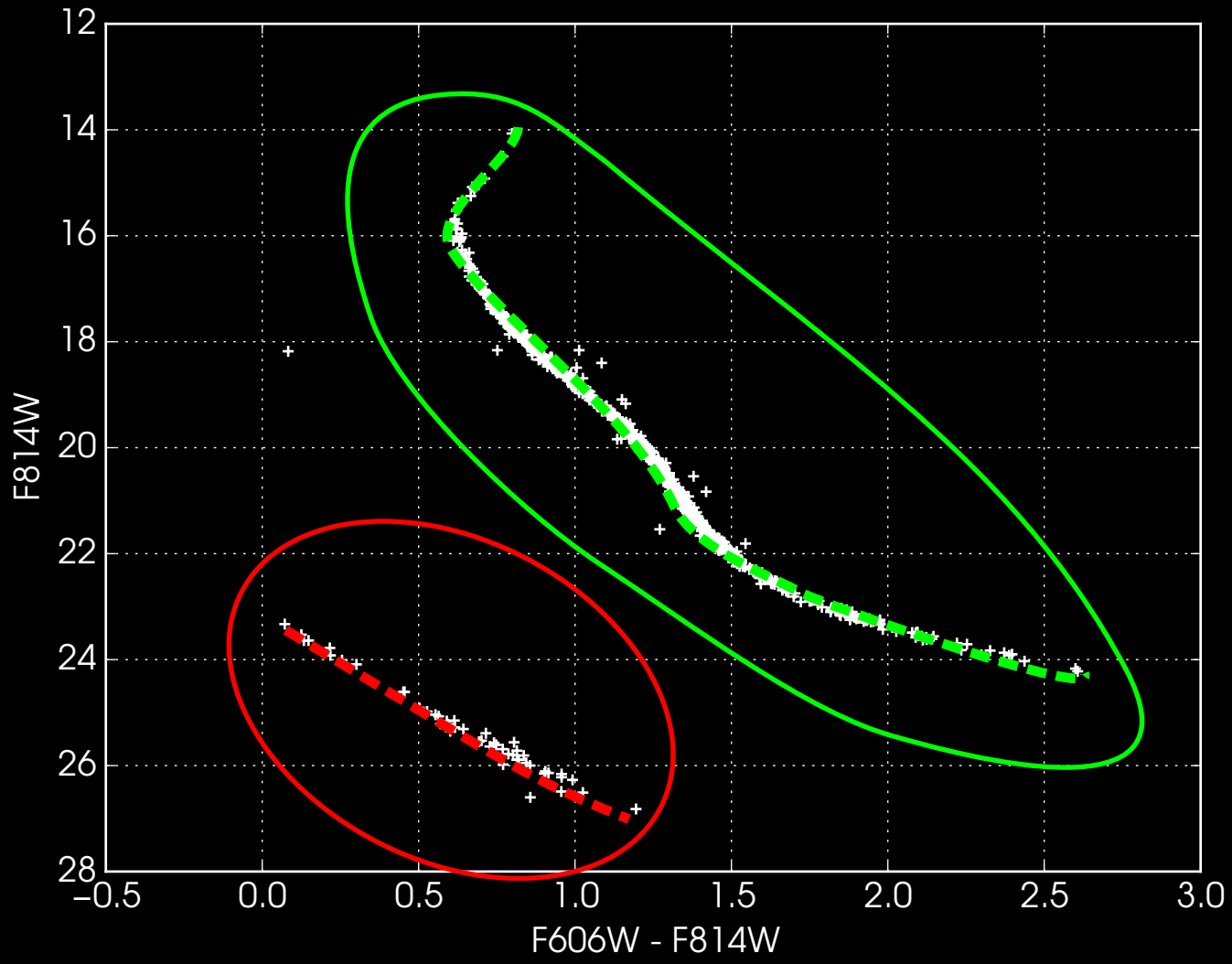
- Generate two regressions
- Identify non-MS for removal

Dealing with non-MS stars



- Generate two regressions
- Identify non-MS for removal

Dealing with non-MS stars



- Generate two regressions
- Identify non-MS for removal

Summary NGC 6397

- Creating a CMD for NGC 697 using *HST* F814W data
- ✓ Removing WD by hand creates great fit to the actual CMD
- ✓ Inputs: {imagette of star} and {imagette of star + magnitude}
- Work in progress:
  - Removal of WD using shape of PSF $\longrightarrow$ concept proven in Kuntzer+16b
  - Generate two regressions (?)
  - Using more data (only $\sim$ 500/2300 stars currently used)
  - Training on simulated data
- Software on https://github.com/kuntzer/sclas

Summary NGC 6397

- Creating a CMD for NGC 697 using *HST* F814W data

✓ Removing WD by hand creates great fit to the actual CMD

✓ Inputs: {imagette of star} and {imagette of star + magnitude}

- Work in progress:
  - Removal of WD using shape of PSF $\longrightarrow$ concept proven in Kuntzer+16b
  - Generate two regressions (?)
  - Using more data (only $\sim$ 500/2300 stars currently used)
  - Training on simulated data
- Software on https://github.com/kuntzer/sclas

Summary NGC 6397

- Creating a CMD for NGC 697 using *HST* F814W data
- ✓ Removing WD by hand creates great fit to the actual CMD
- ✓ Inputs: {imagette of star} and {imagette of star + magnitude}
- Work in progress:
  - Removal of WD using shape of PSF $\longrightarrow$ concept proven in Kuntzer+16b
  - Generate two regressions (?)
  - Using more data (only $\sim$ 500/2300 stars currently used)
  - Training on simulated data
- Software on https://github.com/kuntzer/sclas

Summary NGC 6397

- Creating a CMD for NGC 697 using *HST* F814W data
- ✓ Removing WD by hand creates great fit to the actual CMD
- ✓ Inputs: {imagette of star} and {imagette of star + magnitude}
- Work in progress:
  - Removal of WD using shape of PSF $\longrightarrow$ concept proven in Kuntzer+16b
  - Generate two regressions (?)
  - Using more data (only $\sim$ 500/2300 stars currently used)
  - Training on simulated data
- Software on https://github.com/kuntzer/sclas

PSF Interpolation

using Auto-Encoders and manifold learning

SDSS

PSF requirements and measurement

- Measured shear power spectrum

$$\hat{C} \approx (1 + \mathcal{M})C + \mathcal{A}, \qquad \mathcal{M}, \mathcal{A} \sim \langle |\delta e_{\rm PSF}|^2 \rangle, \langle \delta R^2_{\rm PSF} \rangle$$

- Science requirement for the PSF shape:

size rms (relative)	$\frac{\sigma(R^2)}{\langle R^2 \rangle} \leq 10^{-3}$
ellipticity rms	$\sigma(e_i) \leq 10^{-4}$
ellipticity	$e_i < 0.15$

From Paulin-Henriksson+2008

- Performance needed for measurement

size	$\sim 5\%$
ellipticity	$\sim 1\%$

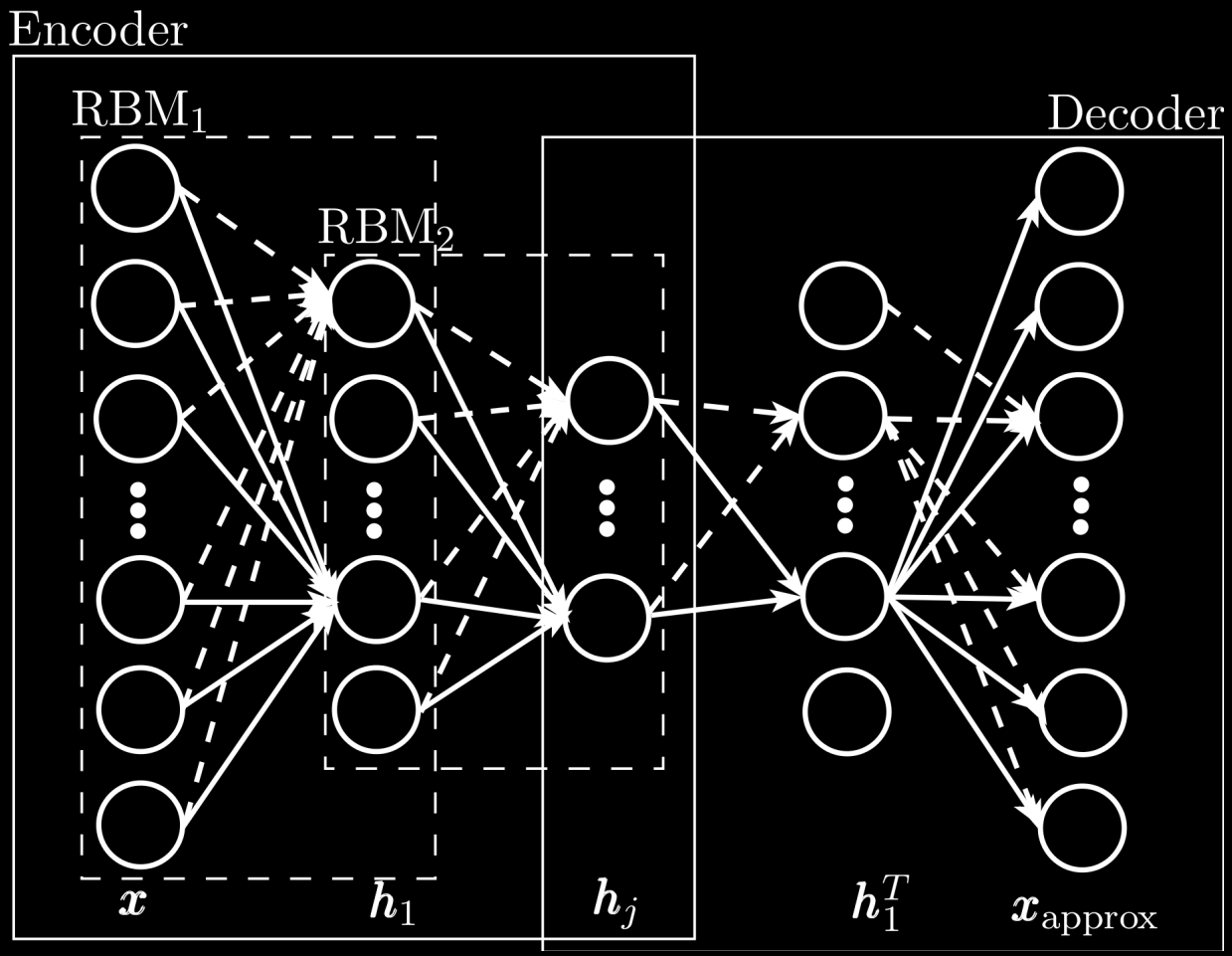
Auto-Encoders

1. Encode the PSF into a few coefficients

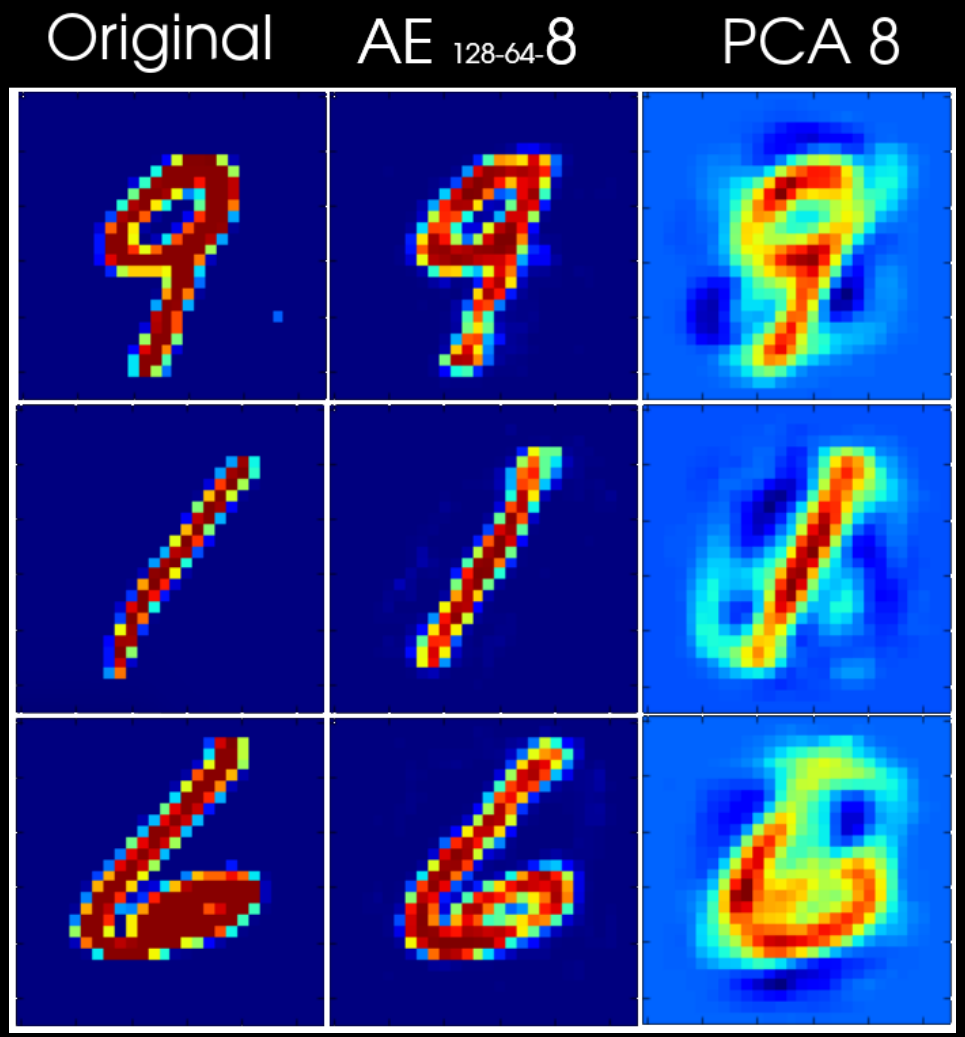
2. Interpolate the coefficients

Auto-Encoders

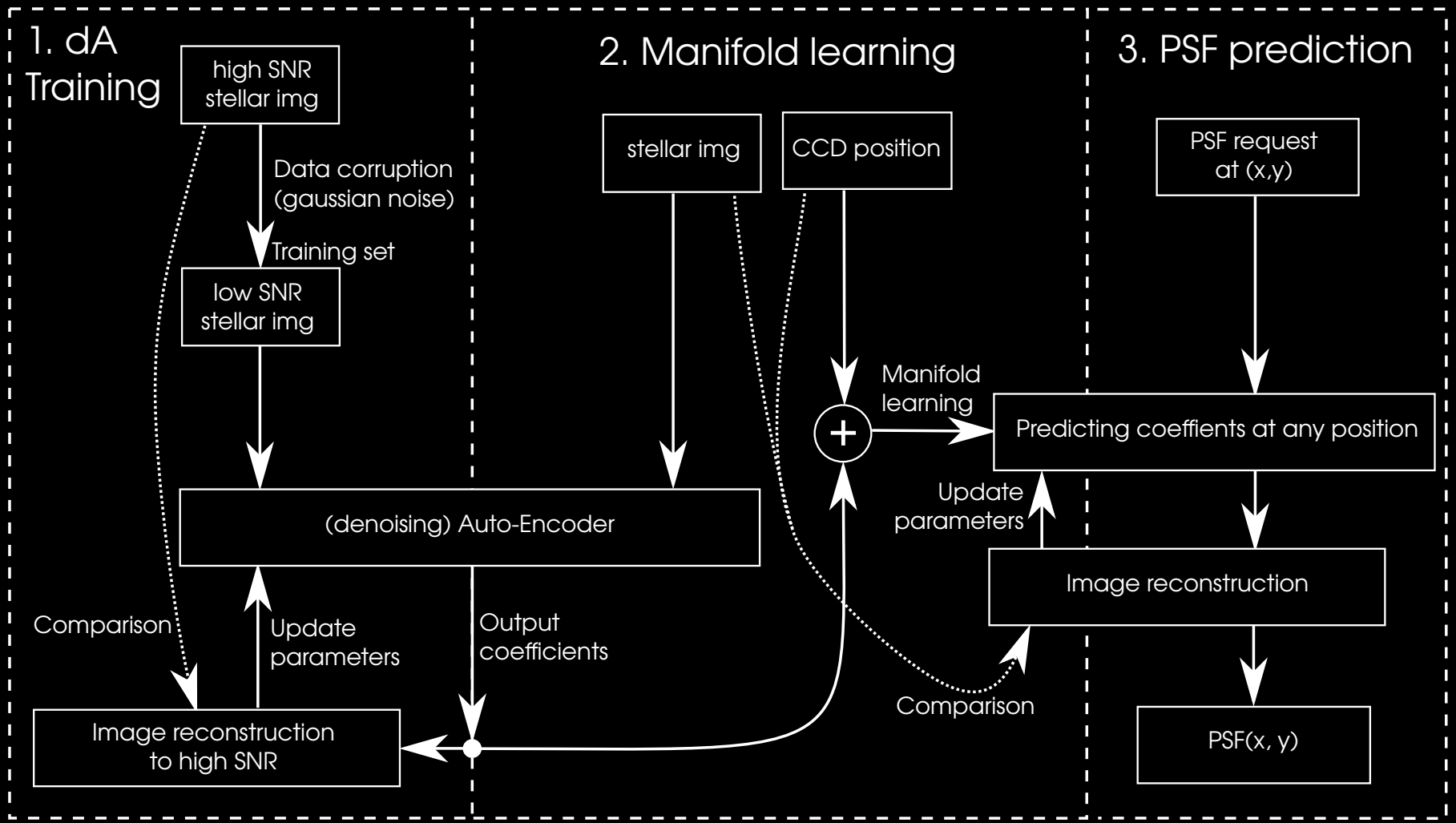
1. Encode the PSF into a few coefficients
2. Interpolate the coefficients



Auto-Encoder performance



Interpolating PSFs



Summary

Summary

PDF available at http://obswww.unige.ch/∼kuntzer/talks/cea_psf_kuntzer.pdf

- Unresolved binaries can be found
  - Unresolved multiples alter the PSF
  - Most of Euclid's binaries have small angular separations
  - Effects on PSF determination is $\mathcal{O}(10^{-4})$
  - "Binary finder" based on systematic biases
  - Auto-correlation, random forests and Bayesian approach (?)

- NGC 6397 Colour-Magnitude diagramme
  - Great fit to the Main Sequence
  - Inputs: {imagette of star} and {imagette of star + magnitude}
  - Can we remove non-MS stars and training on simulations?

- Auto-encoders for PSF interpolation
  - Relatively new and promising ML technique
  - Denoising almost for free
  - Seems to work well on handwritten numbers, harder on stellar images
  - Software on https://github.com/kuntzer/pylae

Summary

PDF available at http://obswww.unige.ch/~kuntzer/talks/cea_psf_kuntzer.pdf

- Unresolved binaries can be found
  - Unresolved multiples alter the PSF
  - Most of Euclid's binaries have small angular separations
  - Effects on PSF determination is $\mathcal{O}(10^{-4})$
  - "Binary finder" based on systematic biases
  - Auto-correlation, random forests and Bayesian approach (?)

- NGC 6397 Colour-Magnitude diagramme
  - Great fit to the Main Sequence
  - Inputs: {imagette of star} and {imagette of star + magnitude}
  - Can we remove non-MS stars and training on simulations?

- Auto-encoders for PSF interpolation
  - Relatively new and promising ML technique
  - Denoising almost for free
  - Seems to work well on handwritten numbers, harder on stellar images
  - Software on https://github.com/kuntzer/pylae

Summary

PDF available at http://obswww.unige.ch/∼kuntzer/talks/cea_psf_kuntzer.pdf

- Unresolved binaries can be found
  - Unresolved multiples alter the PSF
  - Most of Euclid's binaries have small angular separations
  - Effects on PSF determination is $\mathcal{O}(10^{-4})$
  - "Binary finder" based on systematic biases
  - Auto-correlation, random forests and Bayesian approach (?)

- NGC 6397 Colour-Magnitude diagramme
  - Great fit to the Main Sequence
  - Inputs: {imagette of star} and {imagette of star + magnitude}
  - Can we remove non-MS stars and training on simulations?

- Auto-encoders for PSF interpolation
  - Relatively new and promising ML technique
  - Denoising almost for free
  - Seems to work well on handwritten numbers, harder on stellar images
  - Software on https://github.com/kuntzer/pylae

Summary

PDF available at http://obswww.unige.ch/∼kuntzer/talks/cea_psf_kuntzer.pdf

- Unresolved binaries can be found
  - Unresolved multiples alter the PSF
  - Most of Euclid's binaries have small angular separations
  - Effects on PSF determination is $\mathcal{O}(10^{-4})$
  - "Binary finder" based on systematic biases
  - Auto-correlation, random forests and Bayesian approach (?)
- NGC 6397 Colour-Magnitude diagramme
  - Great fit to the Main Sequence
  - Inputs: {imagette of star} and {imagette of star + magnitude}
  - Can we remove non-MS stars and training on simulations?
- Auto-encoders for PSF interpolation
  - Relatively new and promising ML technique
  - Denoising almost for free
  - Seems to work well on handwritten numbers, harder on stellar images
  - Software on https://github.com/kuntzer/pylae

MS stars and filters