Null Tests

Looking for signal in all the wrong places...

Mike Jarvis

Feb. 13, 2017

LSST DESC DE School

Systematic Errors

- **Statistical Errors** are errors that have zero expectation value, and typically become closer to zero as you take more data.
- **Systematic Errors** are errors that have non-zero expectation value.
- LSST cosmology will most likely be **systematics limited**. The statistical errors will be so small that even fairly subtle systematic effects will probably dominate the error budget
- At the very least, **systematic errors will dominate our analysis effort**.

Systematic Errors

- Statistical errors are often called **Noise**
- Systematic errors are often called **Bias**

Measurement = Truth + Noise + Bias

<Measurement> = Truth + <Bias>

Null Tests

- Null tests are measurements whose truth value is expected to be zero, but which still respond to the systematic errors.

- The value of the bias in the null test may or may not be the same value as the bias that affects the real measurement.

Null = 0 + Noise + Bias′

<Null> = <Bias′>

Null Tests

- The goal is to design null tests such that **if <Bias’> ≈ 0, then <Bias> ≈ 0**.

- In other words, whenever **<Bias> ≠ 0**, we would notice this by finding that **<Null> ≠ 0**.

Designing Null Tests

Principle 1:

If a systematic effect has a dependent relationship with some variable and the science does not, then binning with respect to that variable is often a good null test.

Designing Null Tests

Principle 2:

Correlation functions between quantities that should not have any dependent relationship in the absence of systematics often make good null tests.

Designing Null Tests

Principle 3:

If the science should be invariant with respect to some variable, splitting the data on that variable and looking at the difference can be a good null test.

Designing Null Tests

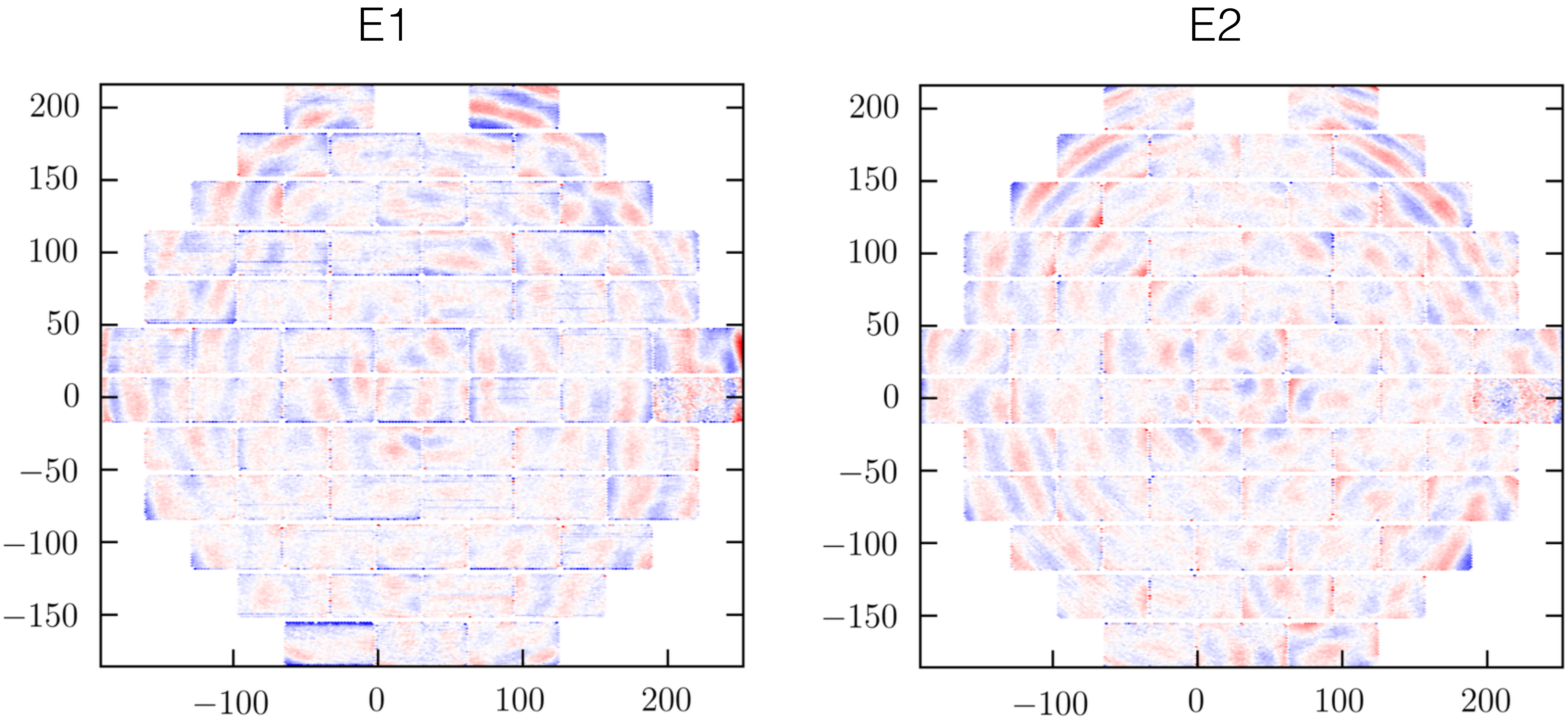
Principle 4:

Be creative. Try to think of non-obvious consequences of the systematic error. This design process is really the essence of science: trying to find ways to disprove a null hypothesis.

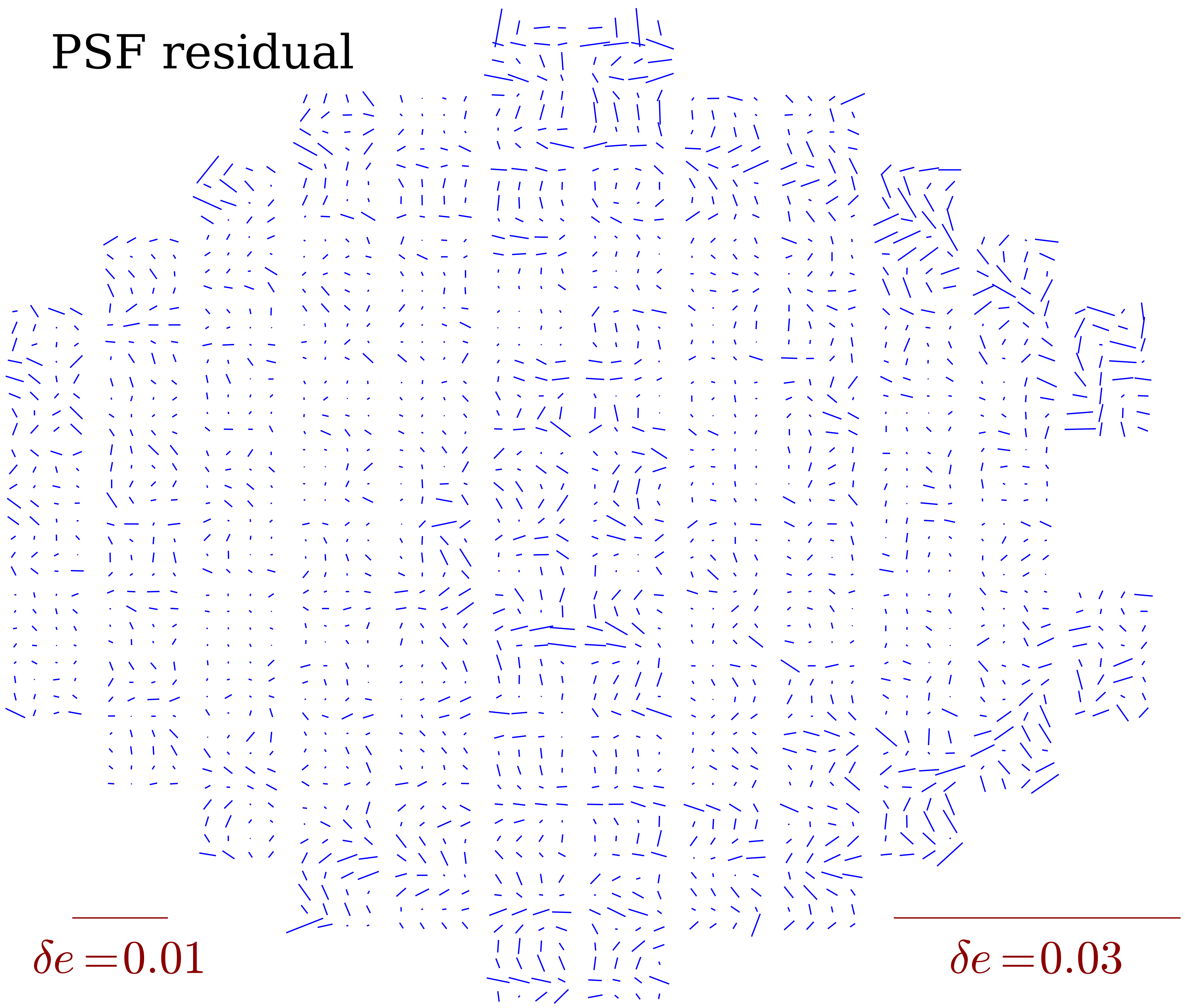
Example 1

The shape of the PSF is partly due to the telescope optics, including aberrations such as coma and astigmatism, whose effects vary across the focal plane.

Design a null test to see if the PSF estimation code is doing a good job of accounting for this variation in the shape of the PSF.

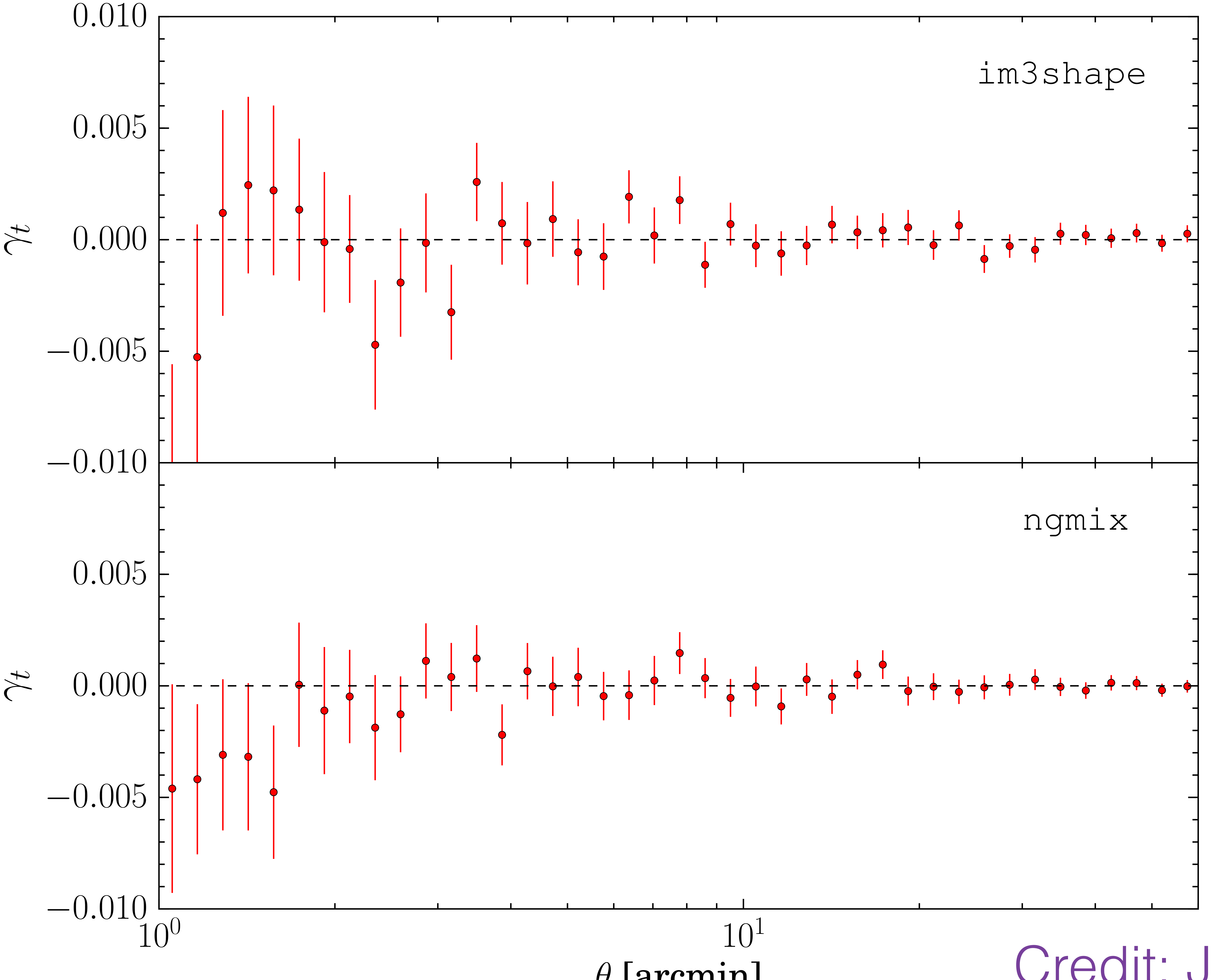


Credit: Sarah Bridle



Credit: Jarvis, et al, 2016

Tangential shear around field centers

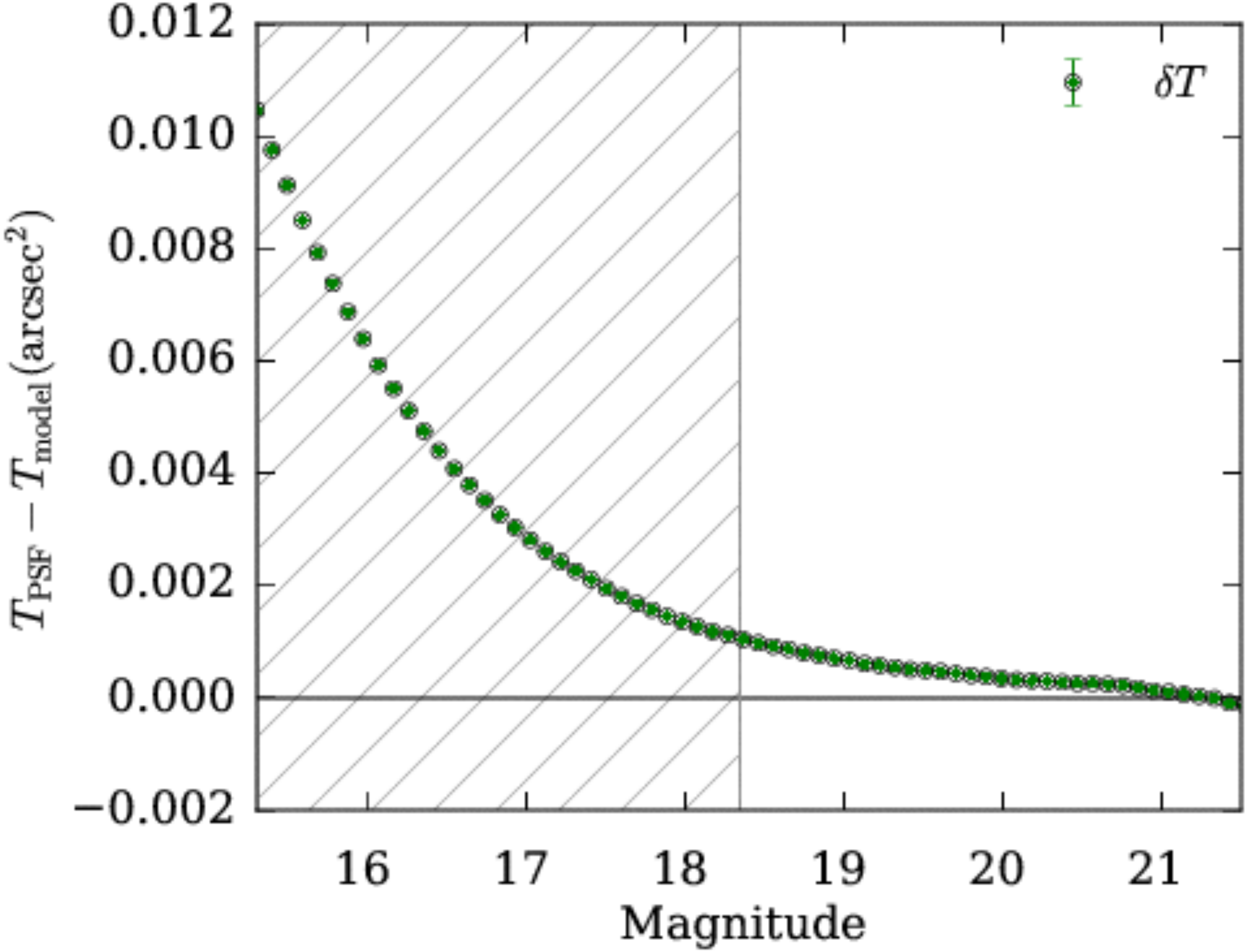


Credit: Jarvis, et al, 2016

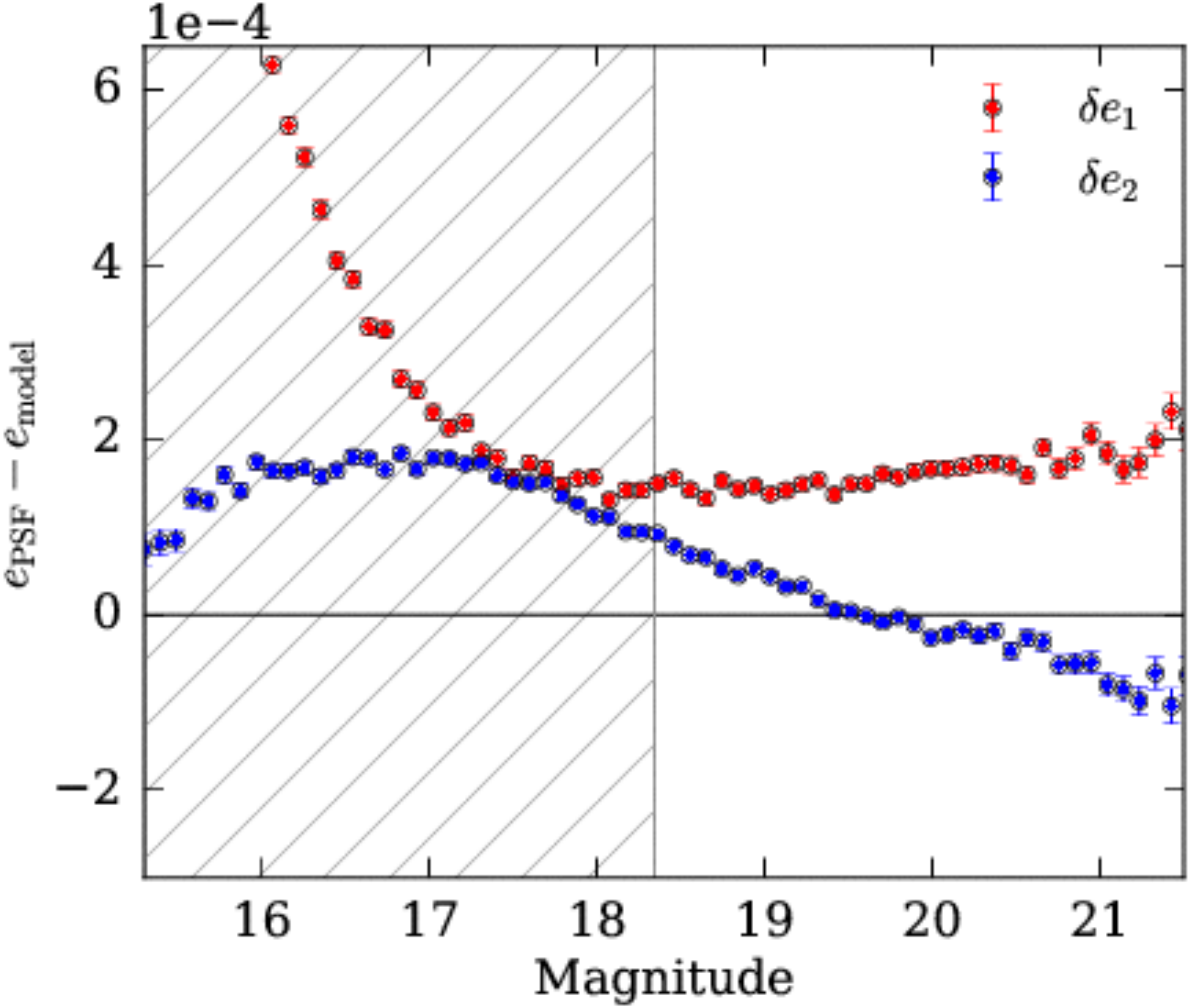
Example 2

When the number of electrons accumulated in a single pixel starts to get large, the electrons repel subsequent electrons incident in that pixel pushing them to neighboring pixels. This is called the brighter-fatter effect.

Design a null test to check that we are properly correcting for this effect.



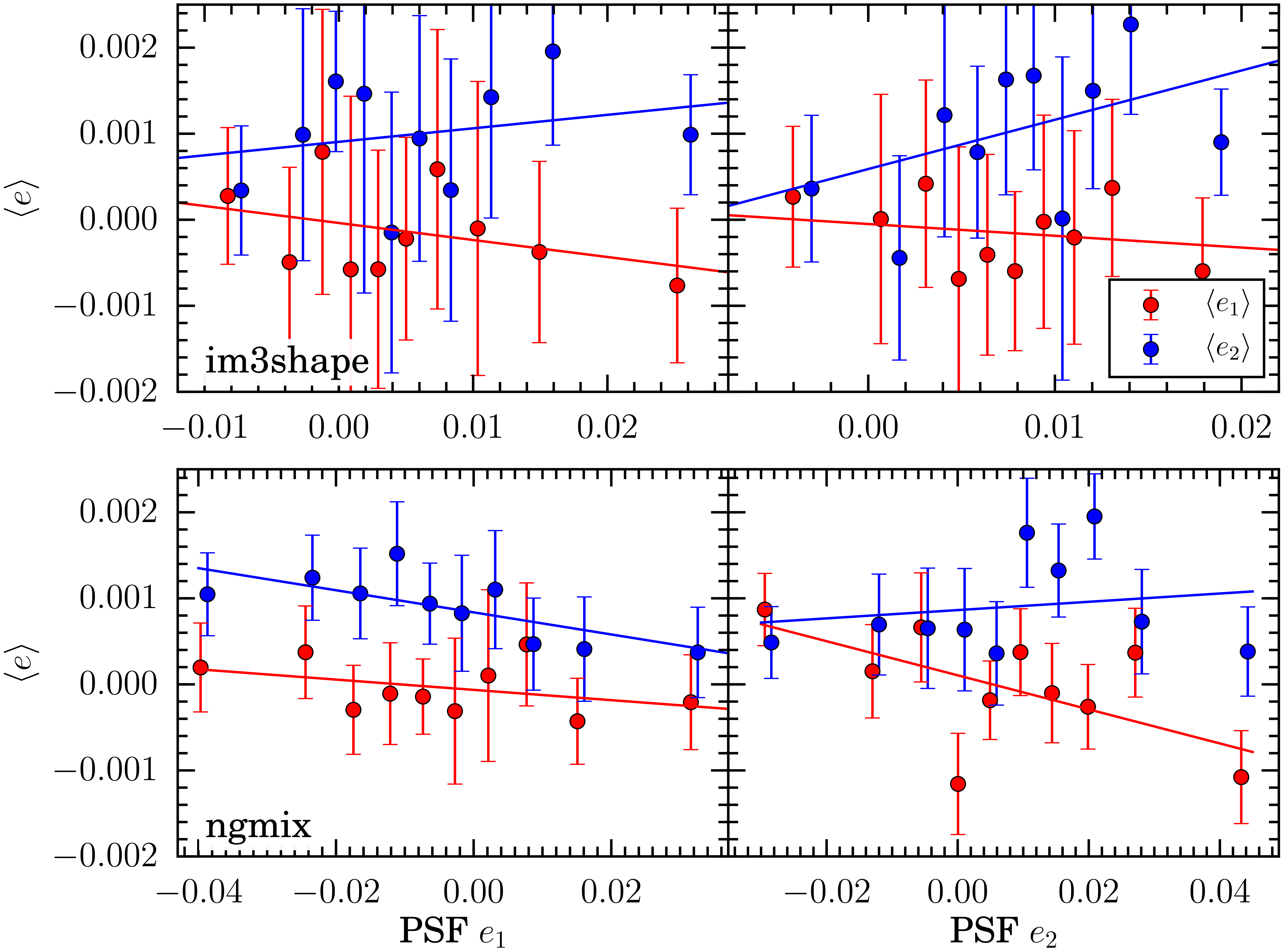
0.012
0.010
0.008
0.006
0.004
0.002
0.000
−0.002
$T_{\rm PSF} - T_{\rm model}$ (arcsec^2)
16
17
18
19
20
21
Magnitude
δT


1e−4
6
4
2
0
−2
$e_{\mathrm{PSF}} - e_{\mathrm{model}}$
16
17
18
19
20
21
Magnitude
δe_1
δe_2

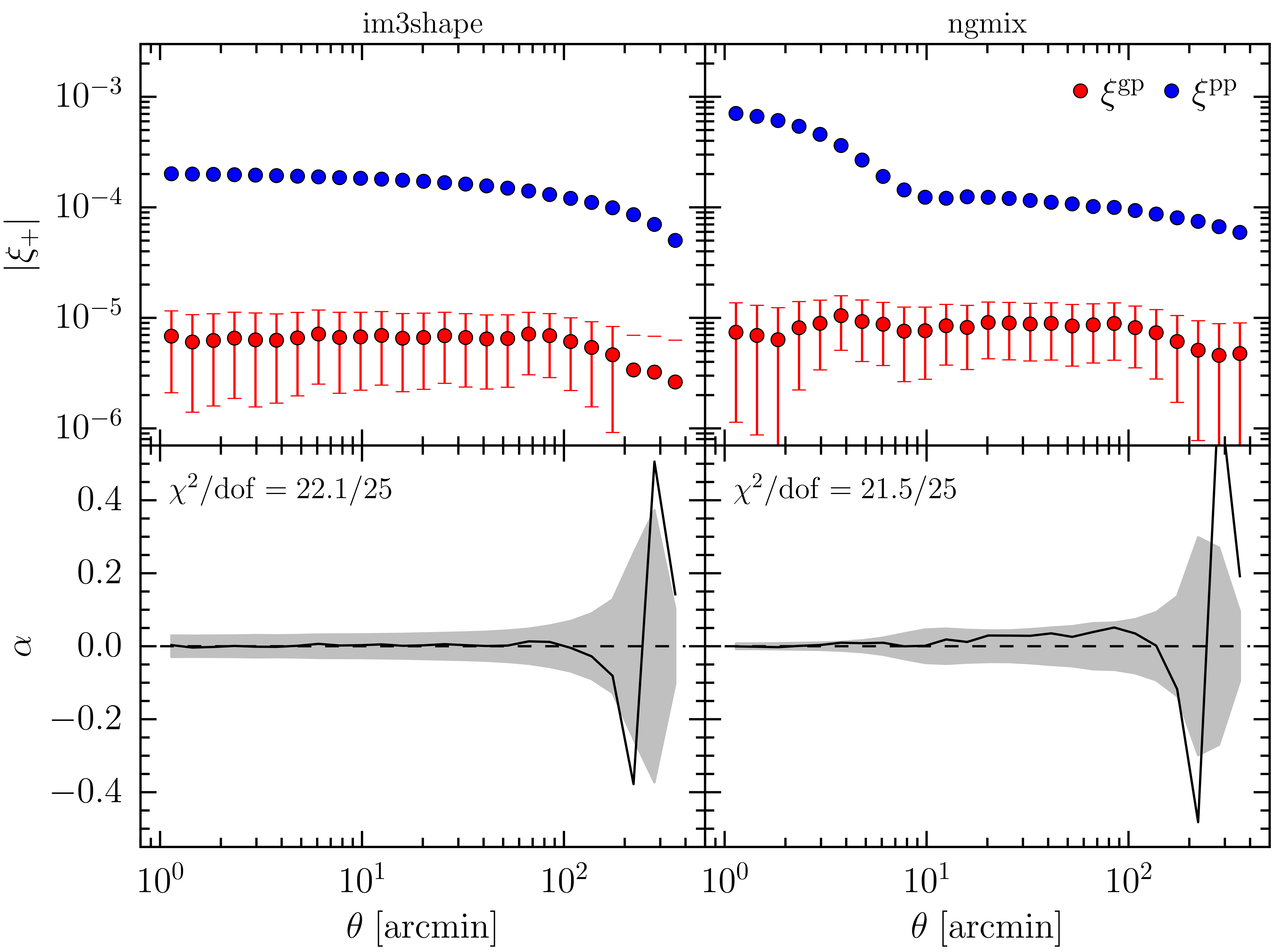
Example 3

One way that shear estimation algorithms can err is to leak some of the PSF shape into the galaxy shapes.

Design a null test that checks if there is any significant PSF leakage in the galaxy shape estimates.



Credit: Jarvis, et al, 2016

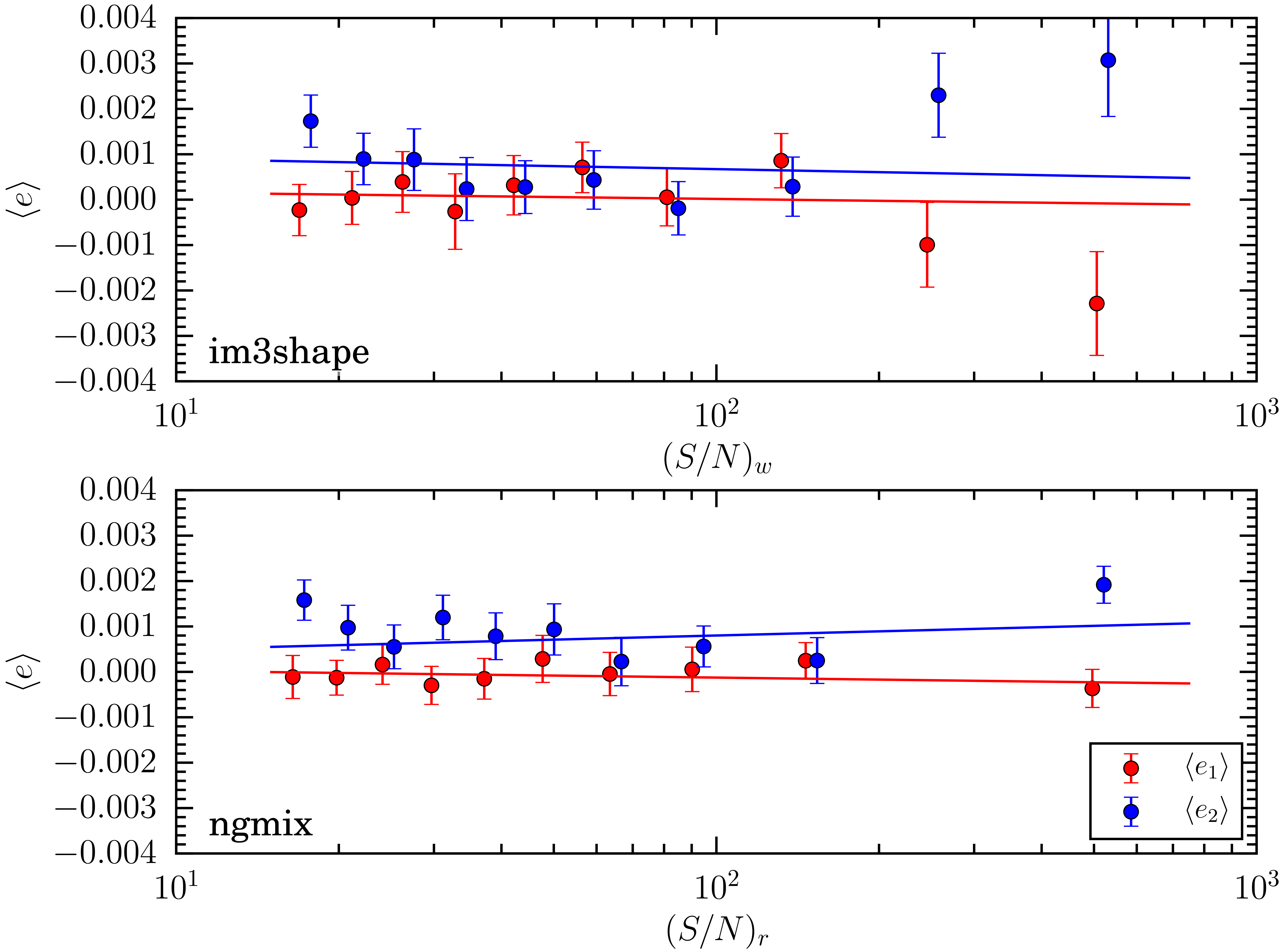


Credit: Jarvis, et al, 2016

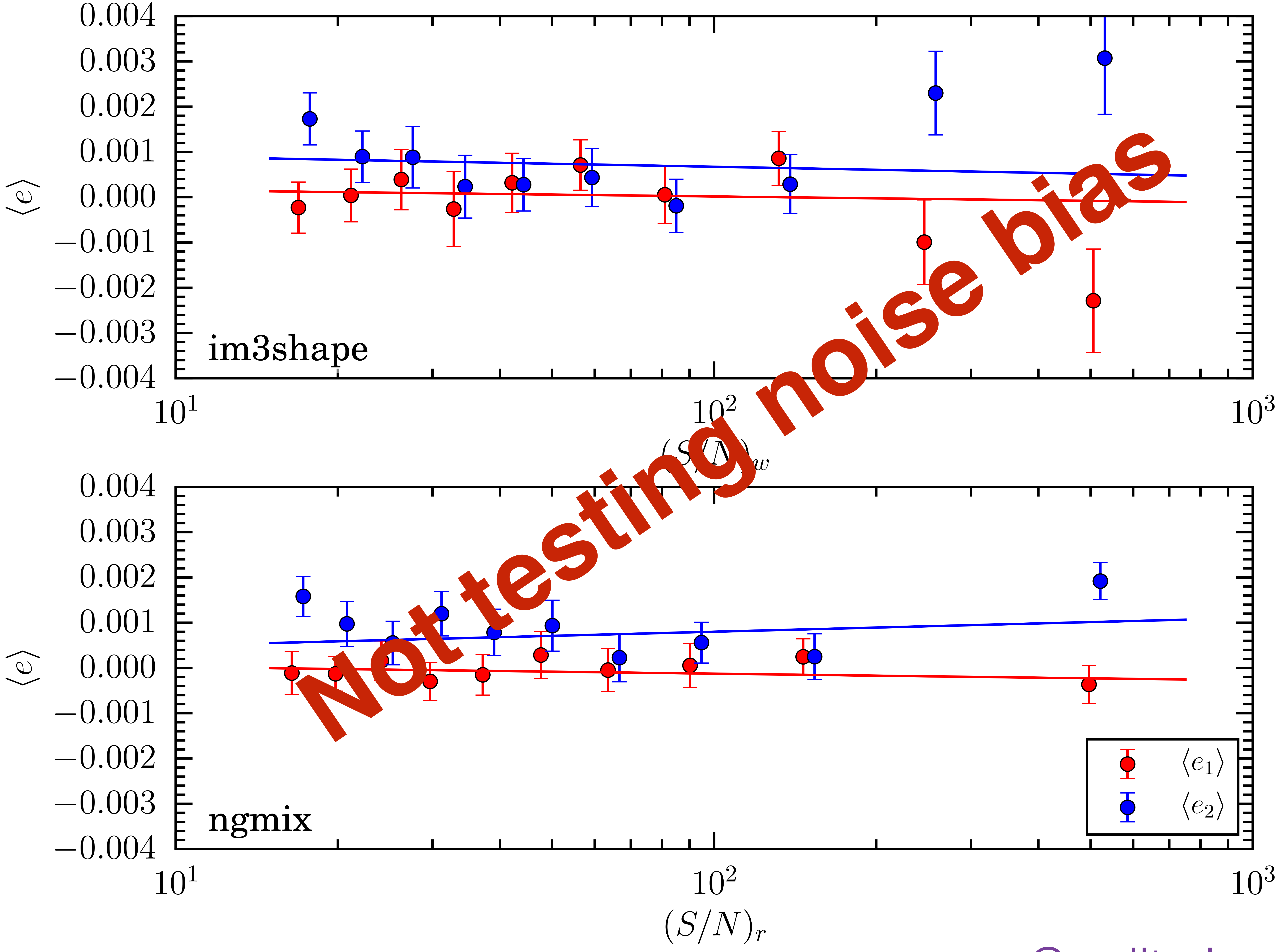
Example 4

Shear estimation at low S/N is hard. In particular, noise bias results from the non-linear parts of the shear estimation process, which turn the statistical noise in the pixel values into a bias in the shear.

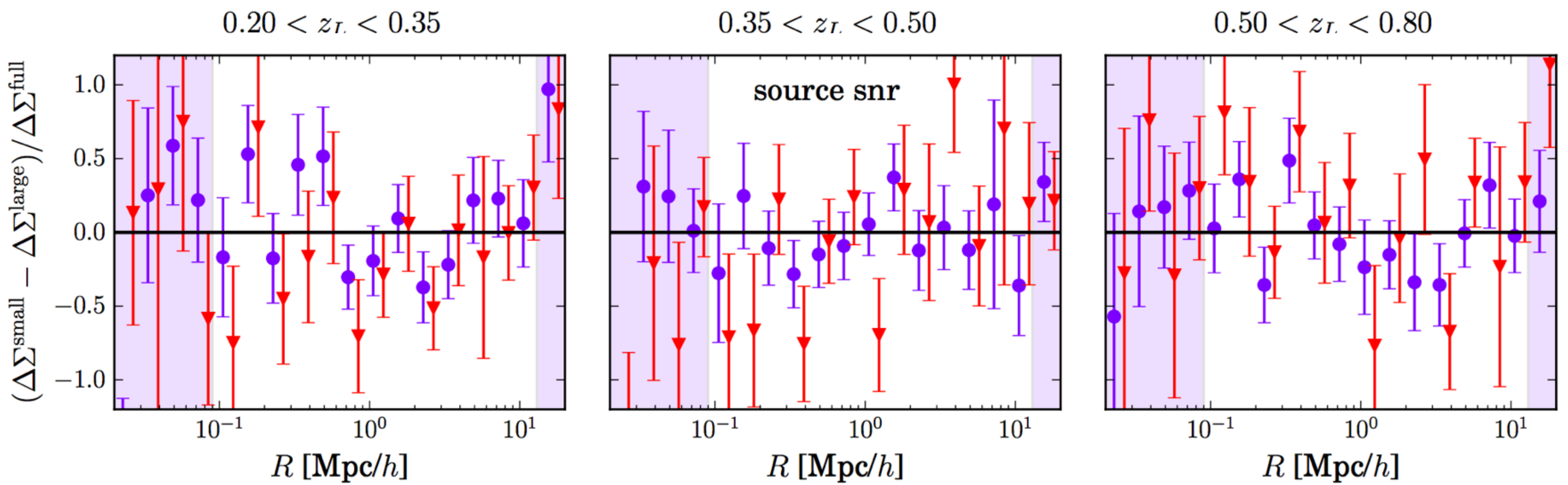
Design a null test to see if there is any residual noise bias in the galaxy shapes.



Credit: Jarvis, et al, 2016



Credit: Jarvis, et al, 2016

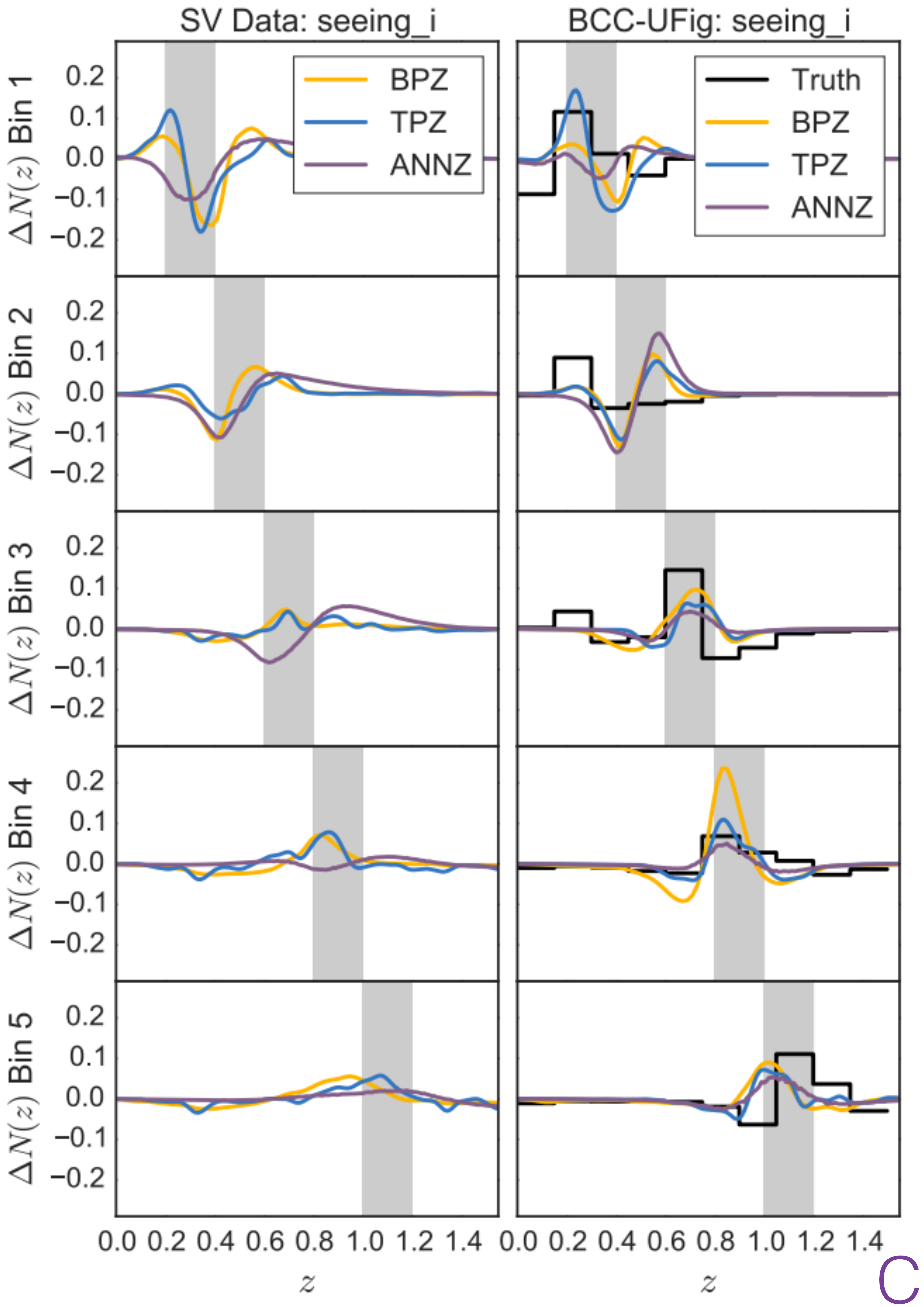


Credit: Clampitt, et al, 2016

Example 5

Photometric errors can be a function of the seeing in the images. This in turn can lead to errors in the colors and consequently biases in the photometric redshifts.

Design a null test to show that variation in seeing does not adversely affect the inferred N(z) across the survey area.



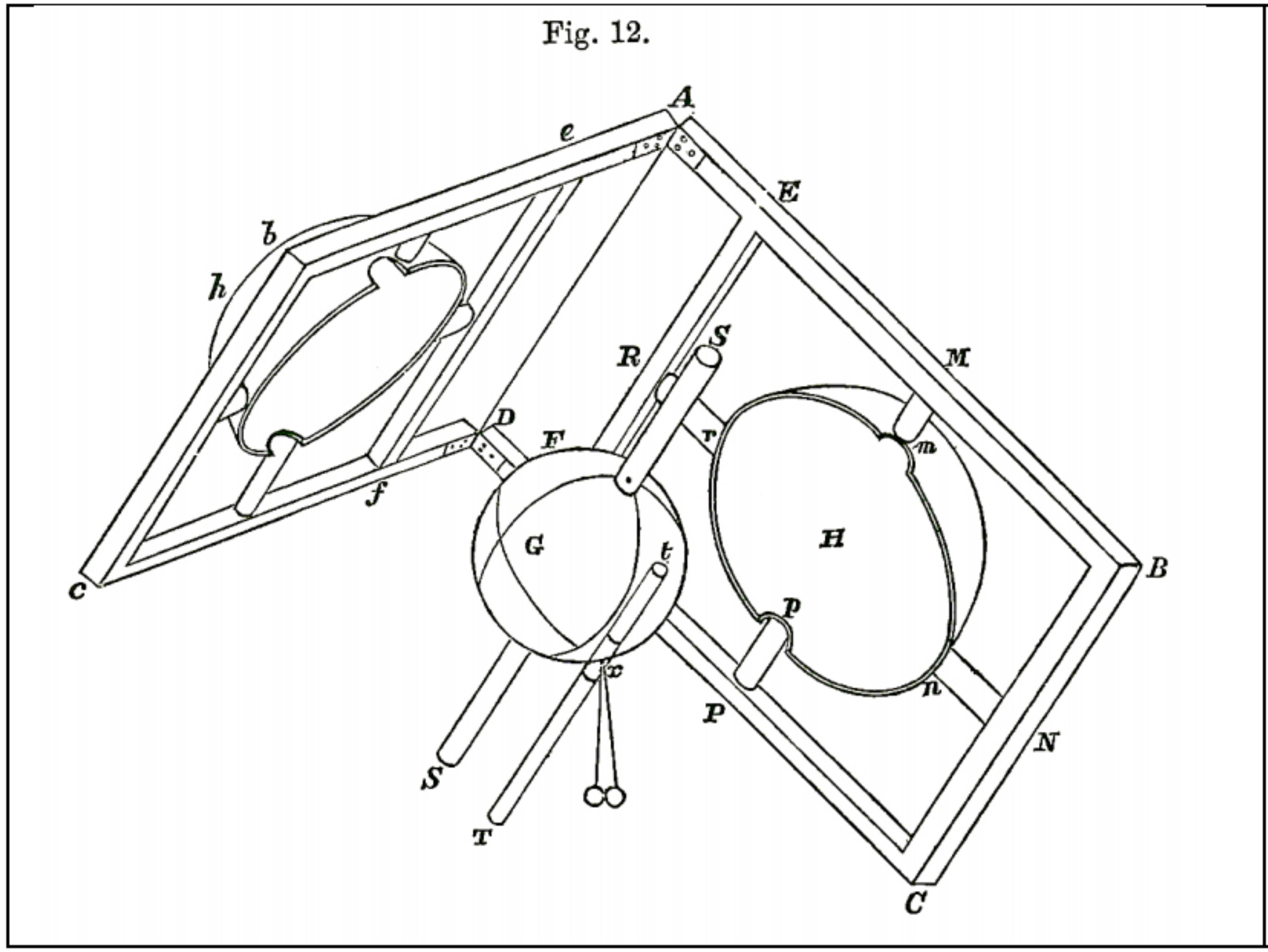
Credit: Leisdedt, et al, 2016

Example 6

One of the first null tests described as such was a test of Coulomb's Law. Specifically that the dependence on distance is really inverse square, not some other power $F \sim r^{-2+\varepsilon}$.

Design a null test to check that the power law index in Coulomb's Law is really 2.

Hint: What measurement should be zero if $\varepsilon=0$, but not otherwise?



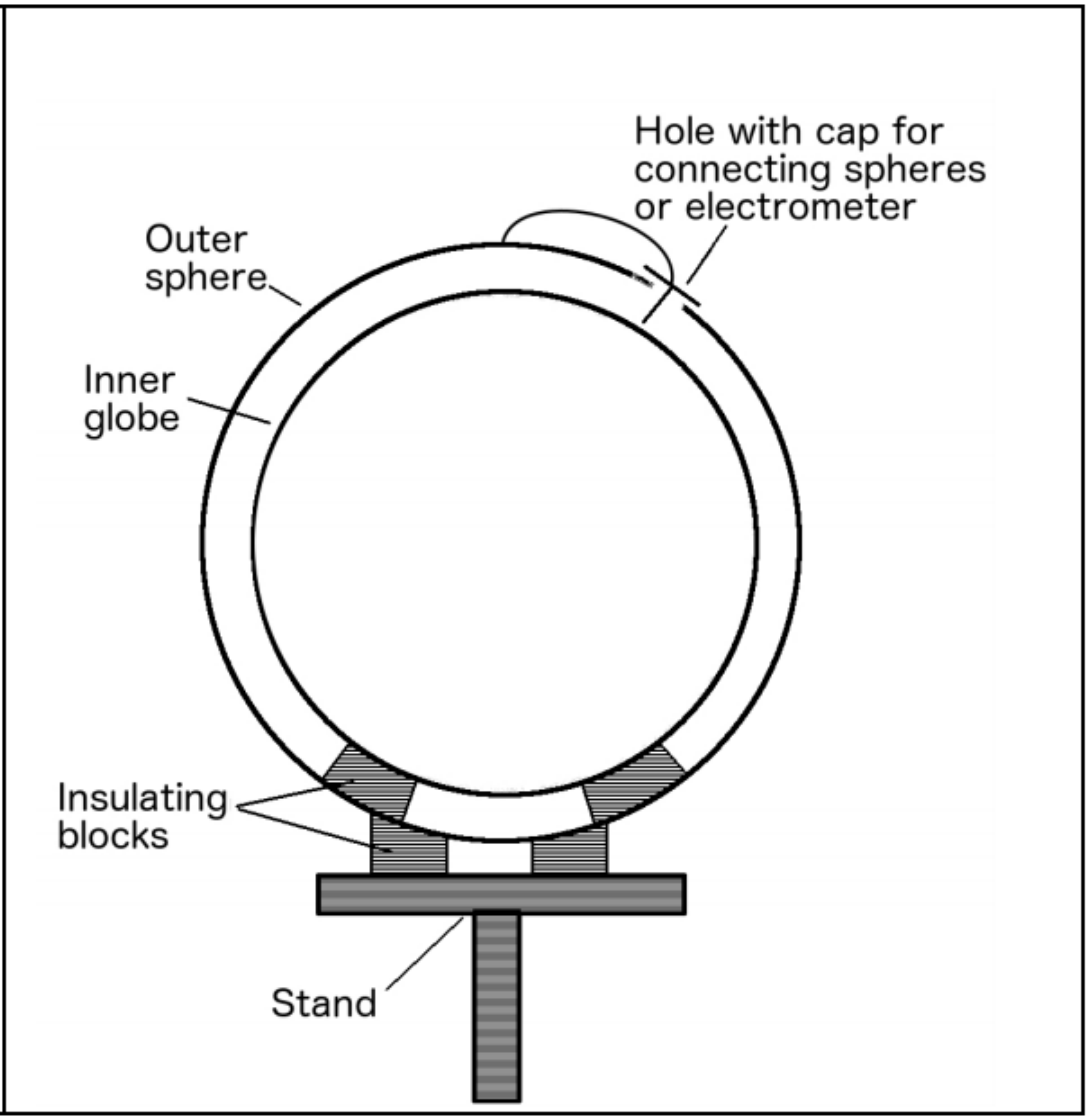


Figure 4. Left: diagram of Cavendish's apparatus showing the hinged frame (AbcDCB), which supported the outer hemispheres (hH), the inner globe (G) supported by an insulator (SS) which stood on the bench, and the pith ball electroscope used for testing (Tt) (Maxwell, 1879 p106). Right: diagram of Maxwell and McAlister's apparatus, showing the inner globe supported on insulating blocks within the outer sphere, which remains closed throughout except for a small capped hole (drawn from the original apparatus shown in Figure 2).

Credit: Falconer, 2016, arXiv:1608.01520

Figure 2. Photograph of Maxwell and McAlister's apparatus . The additional ball hanging from the retort stand has traditionally been considered part of the apparatus, but does not correspond to the brass ball for testing sensitivity of Maxwell's description. Photograph courtesy of the Cavendish Laboratory, Cambridge.

Credit: Falconer, 2016, arXiv:1608.01520

"… let d, be the largest deflexion which could escape observation in the first part of the experiment [the null detection]. Then we know that the potential of the [inner] globe at the end of the first part of the experiment cannot differ from zero by more than

$$\pm \frac{1}{486} \frac{d}{D} V$$

where V is the potential of the shell when first charged…. Now, even in a rough experiment, D was certainly more than 300d. In fact no sensible value of d was ever observed…" (1879 pp418-419; 1881 p79).

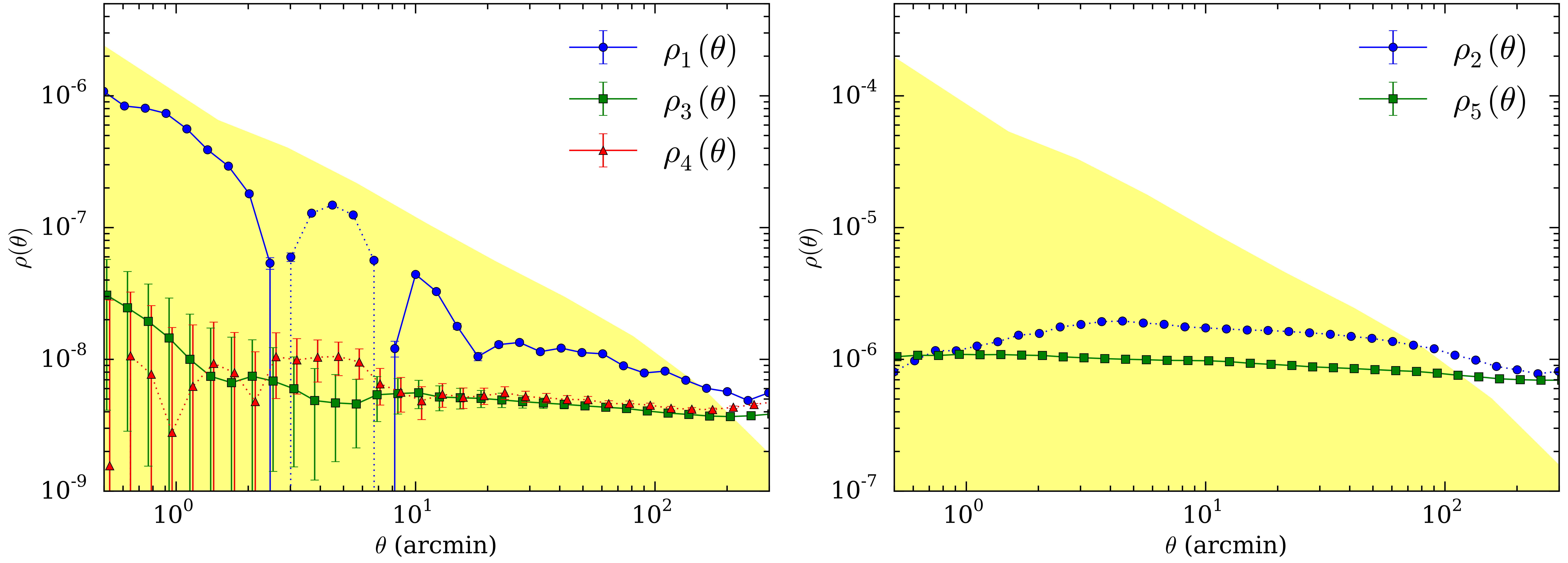
From this, Maxwell and his student McAlister concluded that the power law index n differed from 2 by no more than 1/21600.

Credit: Falconer, 2016, arXiv:1608.01520

Caveats

1. **A null test "failure" is not necessarily important to your science.**

- The systematic error may be measurable at high S/N, but its impact on the science may still be negligible (or at least acceptable).
- You need to propagate any detected systematic error into the relevant science product to see how much it affects the result.



Credit: Jarvis, et al, 2016

Caveats

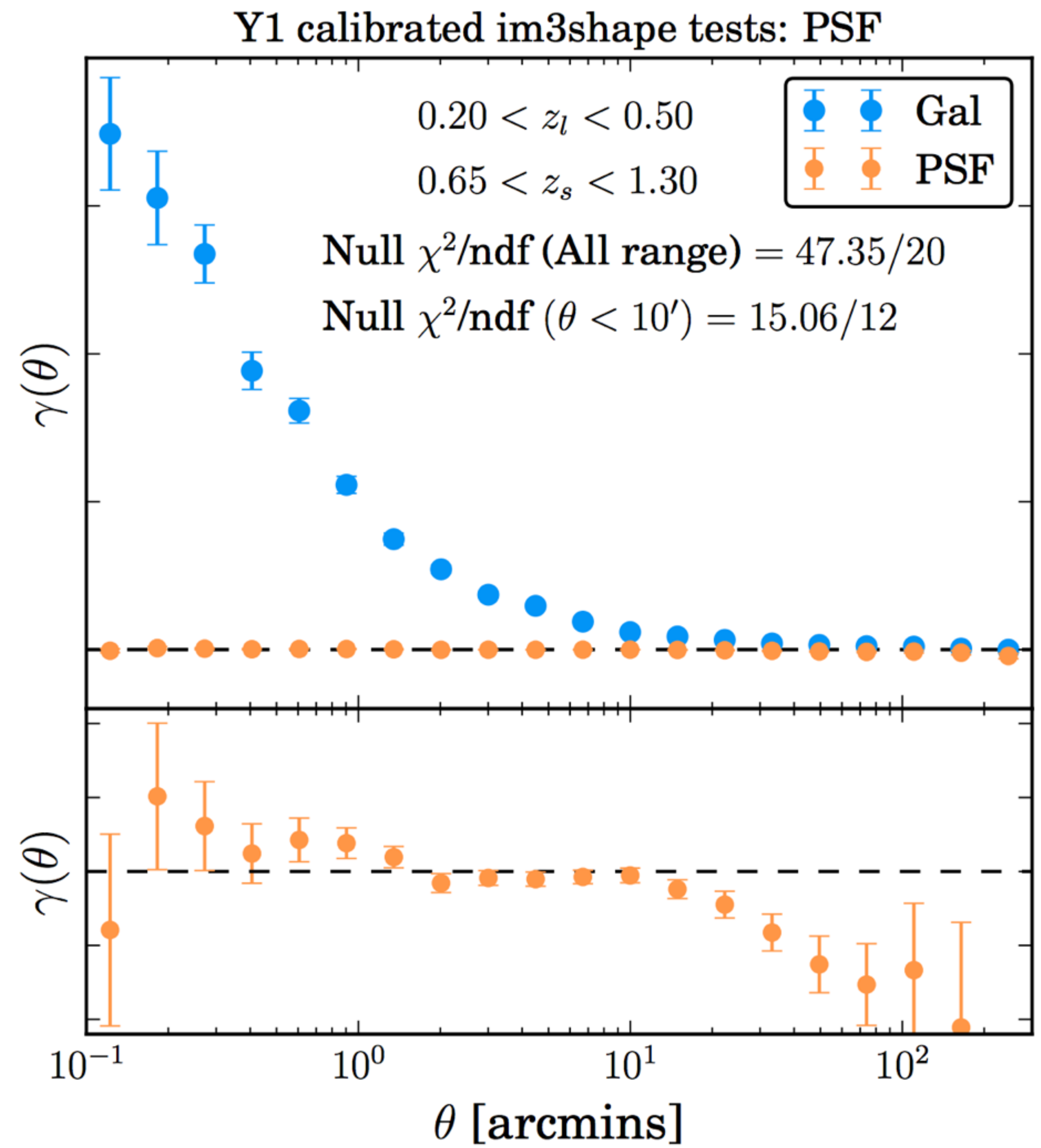
2. **A null test "pass" does not necessarily mean the systematic error being tested does not impact your science.**

- It is hard to design good null tests for many systematic errors.
- Sometimes you can only detect it directly if the bias is very large, but a smaller bias would still be important for your science.

Caveats

3. **Sometimes things you think are null tests don't actually have expectation zero.**

- There can be a tendency to just correlate everything with everything and look at what comes out non-zero.
- For some of these, there are subtle non-systematic-error reasons for the test to be non-null.

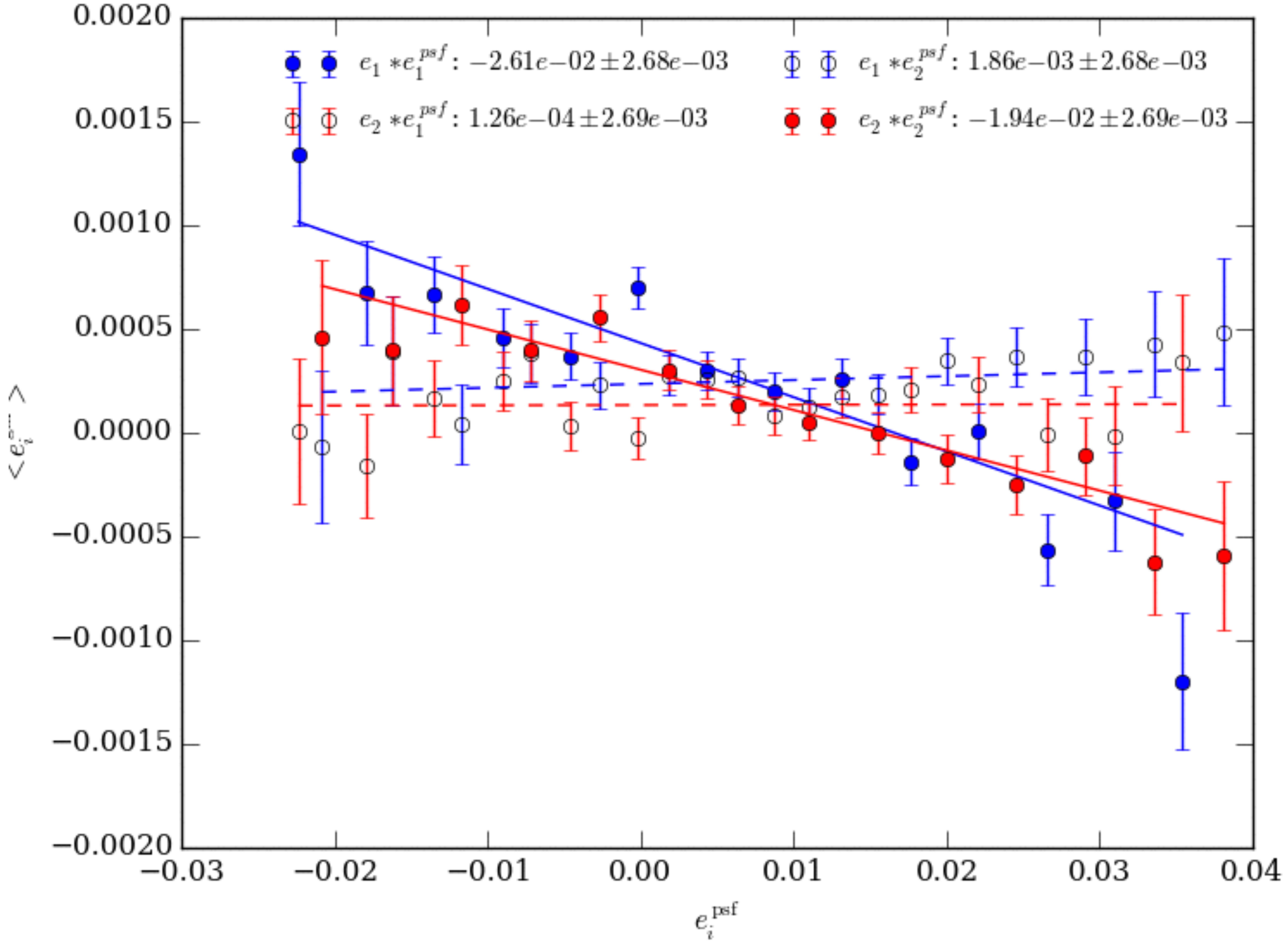


Credit: Judit Prat

Caveats

4. A null test failure might be due to a systematic error that is not the one you thought you were testing.

- It can be easy to go down trails of trying to fix a systematic error, because some null test said it was significant, but where the real cause is something entirely different.

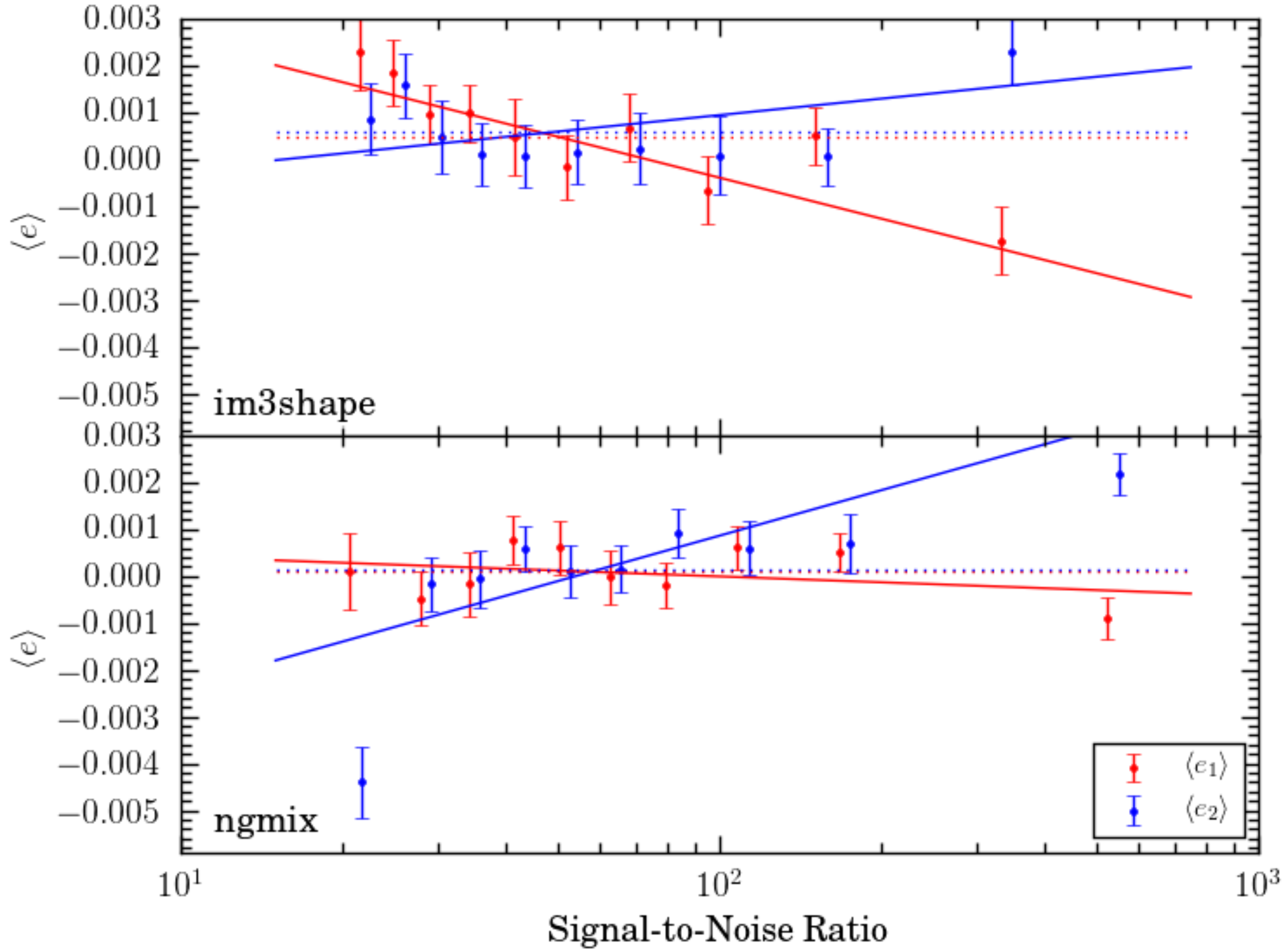

$e_1 * e_1^{psf}$: $-2.61e-02 \pm 2.68e-03$
$e_1 * e_2^{psf}$: $1.86e-03 \pm 2.68e-03$
$e_2 * e_1^{psf}$: $1.26e-04 \pm 2.69e-03$
$e_2 * e_2^{psf}$: $-1.94e-02 \pm 2.69e-03$
0.0020
0.0015
0.0010
0.0005
0.0000
−0.0005
−0.0010
−0.0015
−0.0020
−0.03
−0.02
−0.01
0.00
0.01
0.02
0.03
0.04
e_i^{psf}

Caveats

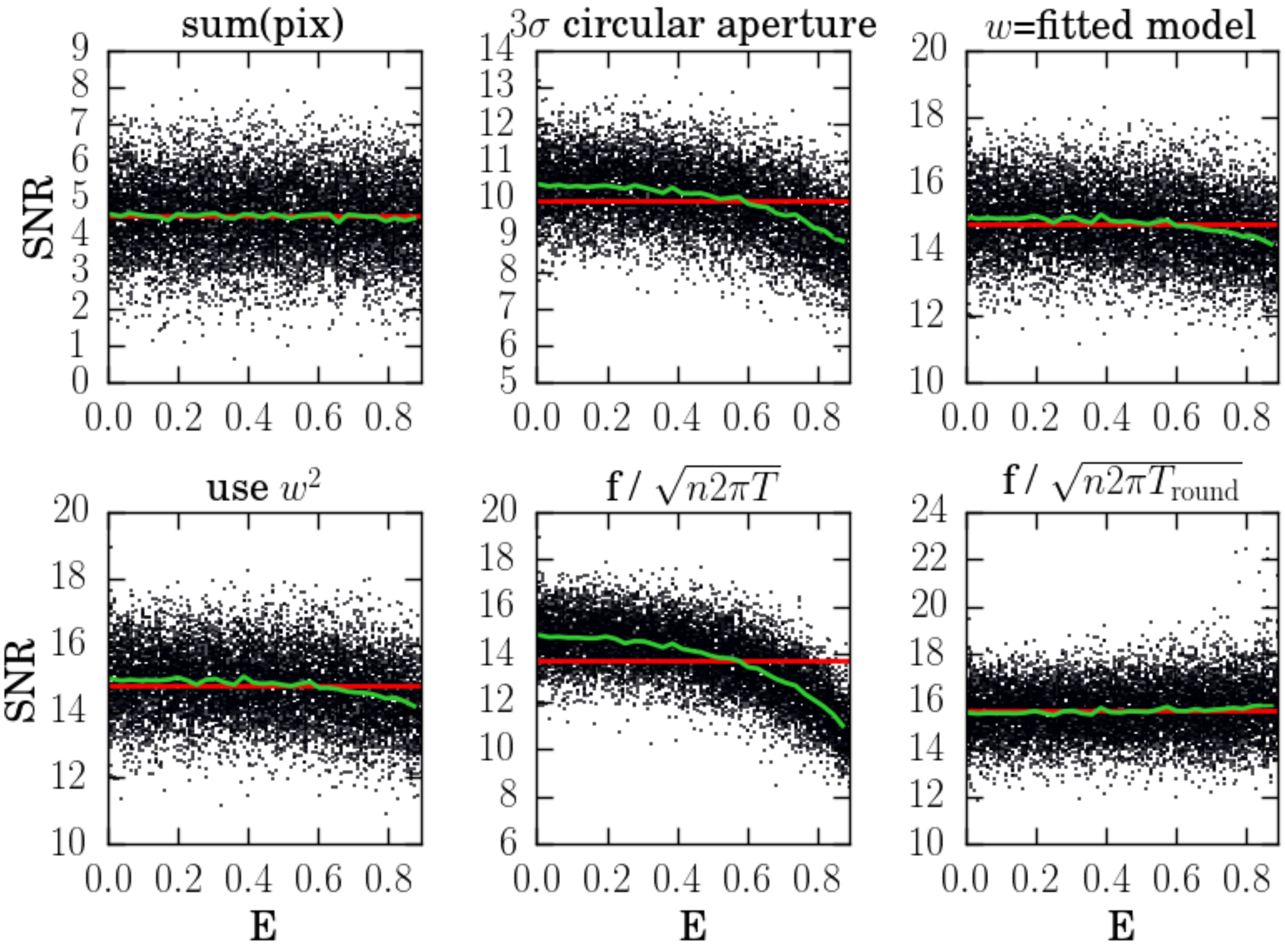
5. **Selection effects can completely mess up your null tests.**

   - The choice of "independent" parameter used for binning may not be completely independent of the measured quantity.

   - If it is not, then you get selection effects at both sides of every bin.



0.003
0.002
0.001
0.000
−0.001
−0.002
−0.003
−0.004
−0.005
$\langle e \rangle$
im3shape
ngmix
$\langle e_1 \rangle$
$\langle e_2 \rangle$
10^1
10^2
10^3
Signal-to-Noise Ratio



sum(pix)
3σ circular aperture
w=fitted model
use w^2
f / $\sqrt{n 2\pi T}$
f / $\sqrt{n 2\pi T_{\mathrm{round}}}$
SNR
SNR
E
E
E

Caveats

6. **Passing all null tests does not mean you don't have any systematics.**

    - It is hard to come up with null tests for every anticipated systematic error.

    - There are also always the Rumsfeldian "unknown unknowns". If you haven't thought of a possible systematic, it could still be lying undetected in your data.

Summary

- Null tests are an invaluable tool for investigating the presence of systematic errors in your data.
- Designing good null tests is a bit of an art, but it is a very useful skill to develop.
- Interpreting null test failures is also tricky and may indicate a problem with the test rather than with the data, so think carefully about what might cause the failure.