Astronomical Surveys: from SDSS to LSST via HSC

Robert Lupton

Princeton University

GREAT Workshop, La Palma, 1st June, 2011

Outline

What's Special about Surveys?

SDSS: Inside the Sausage Factory

HSC: Hyper-SuprimeCam

LSST: the Large Survey Synoptic Telescope

A Brief Summary of Sloan

What's Special about Surveys?

Quality Control

One of the things that we teach our students (and post-docs) is how to look carefully at a set of facts and ask if they makes sense.

In the case of theory, this means asking exactly what the New Discovery depends on, and whether its foundations are sound

In the case of data, this means asking if the Fascinating Result du Jour is an artifact of the instrument or of the reduction.

This is difficult in the context of a survey:

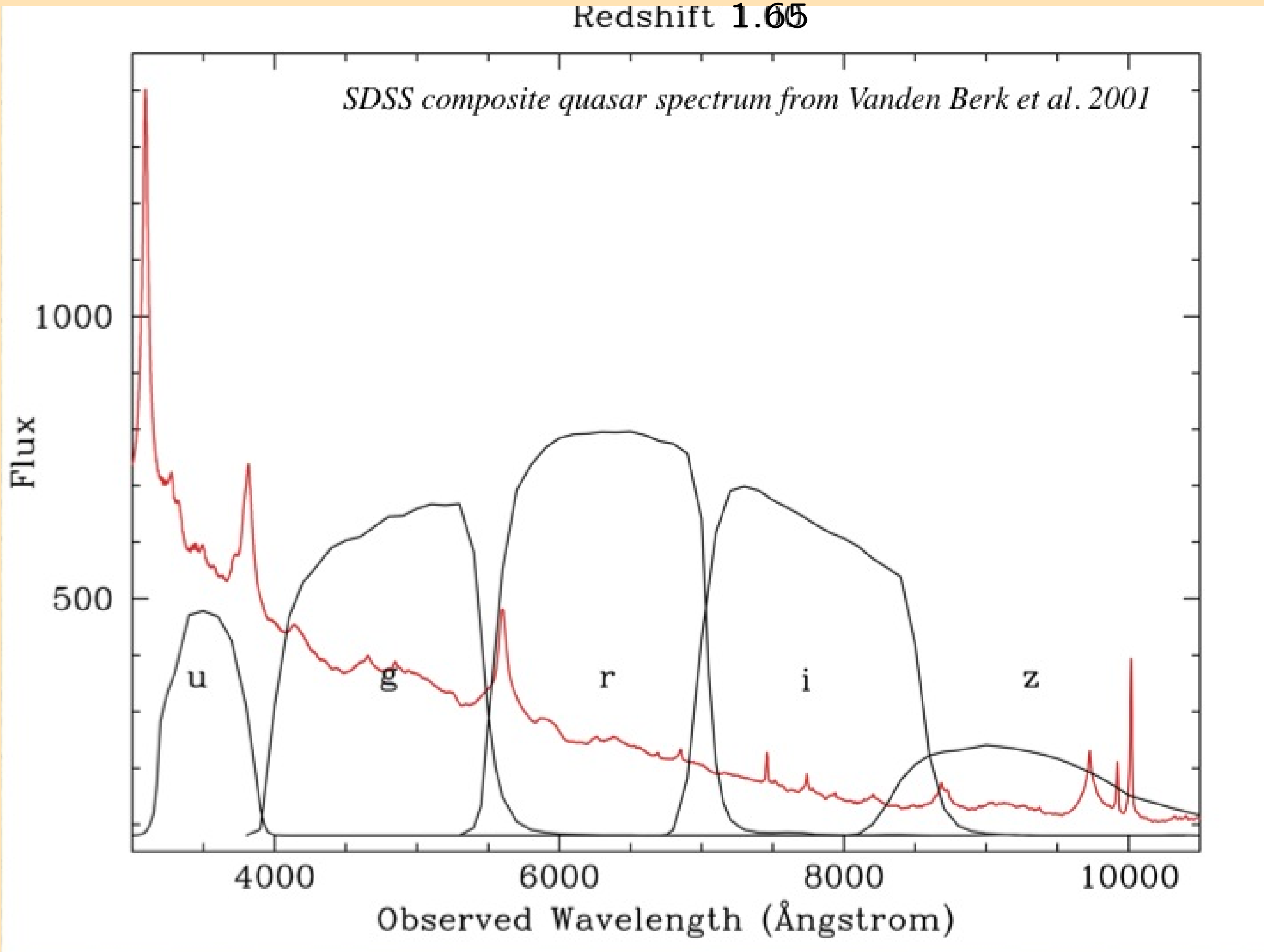
- There's too much data for humans to look at
- The consumer is far removed from the raw data
- Large datasets make it possible to study rare events; glitches look like rare events

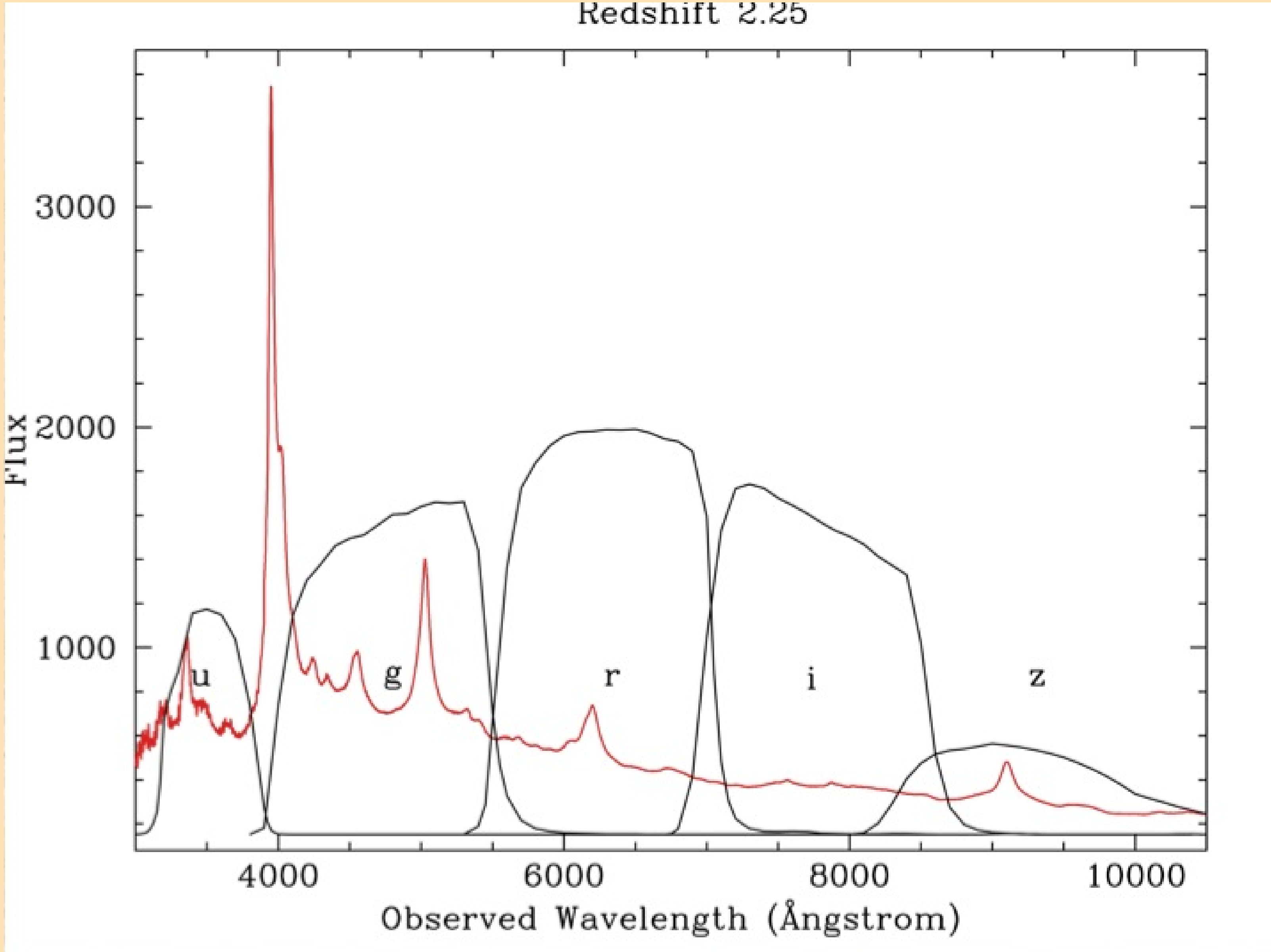
Then there's the problem of how to let the astronomical public what they should trust and where they should tread warily.

Inside the Sausage Machine

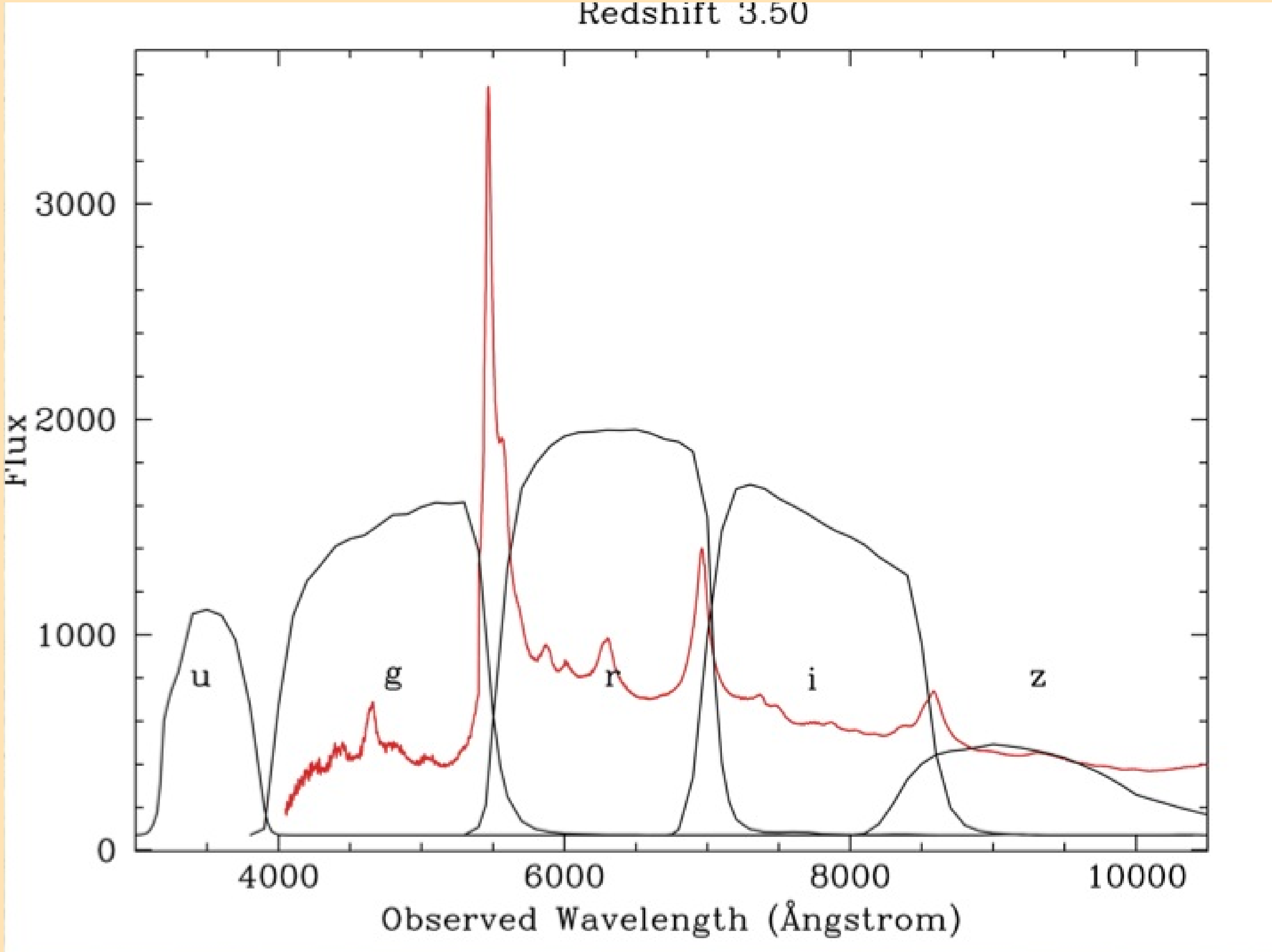
Example: Finding $z \sim 6$ Quasars

How does a photometric survey find Quasars?

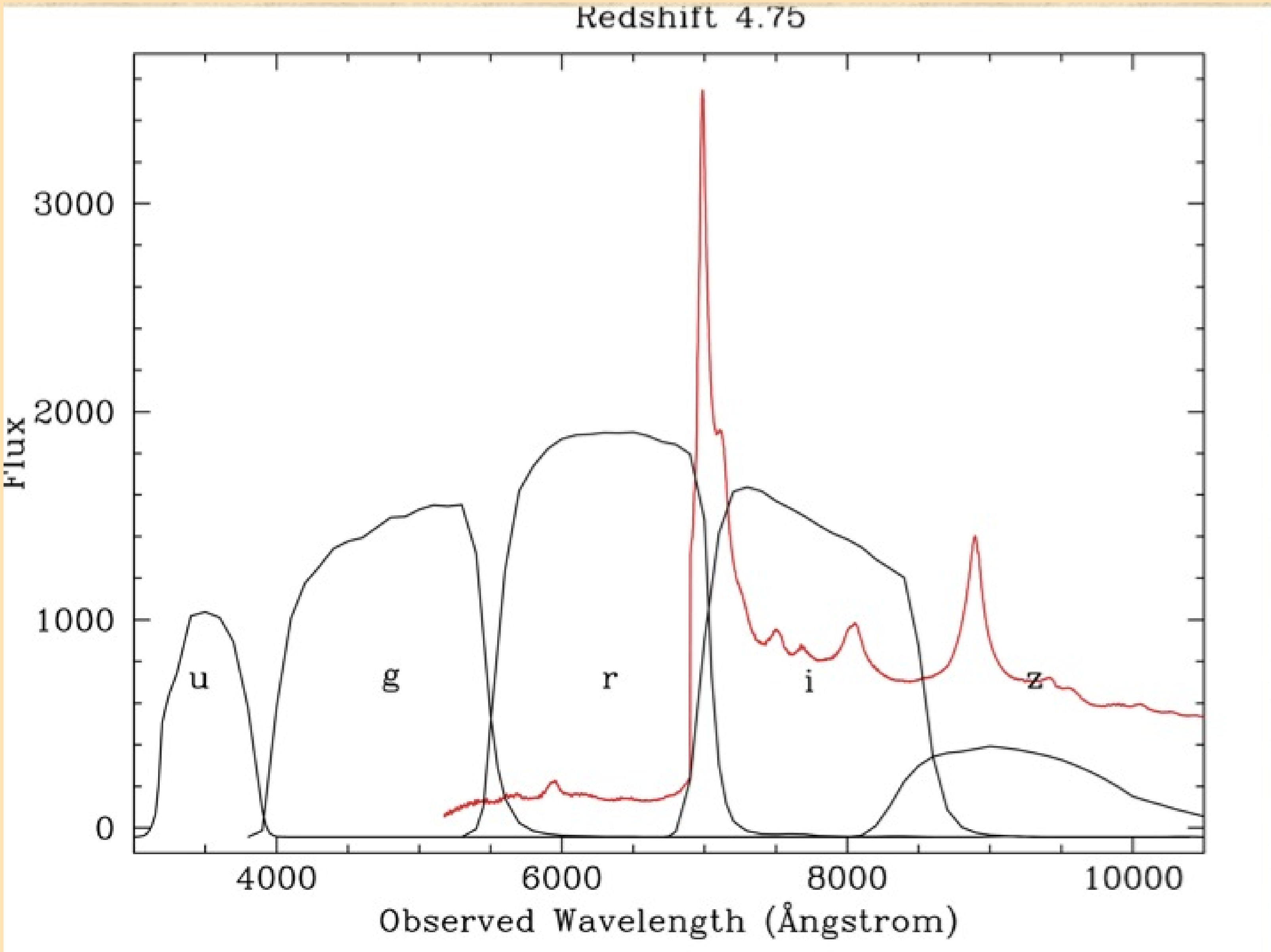

Redshift 1.65
SDSS composite quasar spectrum from Vanden Berk et al. 2001
1000
500
Flux
u
g
r
i
z
4000
6000
8000
10000
Observed Wavelength (Ångstrom)

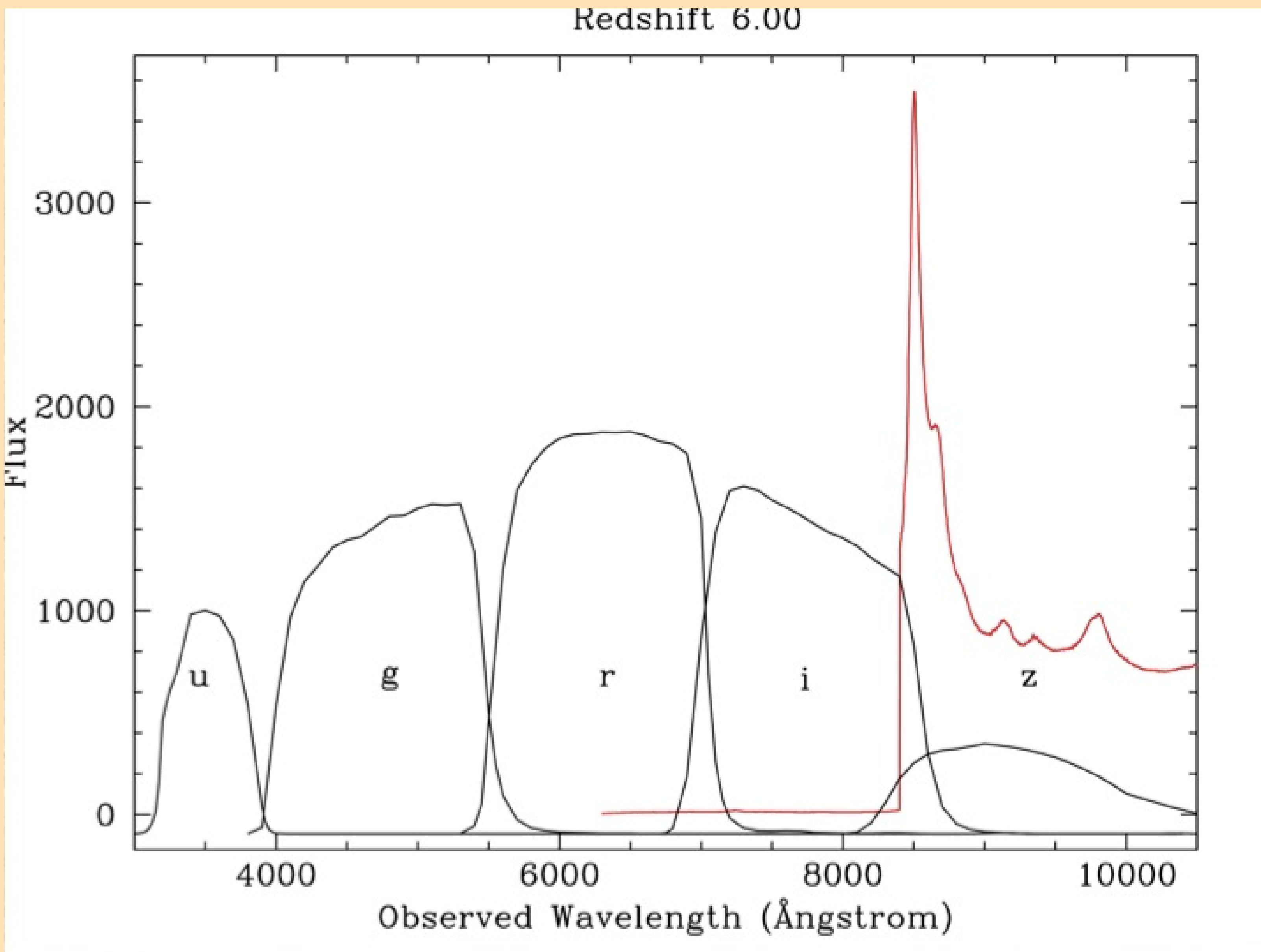


Redshift 2.25
Flux
3000
2000
1000
u
g
r
i
z
4000
6000
8000
10000
Observed Wavelength (Ångstrom)



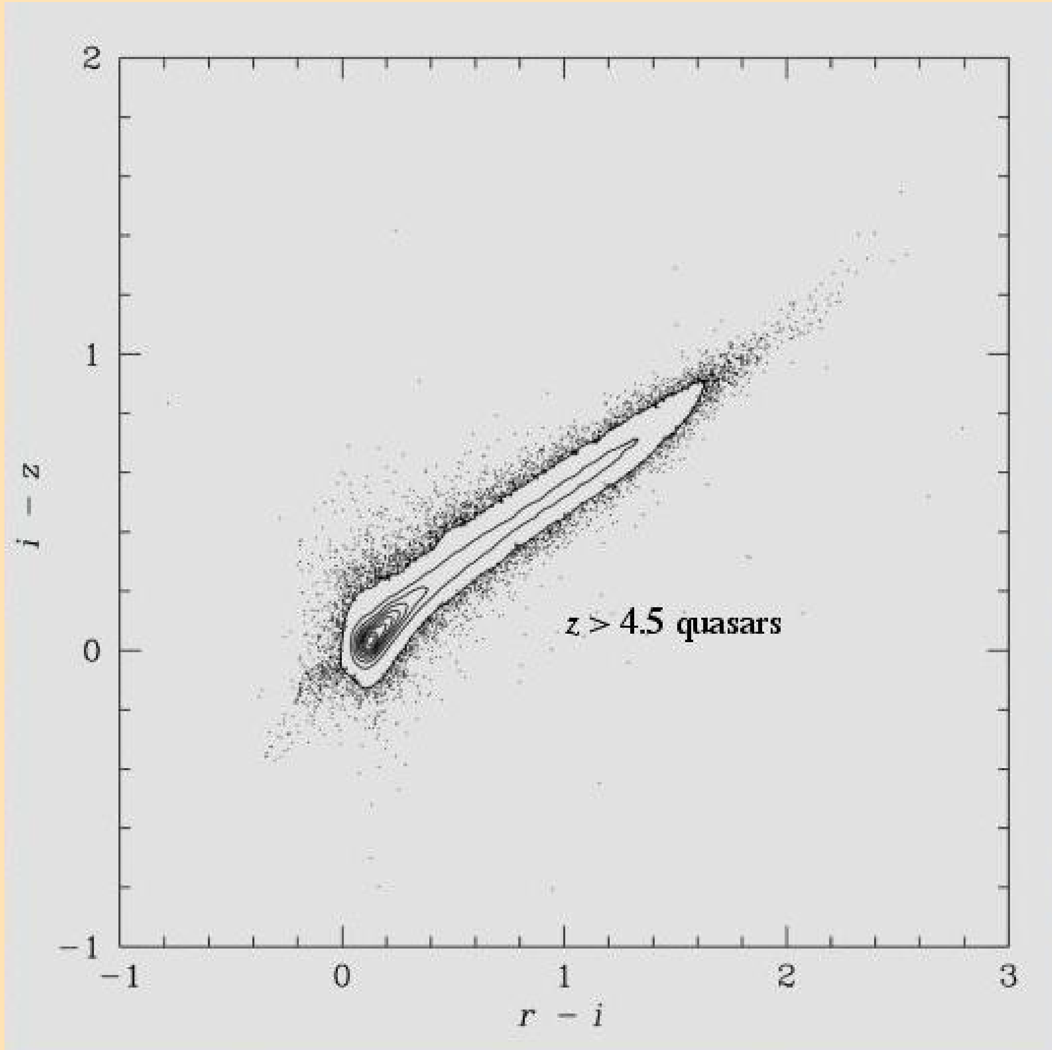
Redshift 3.50
Flux
3000
2000
1000
0
4000
6000
8000
10000
Observed Wavelength (Ångstrom)
u
g
r
i
z


Redshift 4.75
Flux
3000
2000
1000
0
4000
6000
8000
10000
Observed Wavelength (Ångstrom)
u
g
r
i
z



Redshift 6.00
Flux
3000
2000
1000
0
u
g
r
i
z
4000
6000
8000
10000
Observed Wavelength (Ångstrom)

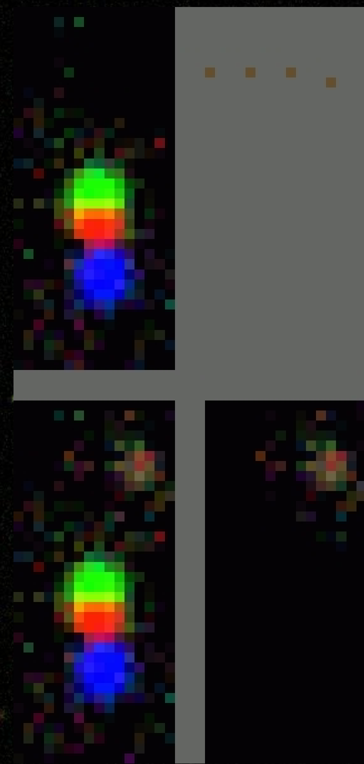
Colour Selection of Quasars

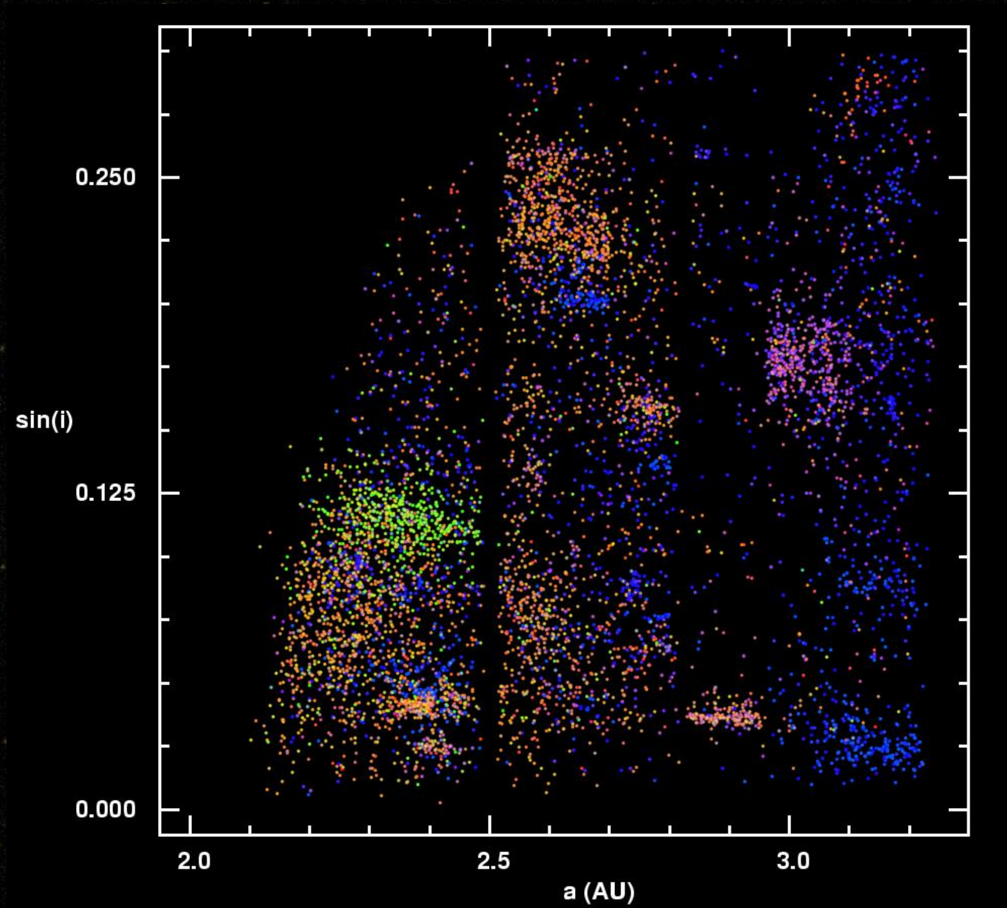


Why isn't this Easy?


Objects Are Blended

Objects Move





The semi-major axis v. (proper) inclination of a sample of known asteroids detected by SDSS

The PSF ϕ can be Complicated

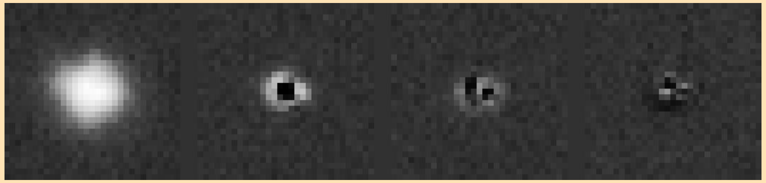
Two solutions:

- Normalise the seeing to some canonical form and value (cf. A&L image subtraction)
  - Involves some measure of deconvolution (or loss of S/N)
  - Slower, more complex code
- Estimate the seeing at the position of each object
  - Fast; a simple linear reconstruction at position of each object
  - The seeing is still variable across the frame

We chose the latter:

- KL decompose the bright stars in the frame, giving a number of basis functions (typically 3 or 4):

$$\phi = \sum_{\alpha=0}^{n-1} A^{(\alpha)} K^{(\alpha)}$$



$K^{(\alpha)}$

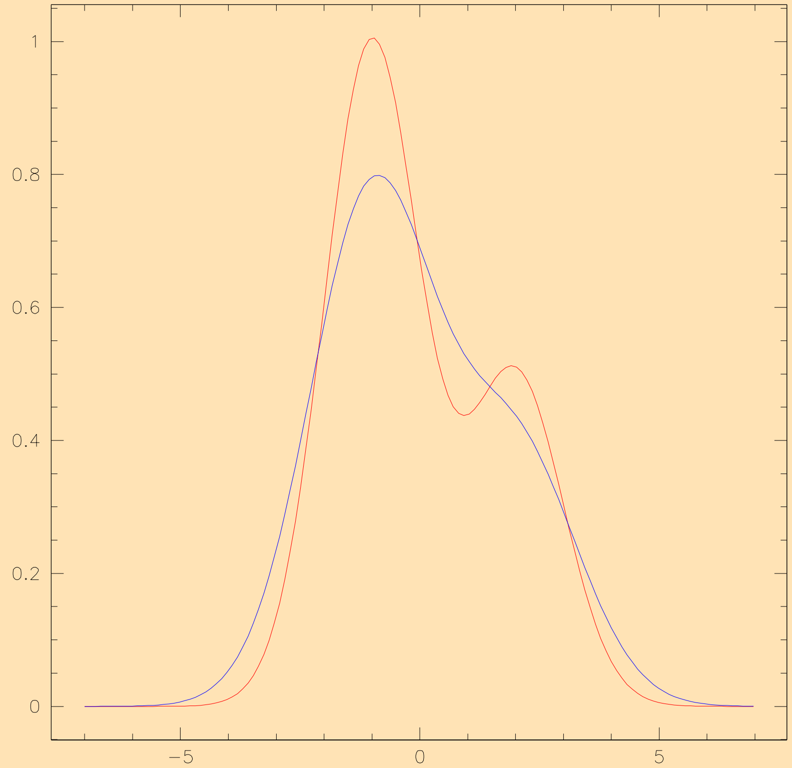
- Write the $A^{(\alpha)}$ as low-order polynomials in x, y:

$$\phi(x, y) = \sum_{\alpha=0}^{n-1} \sum_{r=0}^{n_r-1} \sum_{s=0}^{n_s-1} a_j^{(\alpha)} x^r y^s K^{(\alpha)}$$

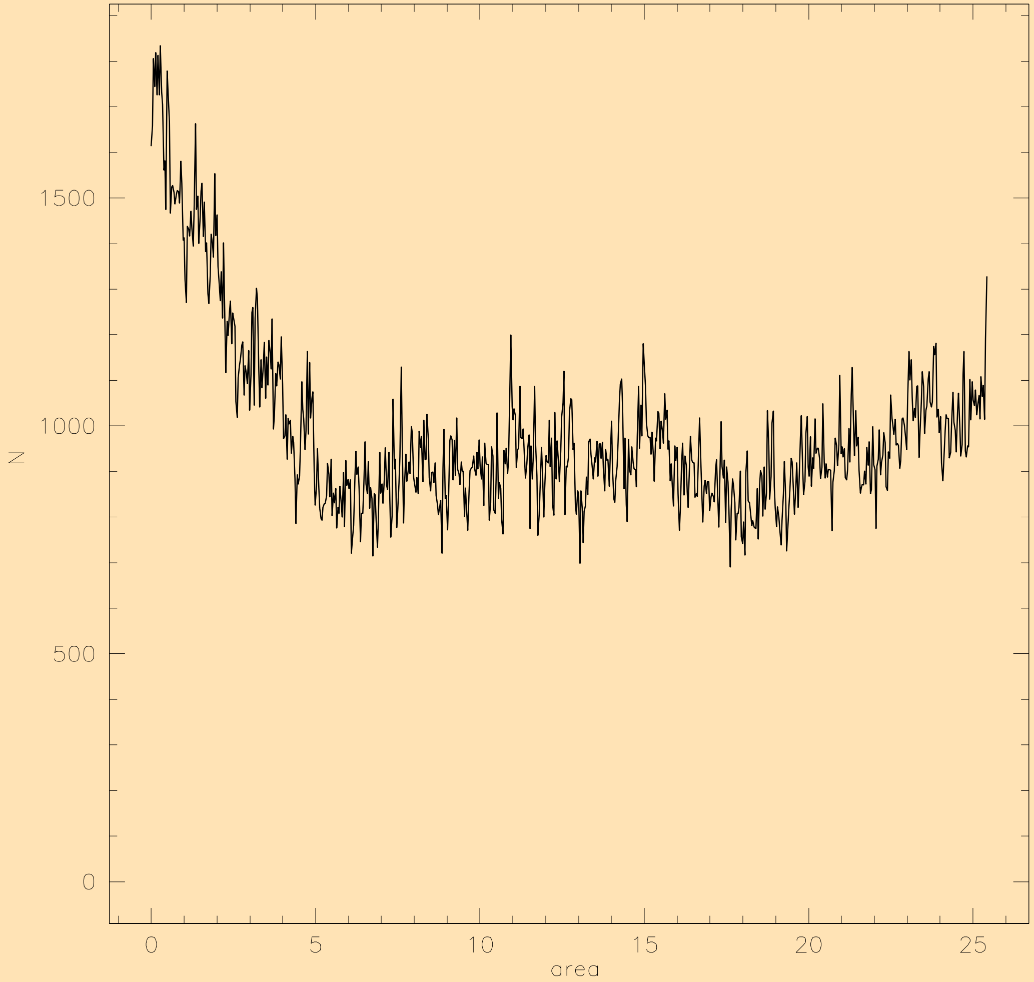
If you combine the last three points:

- blending
- moving
- variable seeing

it is not obvious how to build a catalogue out of a set of observations.

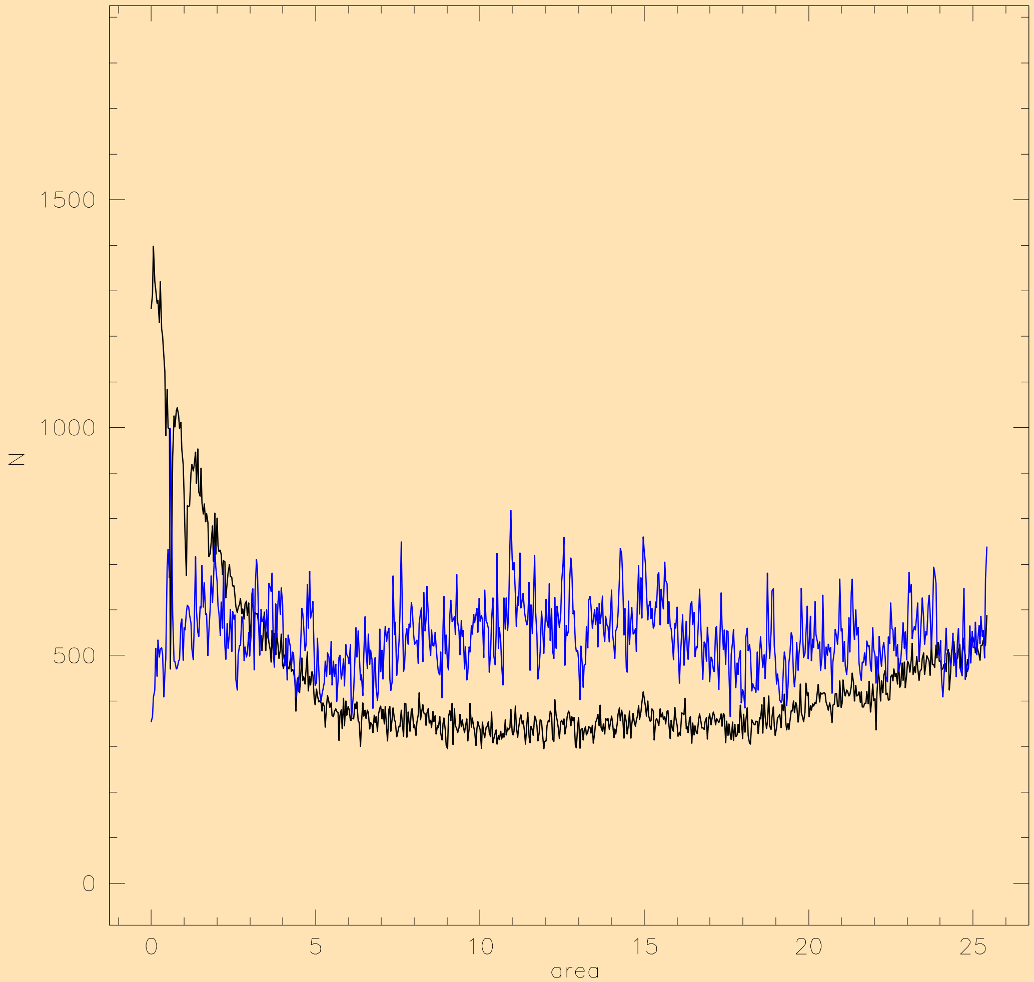


Not all Objects are Point Sources



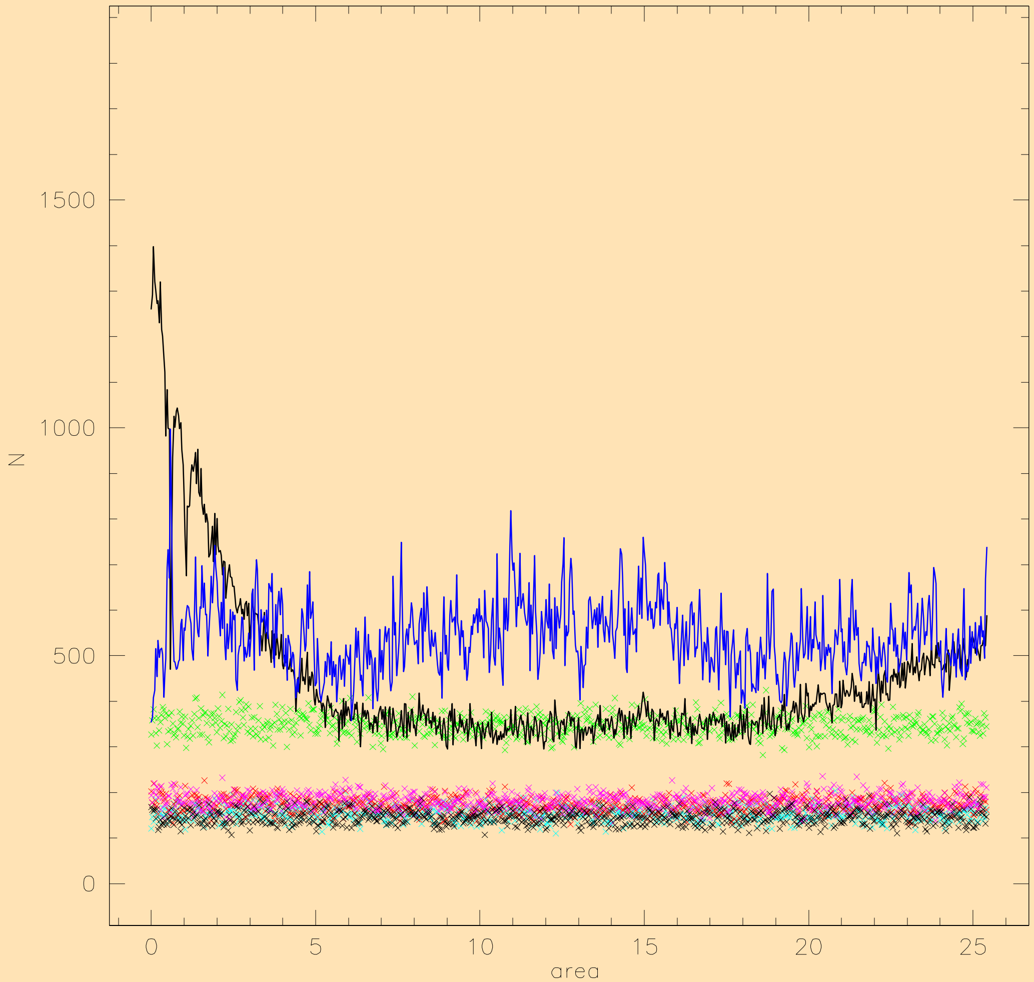
All Objects

Not all Objects are Point Sources



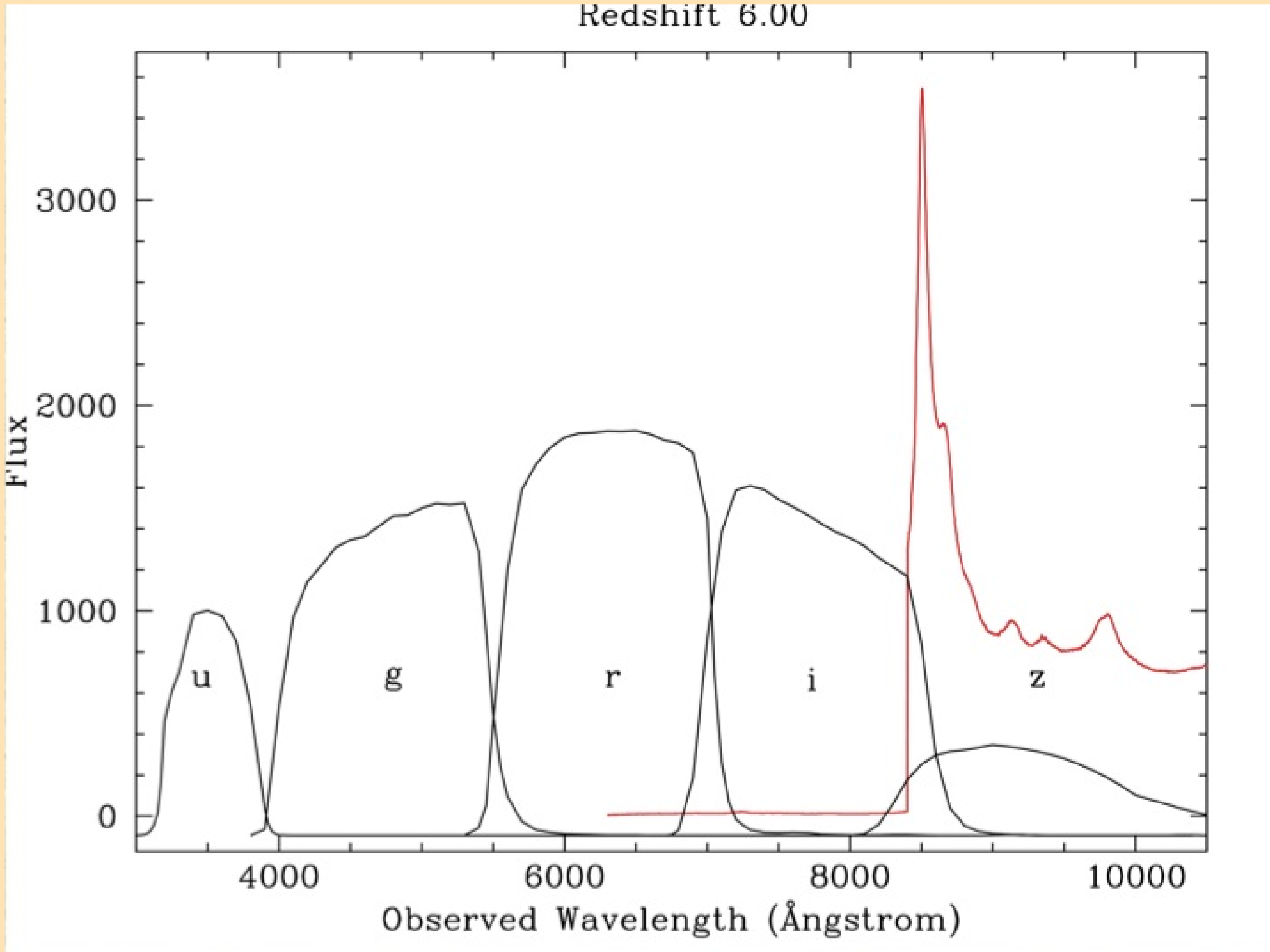
Stars and Galaxies

Not all Point Sources are Point Sources



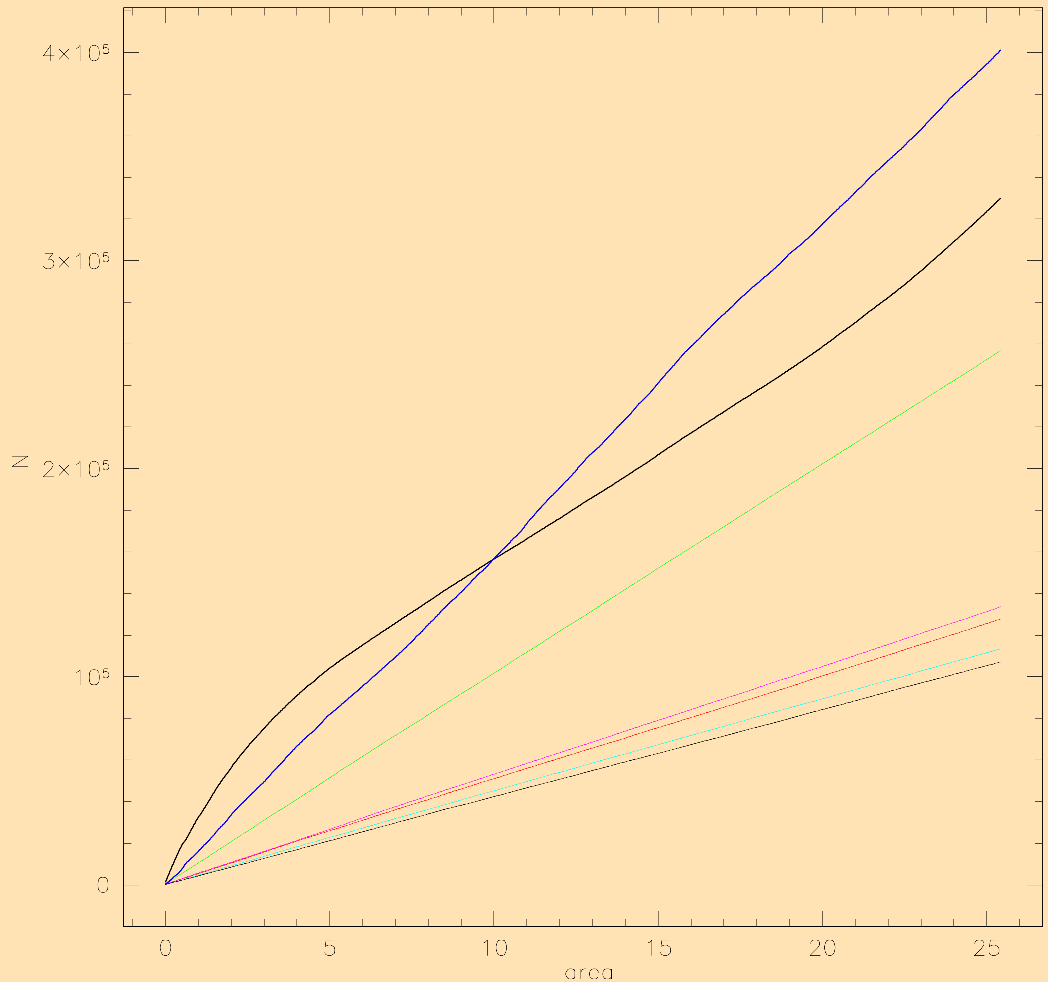
Stars and Galaxies and Cosmic Rays

$z \sim 6$ QSOs



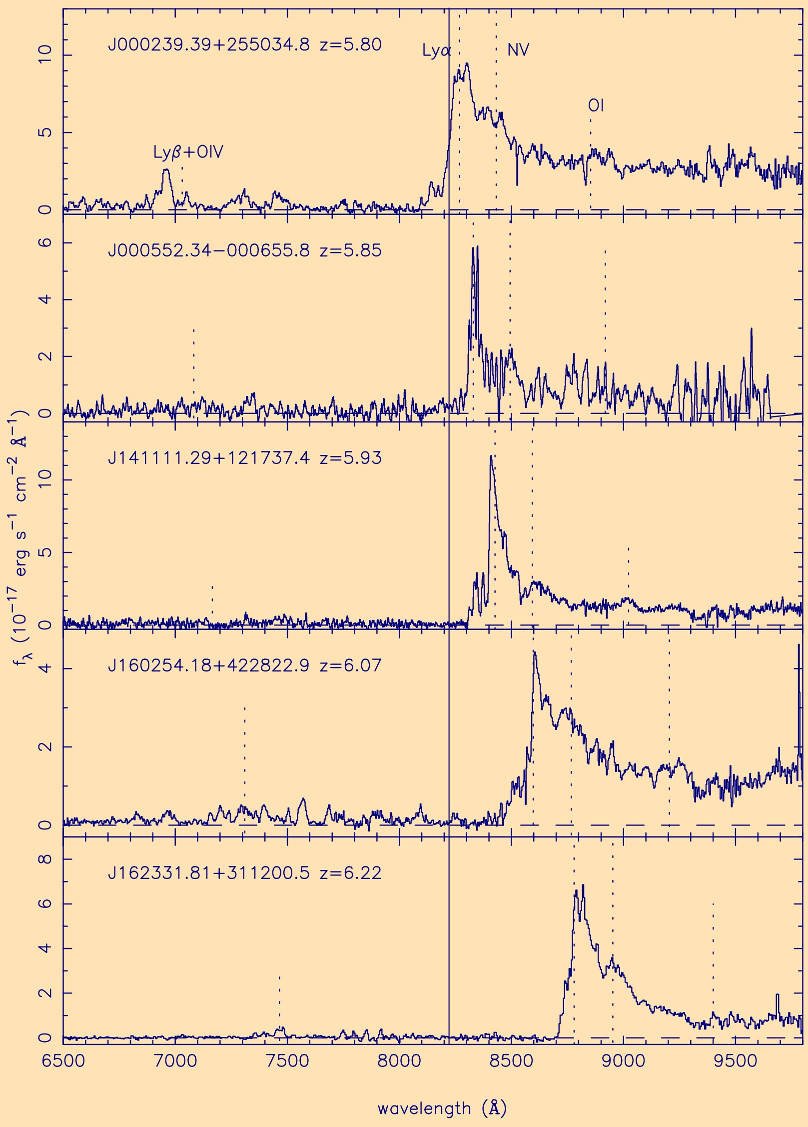
Redshift 6.00
3000
2000
1000
0
Flux
u
g
r
i
z
4000
6000
8000
10000
Observed Wavelength (Ångstrom)

Not all Point Sources are Point Sources

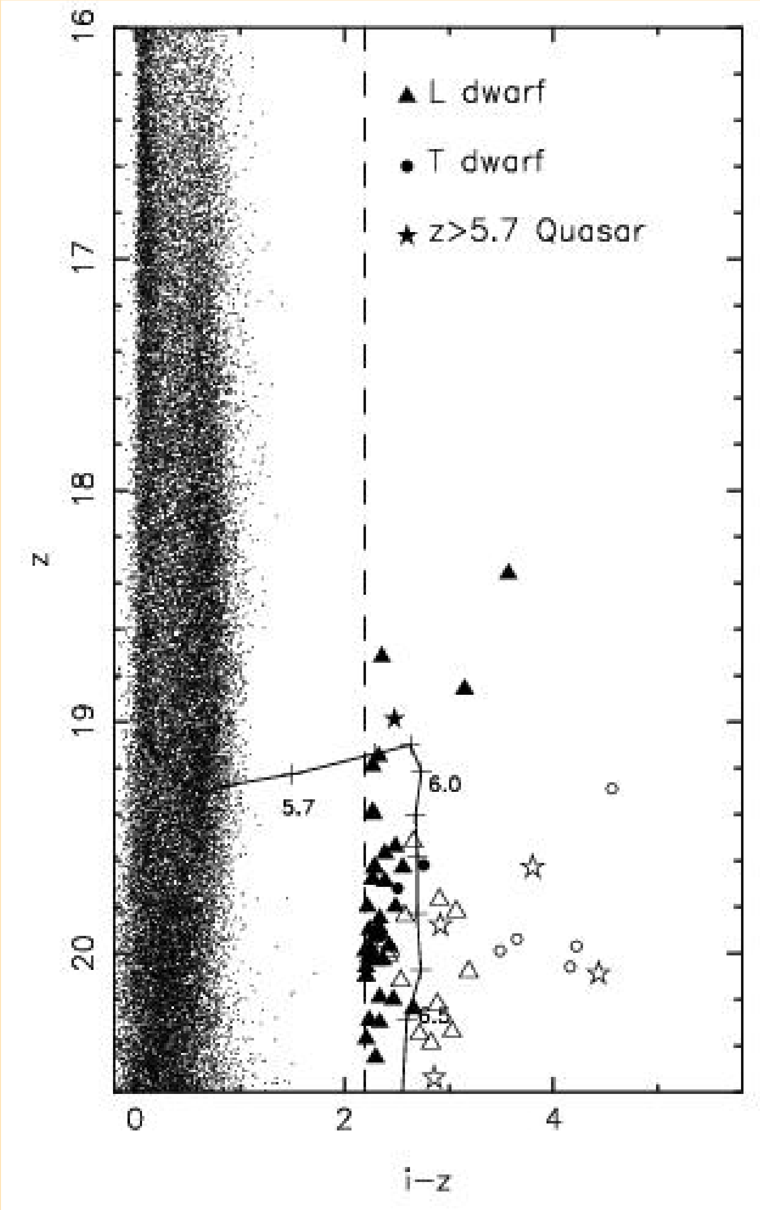


Stars and Galaxies and Cosmic Rays (cumulative)

High-z Quasars are not very Bright

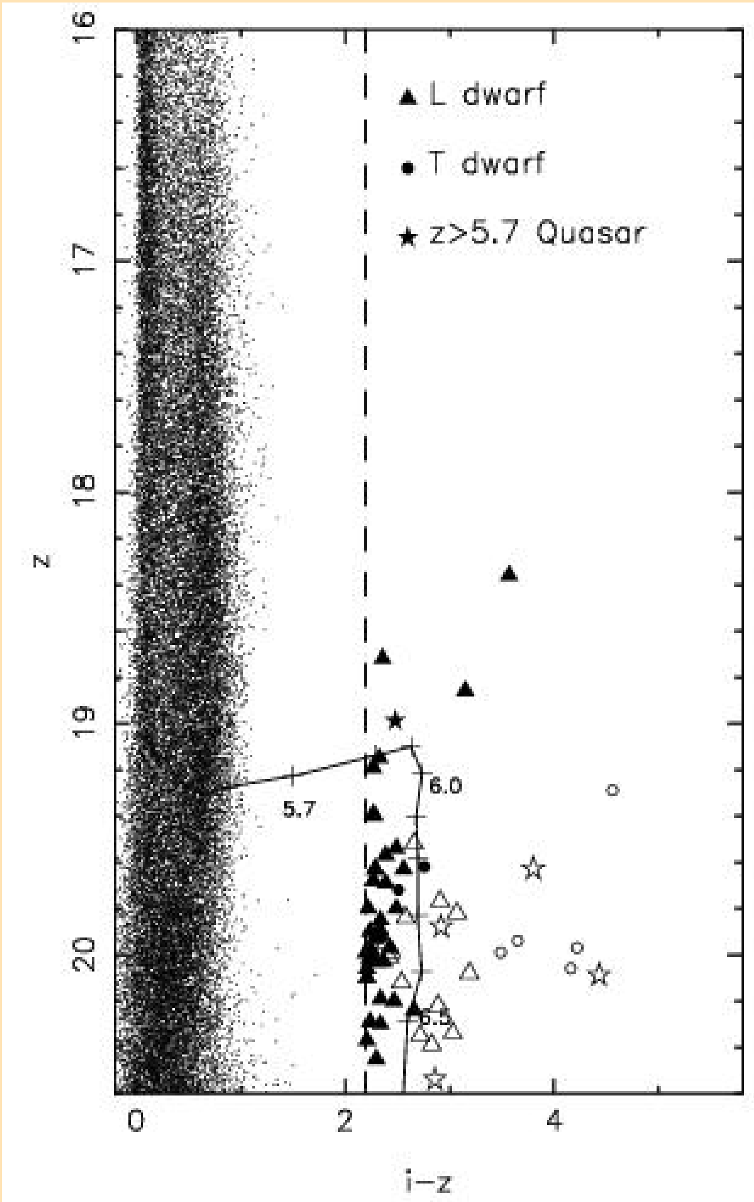


High-z Quasars are not very Bright



Filled symbols: $S/N|_i > 4$

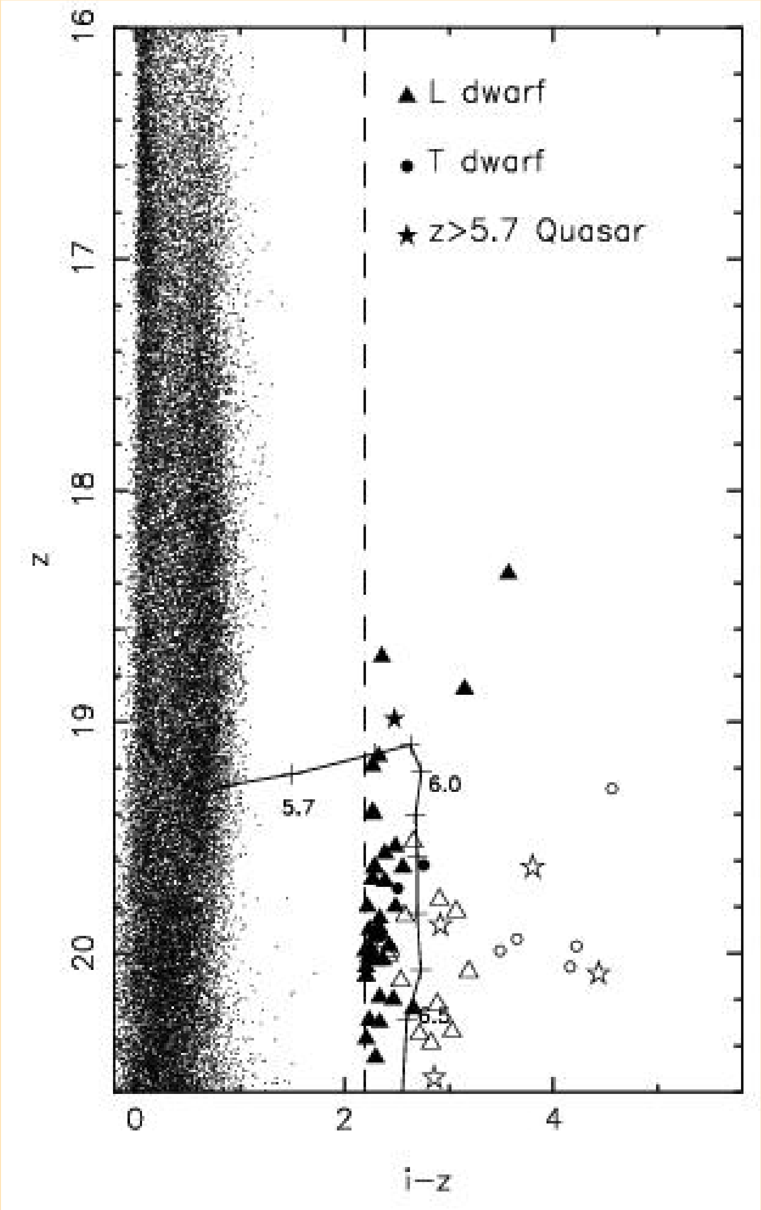
High-z Quasars are not very Bright



Filled symbols: $S/N|_i > 4$

The small dots are 5×10^4 stars (from $\sim 10\,\mathrm{deg}^2$ of sky)

High-z Quasars are not very Bright

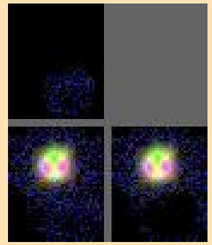


Filled symbols: $S/N|_i > 4$

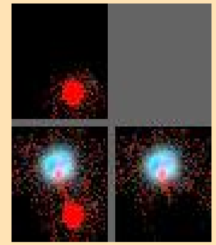
The small dots are 5×10^4 stars (from $\sim 10\,\mathrm{deg}^2$ of sky)

The quasars (and L/T stars) are from $\sim 1700\,\mathrm{deg}^2$

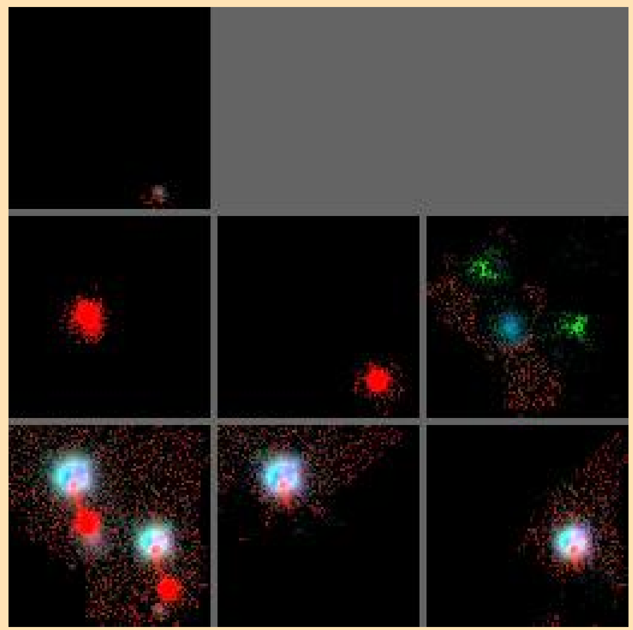
All that Glistens isn't Gold



ugr



riz



riz

Is Anything Left?

gri

riz

The Next Generation of Imaging Surveys

How could you possibly do better than SDSS?

- More sky coverage
- More epochs
  - Deeper photometry
  - More reliable photometry
  - Variability
  - Motions/Parallaxes
    1mas/year $\equiv$ 5 km/s at 1 kpc
- More bands (or redder bands)

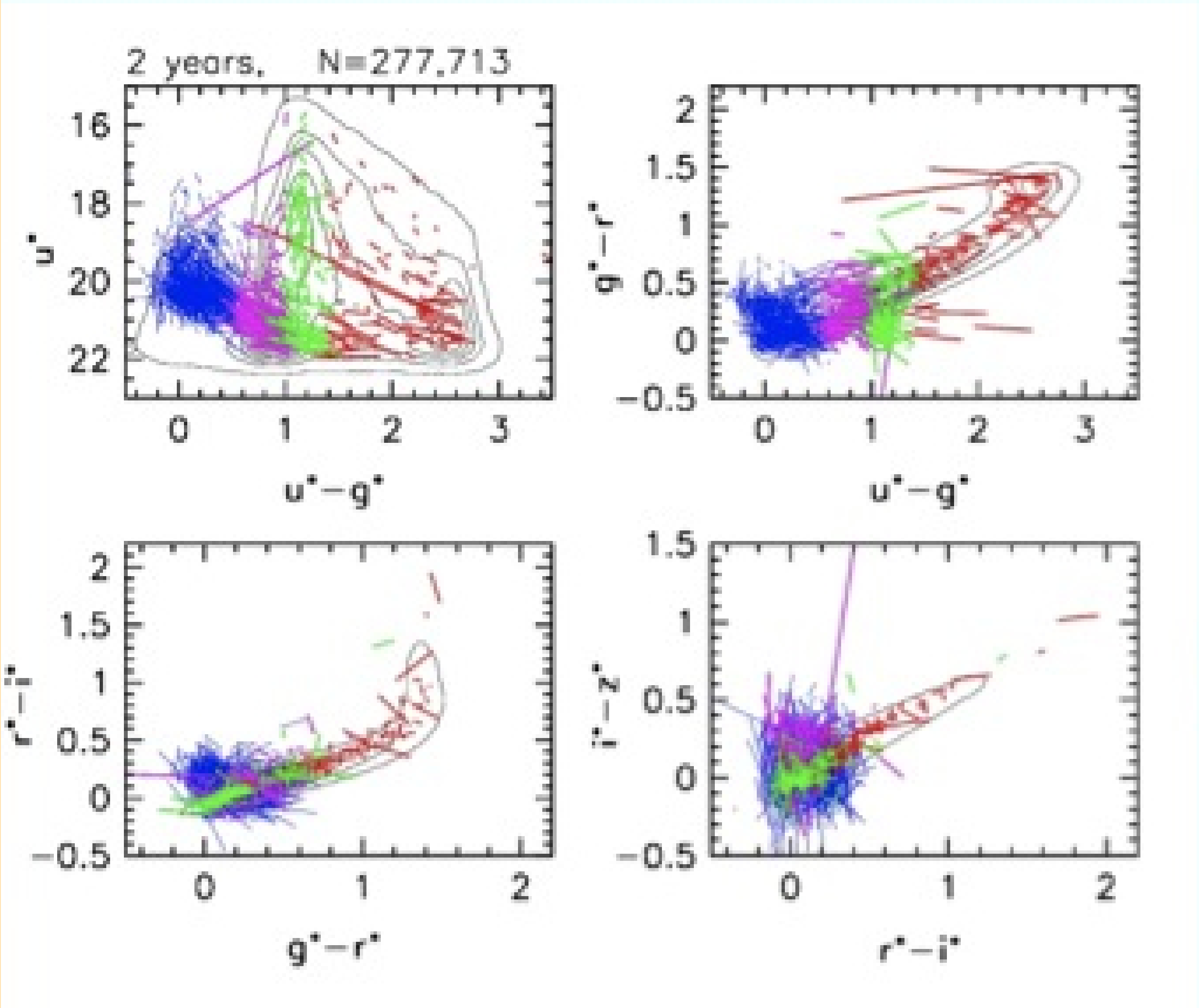
The Next Generation of Imaging Surveys

How could you possibly do better than SDSS?

These translate into:

- Dark Energy: Weak Lensing, BAO, SNe Ia
- Evolution of galaxies at high redshift
- Milky Way structure out to M31
- New classes of explosive events
- Populations of small bodies (near and far) in the Solar System.
- ...

E.g. Variability from SDSS



$\Delta t \sim 2$ years

Hyper-SuprimeCam (HSC)

Next generation surveys include:

	Aperture	FoV	Median FWHM	QE @ $1\mu m$	Nights
	m	deg^2	asec		year^{-1}
SDSS	2.5	1.5	1.2	0.05	365.24
PS1	1.8	7.0	1.0	0.2	365.24
DES	4.0	3.0	1.0	0.5	180
HSC	8.2	1.8	0.7	0.4	??
PS4	3.6	7.0	1.0	0.2	365.24
LSST	8.4	9.6	0.7	0.4	365.24

HSC: a Japan–Taiwan–Princeton (!) collaboration

PI: 宮崎聡 Satoshi Miyazaki

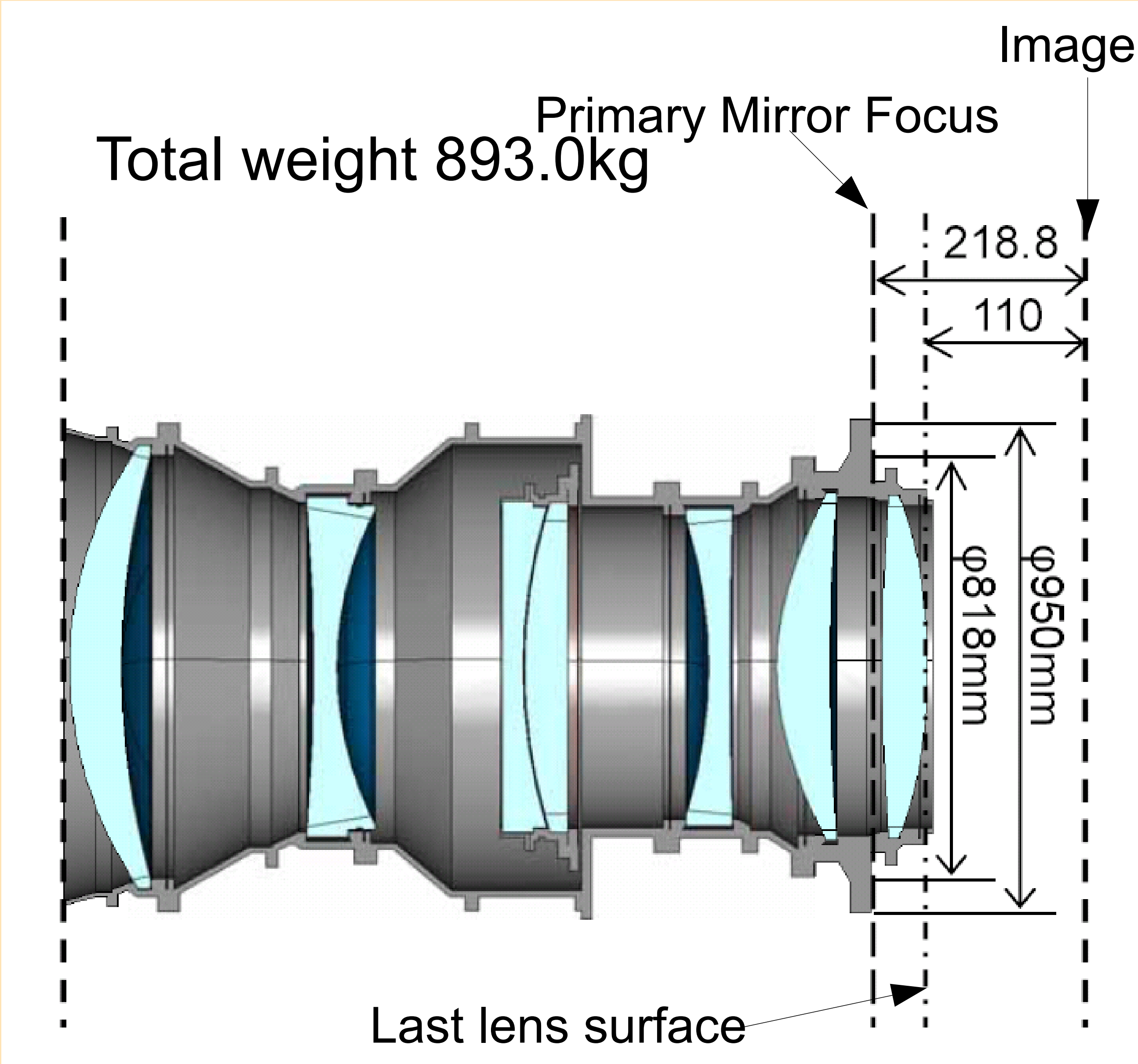
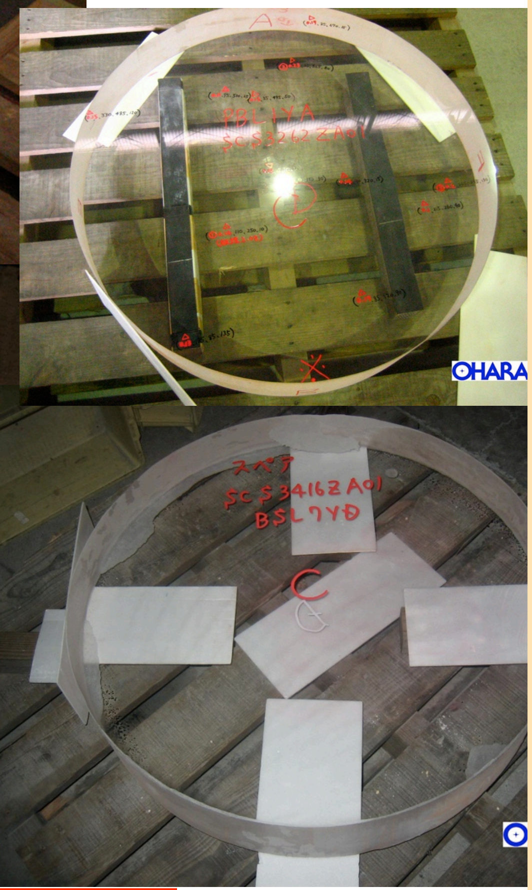
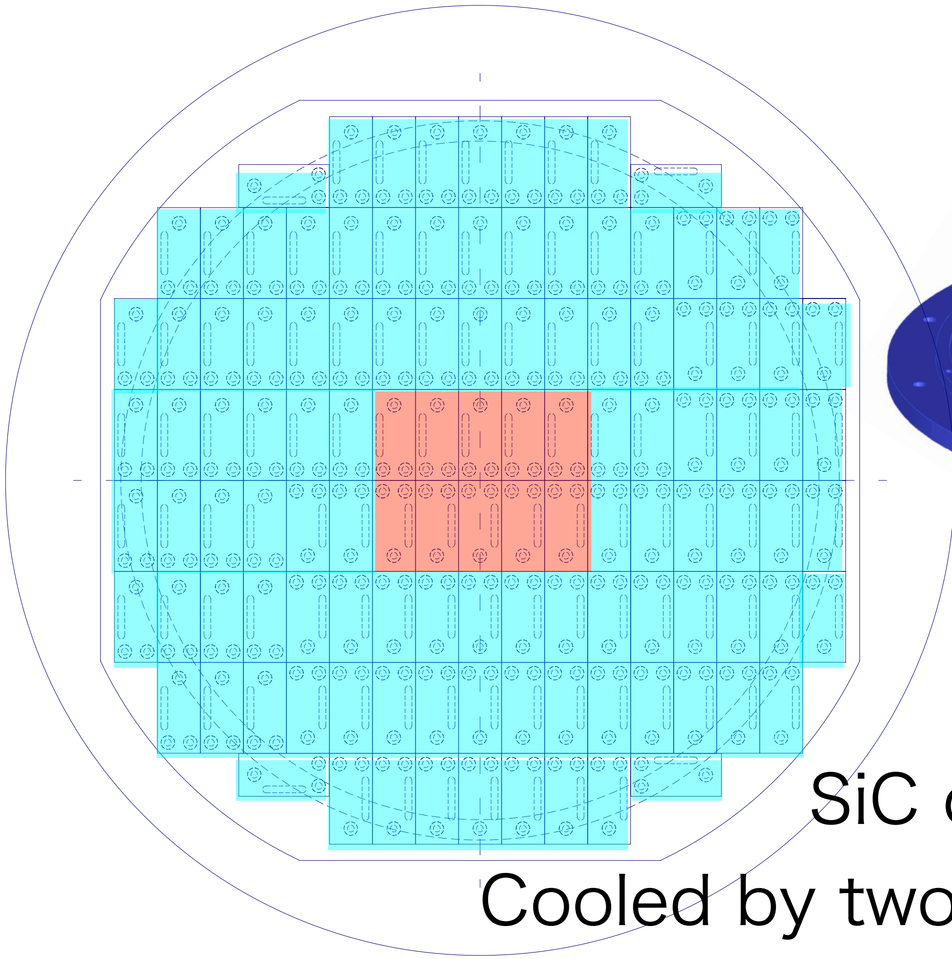


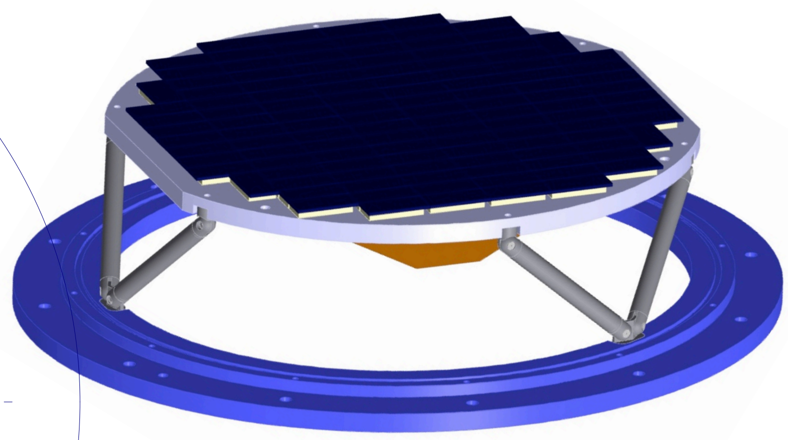
Image
Primary Mirror Focus
Total weight 893.0kg
218.8
110
φ818mm
φ950mm
Last lens surface

Silica: Corning
i-lines: OHARA
(homo < 5 ppm)



HSC Focal Plane



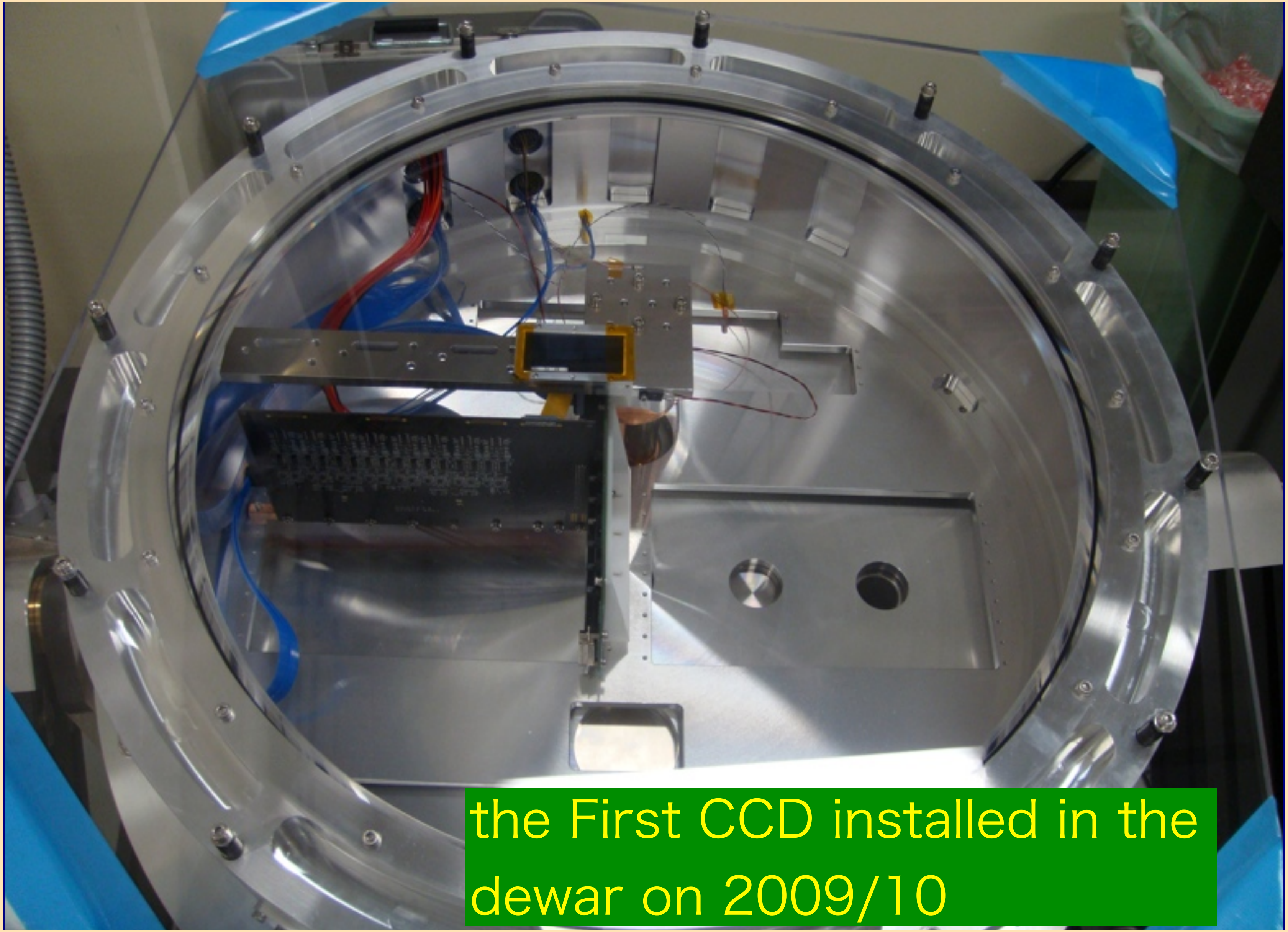


112 + 4 Guides

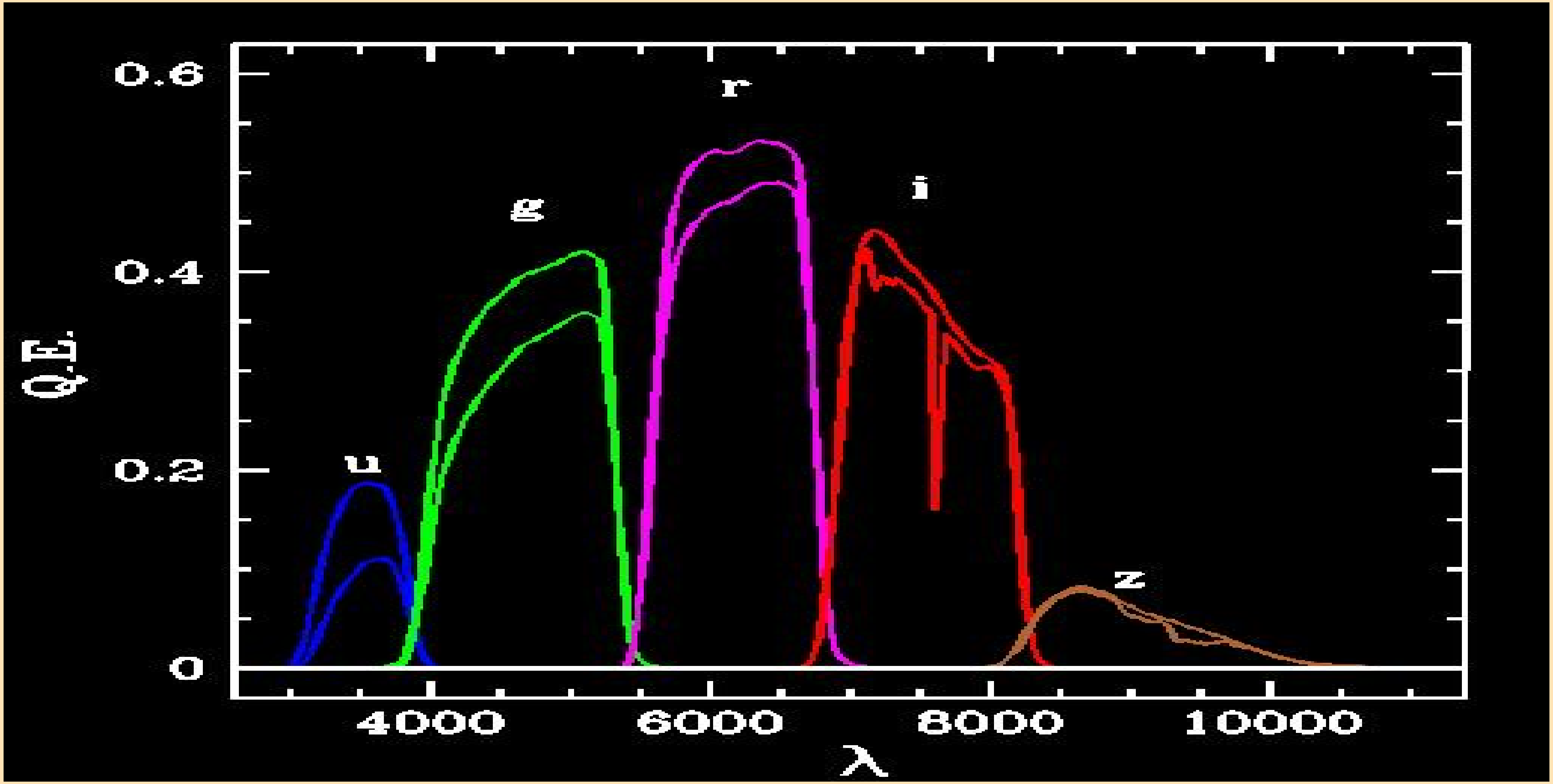
SiC cold plate

Cooled by two pulse tube coolers

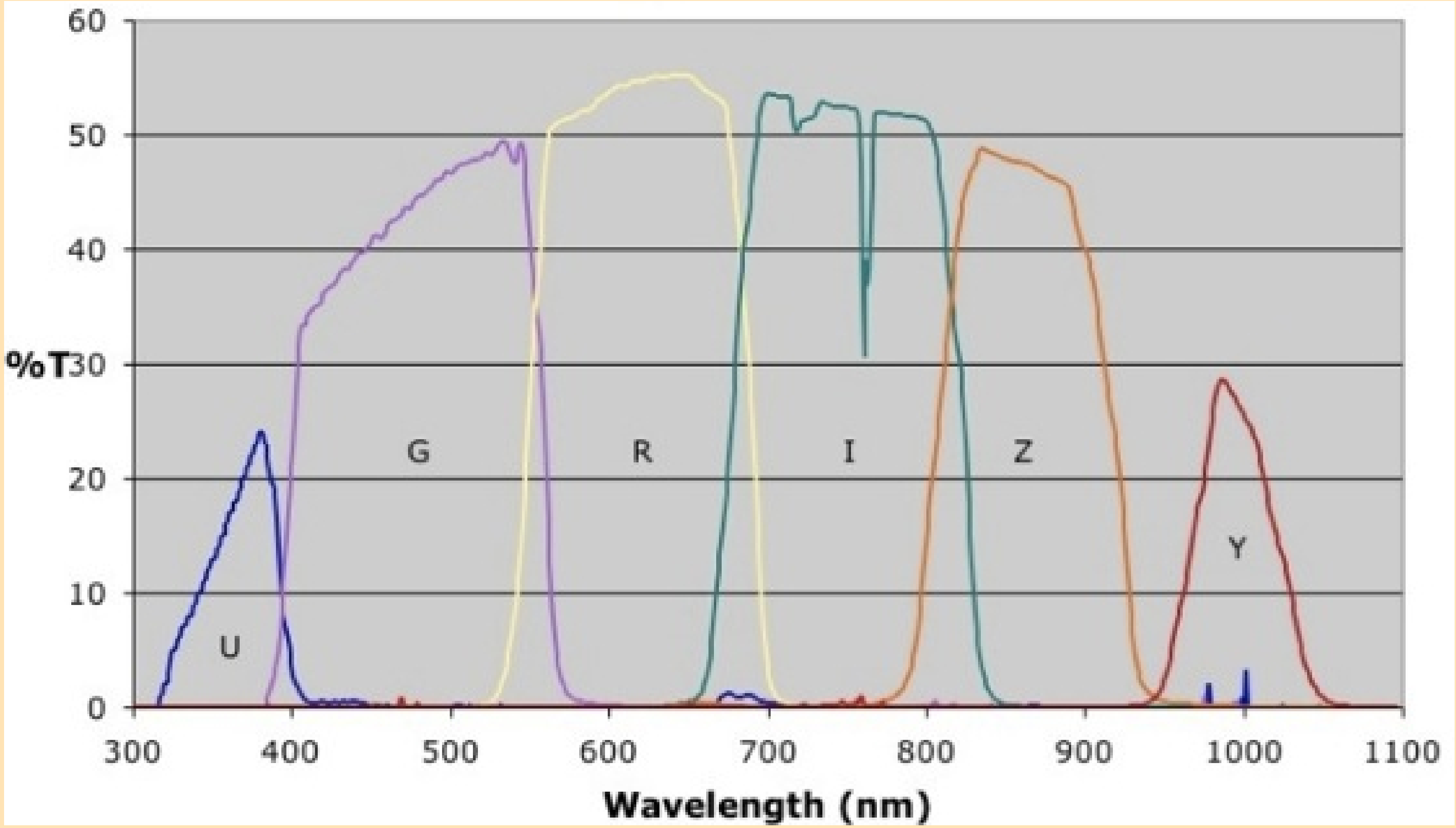
45 W@-100 C each



the First CCD installed in the dewar on 2009/10



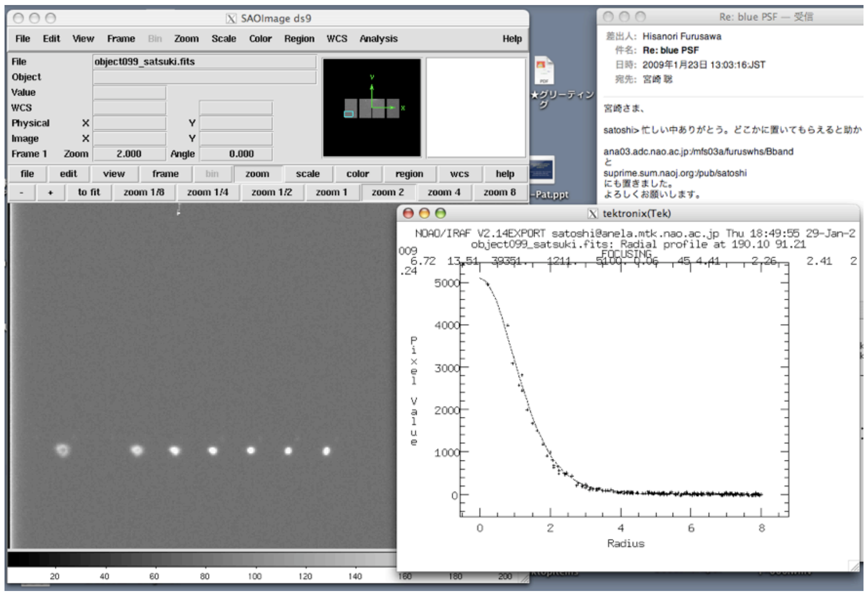
SDSS

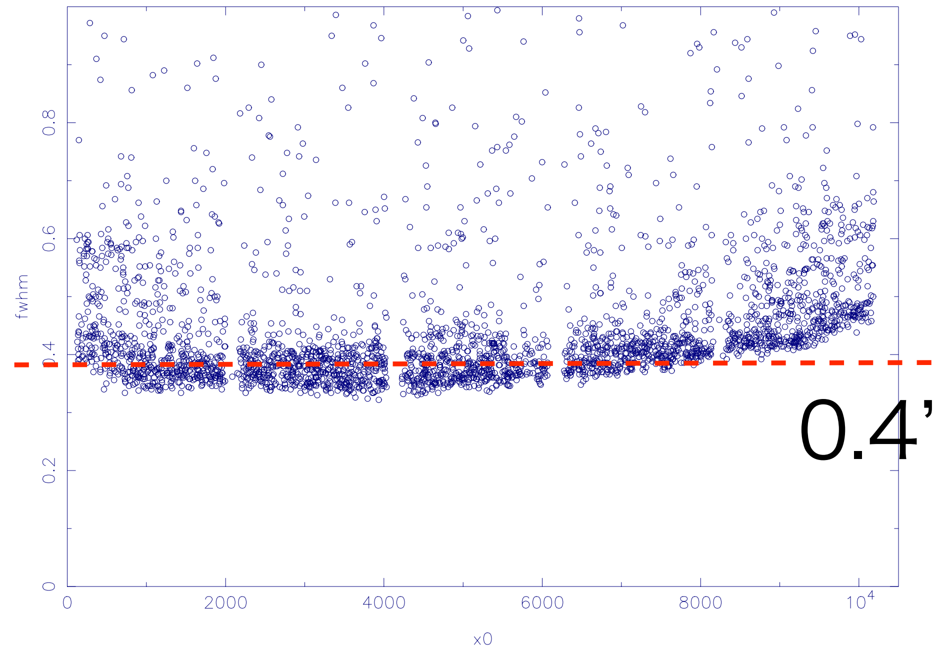


%T
Wavelength (nm)
U
G
R
I
Z
Y
0
10
20
30
40
50
60
300
400
500
600
700
800
900
1000
1100

B band

z' band





0.48'' FWHM

0.38'' FWHM

Seeing at Subaru

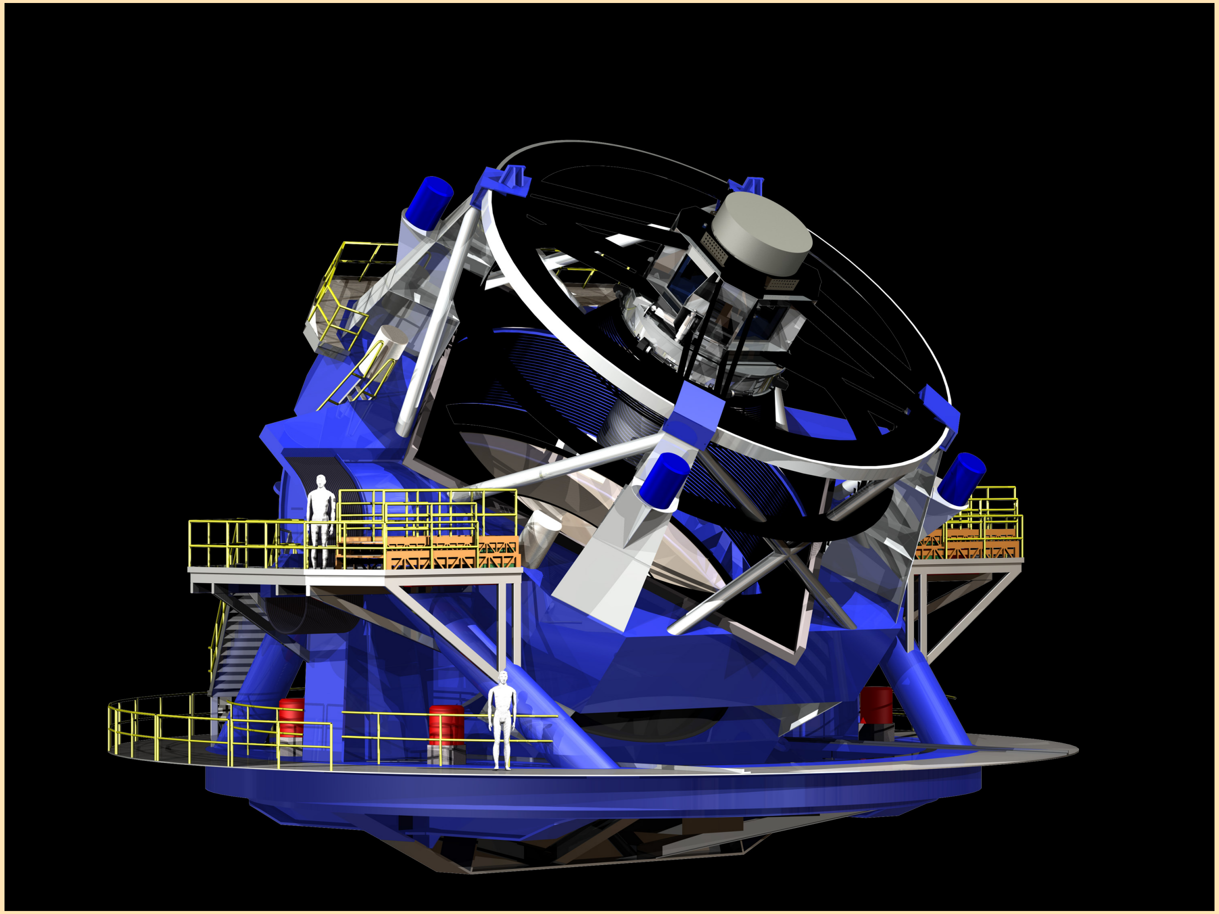
LSST

A number of (US) national decadal surveys have called for a dedicated wide-field telescope:

The LSST project is a collaboration of many university groups and national laboratories to build a large wide-field telescope.

Brookhaven National Laboratory (BNL), California Institute of Technology, Carnegie Mellon University, Columbia University, Google, Inc., Harvard-Smithsonian Center for Astrophysics, Johns Hopkins University, Las Cumbres Observatory, Inc., Lawrence Livermore National Laboratory (LLNL), National Optical Astronomy Observatory, Penn State University, Princeton University, Purdue University, Research Corporation, Rutgers University, Stanford Linear Accelerator Center, Stanford University -Kavli Institute for Particle Astrophysics and Cosmology, The University of Arizona, University of California at Davis, University of California at Irvine, University of Illinois at Champaign-Urbana, University of Pennsylvania, University of Pittsburgh, University of Washington ...

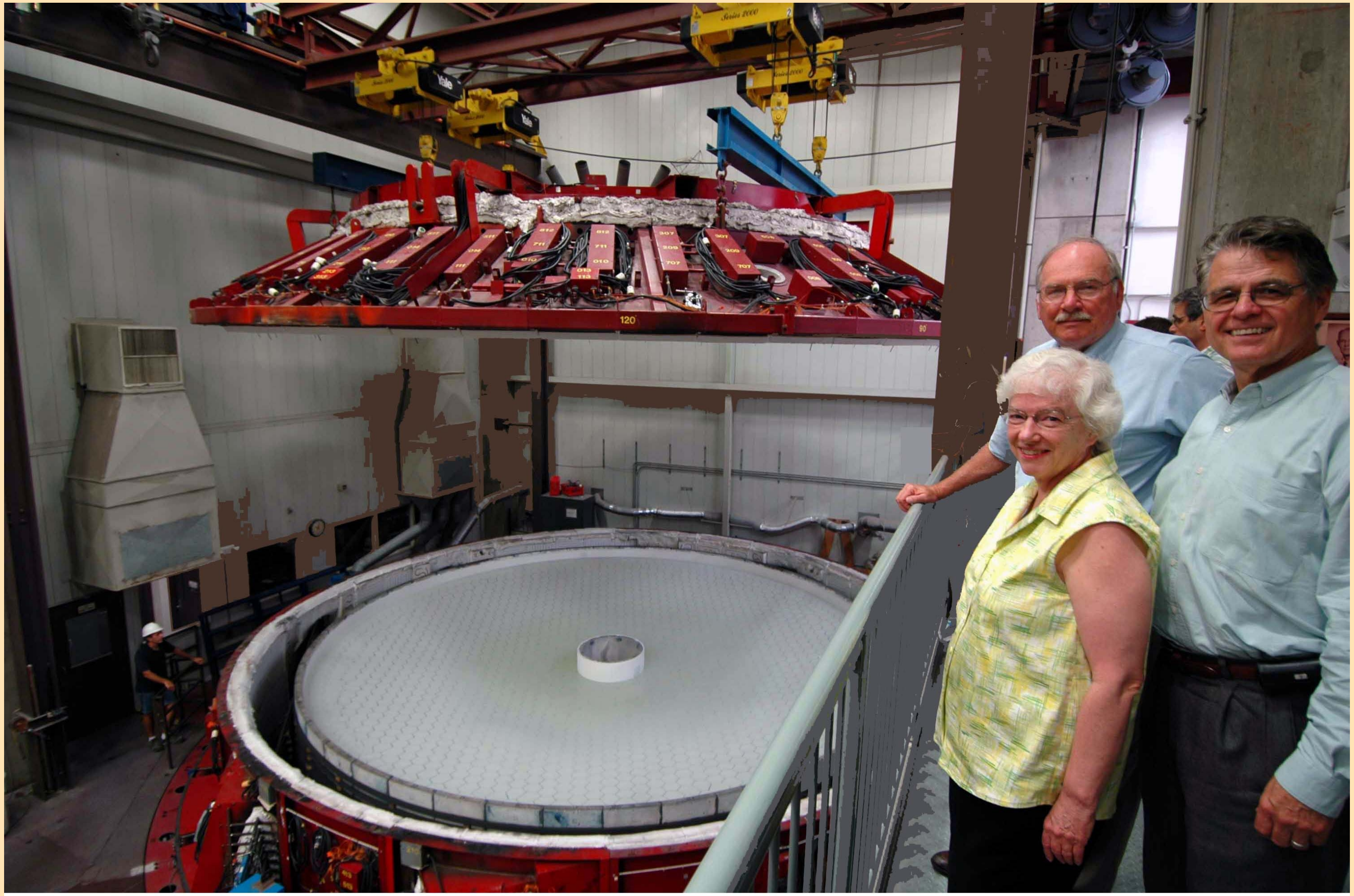
The primary will be 8.4m in a compact telescope



on Cerro Pachón in Chile


LSST-webcam 2 2011-03-08 12:56:02

The optical design is unusual:



To make use of such a large field (9.6 deg^2) requires a large camera...

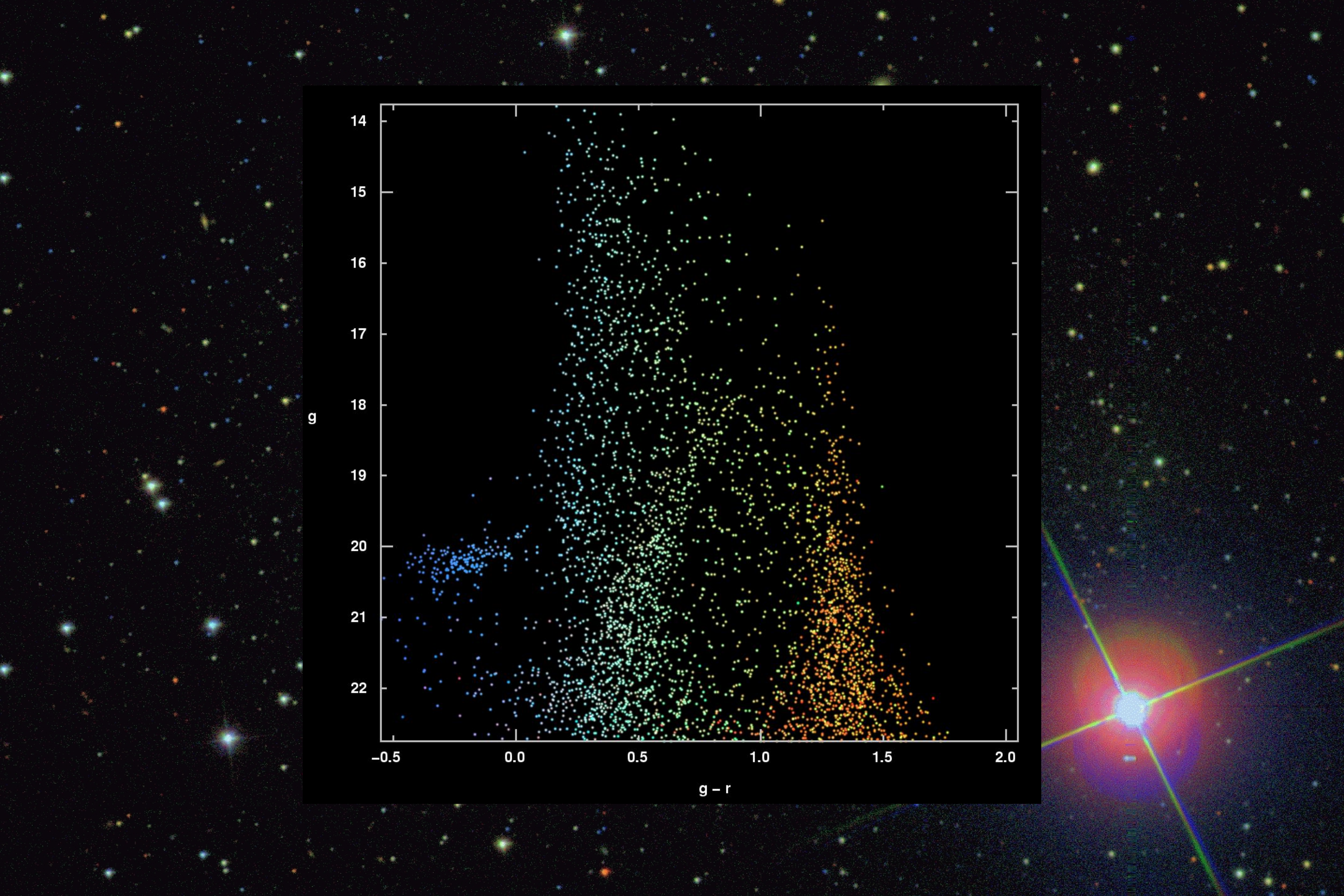
- 390M$ (2006); 45M$/year operating costs
- NSF, DoE, Private, ...

- Main survey will cover 20,000 deg^2
- Over 300 15s exposures in g, r, i, z, y
- 2σ depths after a pair of 15s exposures are 23.9 (u), 25.0 (g), 24.7 (r), 24.0 (i), 23.3 (z), 22.1 (y)
- At end of the survey, 26.2, 27.4, 27.6, 26.9, 26.1, and 24.8

A Series of Science Collaborations:

- *Weak lensing* Bhuvnesh Jain and Dave Wittman
- *Strong lensing* Phil Marshall
- *Supernovae* Michael Wood-Vasey
- *Large-scale structure/BAO* Andrew Hamilton
- *AGN* Niel Brandt
- *Galaxies* Harry Ferguson
- *Galactic structure* James Bullock and Beth Willman
- *Stellar populations* Abi Saha
- *Variability and transients* Shri Kulkarni and Lynne Jones
- *Solar system* Steve Chesley


The End



14
15
16
17
18
19
20
21
22
g
−0.5
0.0
0.5
1.0
1.5
2.0
g − r

What's involved in handling the next generation of data?

- Hardware
  - Disk
  - Processors and Memory. GPUs? Cell Processors?
- Software
  - Algorithms
  - Software Engineering and Techniques
  - Sociology

Software Engineering and Techniques

- Languages (C++ and python?)
- Data types (objects)
- Build systems (or, I hate libtool; LSST uses scons)
- Versioning
- Process management (Naïve ssh? GRID? custom MPI?)
- Fault tolerance
- Provenance
- Testing (regression; science; coverage)
- Data Challenges

Sociology

- People
- Careers
- Collaborating at the algorithms level
- Collaborating at the code level
- Deciding what's the responsibility of different Scientists

Processing Polychromatic Sets of Images

A currently popular approach is to resample the various exposures to a common grid and sum the resulting images with some weighting/filtering. However:

- Correlated noise
- Sampling
- Discontinuous PSFs
- No opportunity for non-linear analysis in the processing (e.g. 3σ clips).
- Average over moving/variable objects

On the other-hand, it has the great advantage of being computationally relatively simple and cheap.

An easy alternative is to process each exposure separately, and add the resulting measurements.

- Only objects detected in at least one frame are measured
- There is no guarantee that the same objects will be detected in each exposure
- It seems unlikely that the errors in all measurements (e.g. galaxy effective radii) will scale as $\sqrt{N}$.

There are ways around some of these problems; for example, we could *detect* on a coadded frame and then use this master catalogue to measure each of the input images.

Does this sound familiar to users of DAOPHOT?

A new generation of analysis codes should:

- Never resample the data
- Analyse stacks of data (taken in multiple bands) as a series of samples of the sky, rather than attempt to generate a single image.
- Make full use of the per-exposure PSF information
- Preserve variability information (astrometric and photometric)
- Use some standard software framework

(Semi-?) Open algorithmic questions

- Estimating the PSF and its spatial structure
- Detecting objects (resolved/trailed; χ^2 image or given SED or ...)
- Deblending of stars and galaxies
- Shape measurements
- (Galaxy) photometry


The End

How should I coadd a set of images?

Given a set of images of the same part of the sky, how should I add them to obtain a deeper image?

- How far does $\sqrt{N}$ take you?
- What's a good algorithm?
- Is there an optimal algorithm?

There are (at least) three ways to think about adding images:

- Add the images together
- Estimate a picture of the Universe
- Estimate the properties of the Universe

Estimating a Picture of the Universe

If we take the middle tack, we can write down the ML estimate of the Universe U given an image, I, and a (known) PSF, ϕ:

$$I(k) = U(k) \times \phi(k) + \epsilon(k)$$

Let us assume that all objects are fainter than the sky, so ϵ is an $N(0, \sigma^2)$ variate.

$$\ln \mathcal{L} \propto -\sum_i \ln \sigma_i - \frac{1}{2} \sum_i \left(U \phi_i - I_i\right)^2 / \sigma_i^2$$

so, differentiating with respect to the Universe,

$$U(k) = \frac{\sum_i I_i \phi_i / \sigma_i^2}{\sum_i \phi_i^2 / \sigma_i^2} \equiv \frac{D(k)}{P(k)}$$

An Optimal Algorithm

$$U(k) = \frac{D(k)}{P(k)}$$

$$D(k) \equiv \sum_i I_i \phi_i / \sigma_i^2; \qquad P(k) \equiv \sum_i \phi_i^2 / \sigma_i^2$$

I.e.

$$U(x) = D(x) \otimes^{-1} P(x)$$

where

$$D(x) = \sum_i I_i \otimes \phi_i / \sigma_i^2; \qquad P(x) = \sum_i \phi_i \otimes \phi_i / \sigma_i^2$$

Is this Wise?

Probably not.

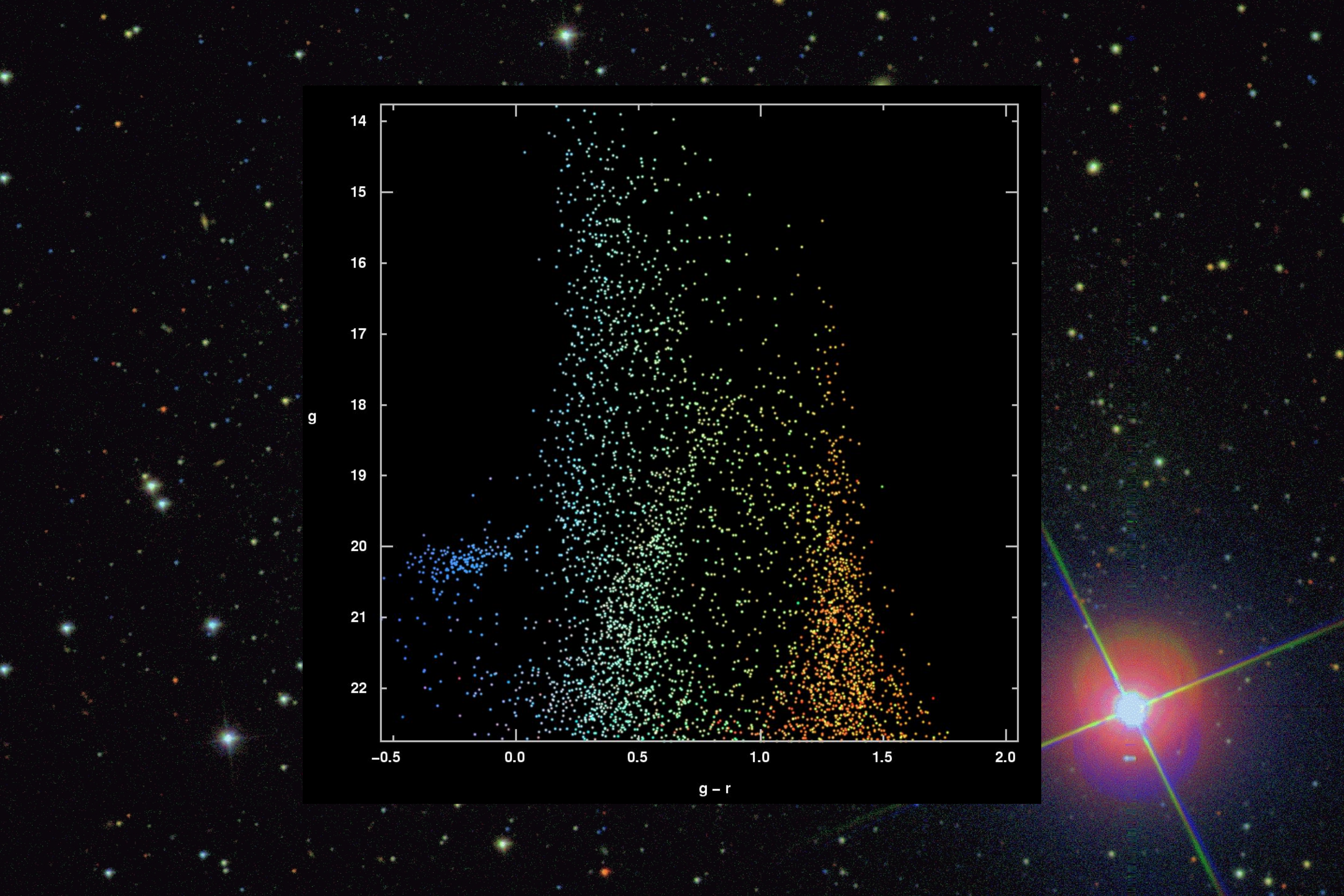
Estimate the properties of the Universe

This is straightforward for e.g. PSF magnitudes.

Harder problems include:

- Sky estimation
- Object detection
- Deblending
- Shape measurements


The End



g
14
15
16
17
18
19
20
21
22
–0.5
0.0
0.5
1.0
1.5
2.0
g – r