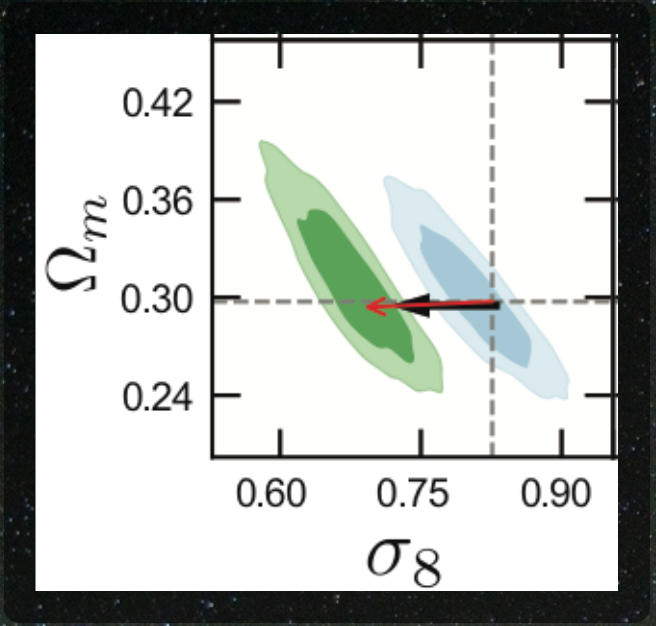


“Blinding multi-probe cosmological experiments”

JM, Gary Bernstein, Dragan Huterer, Franz Elsner, Elisabeth Krause, Aaron Roodman, et al [DES Collaboration]

*MNRAS **494**, no.3, 4454-4470 (2020)*

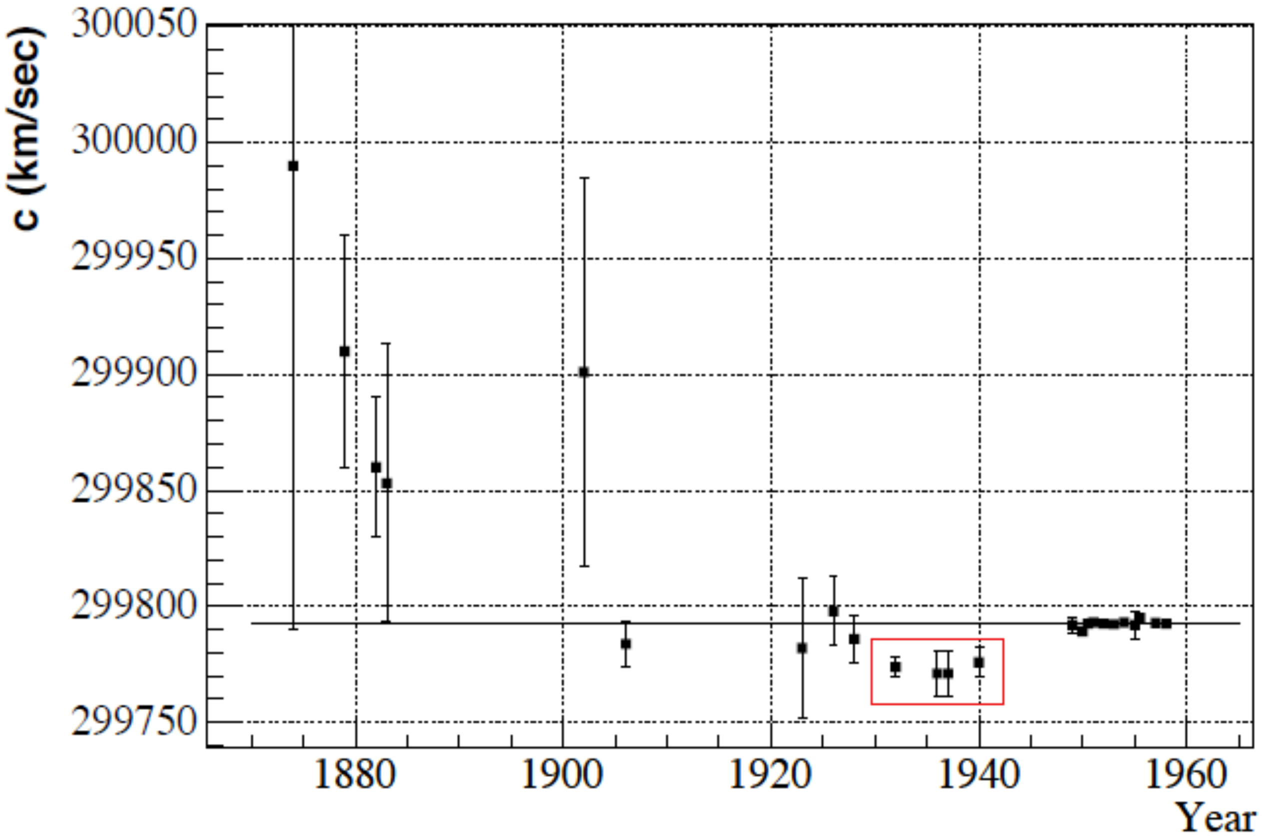
https://arxiv.org/abs/1911.05929

Jessie Muir, Porat Fellow @ KIPAC/Stanford

GCCL seminar, July 17 2020

DARK ENERGY SURVEY

Photo credit: Ross Cawthon (Wisconsin)



a methane absorption line frequency (6). One striking feature is the 17 km/sec shift between the series of experiments from 1930–1940 and later determinations. A fascinating post-mortem on the systematic uncertainties in these experiments (7), noting the different techniques used in the four low results, speculates about one of the sources of bias:

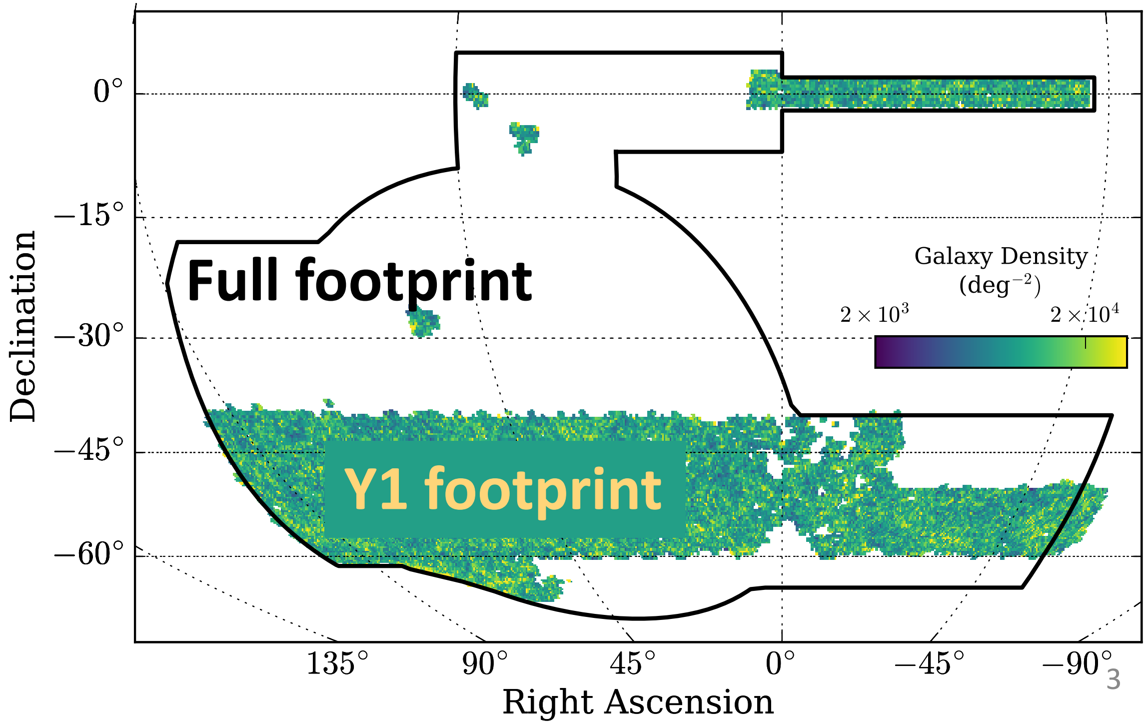
> the investigator searches for the source or sources of such errors, and continues to search until he gets a result close to the accepted value.
>
> *Then he stops!*

Klein and Roodman 2005 review:, “Blind analysis in nuclear and particle physics.”

The Dark Energy Survey

Funded by

- 5000 sq. degree imaging survey using 4m Blanco telescope @CTIO in Chile
  - 6 years of observing ended Jan. 2019
- Y1: **1300 sq. deg** at 40% depth
- Y3: 5000 sq. deg, 50% depth
  - Data processed, analysis in full swing
- Collaborating institutions:

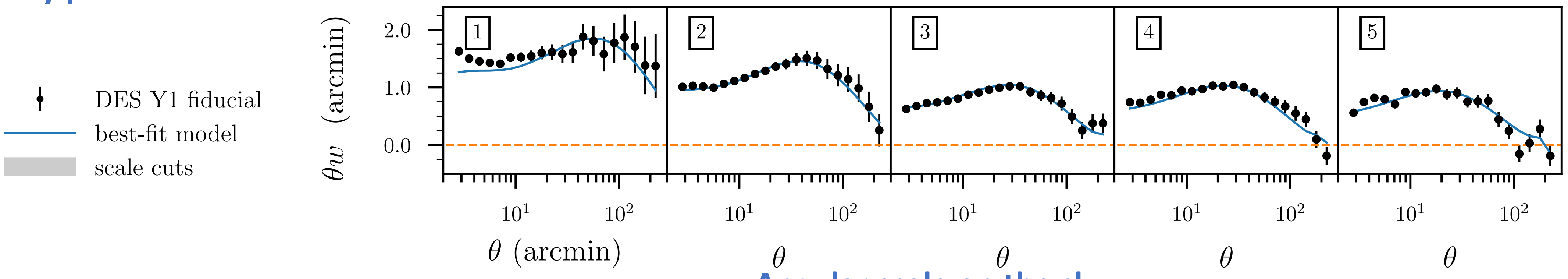


DES key results come from the combined analysis of galaxy clustering and weak lensing.

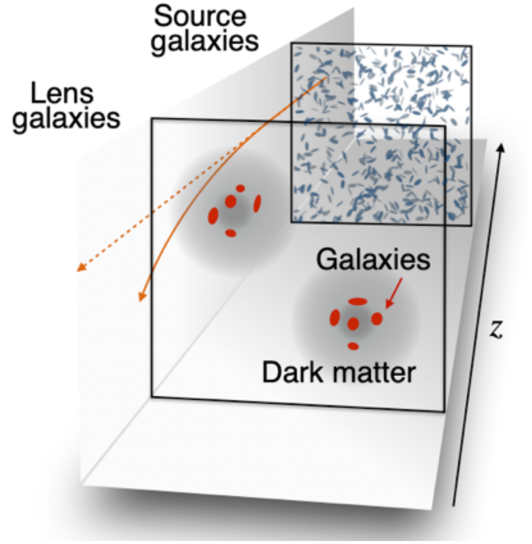
Plots from DES Collaboration 2017, arXiv:1708.01530

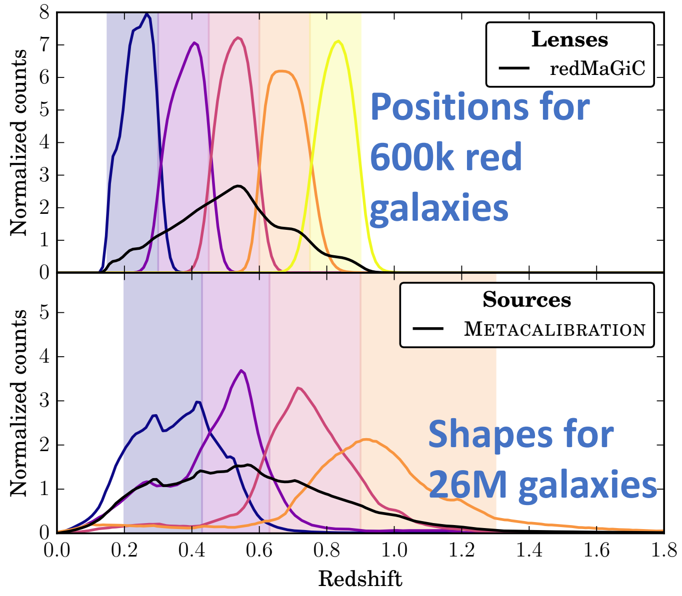
2pt correlations between galaxy positions for DES Y1

(closest) Bins along the line of sight (farthest)



Angular scale on the sky



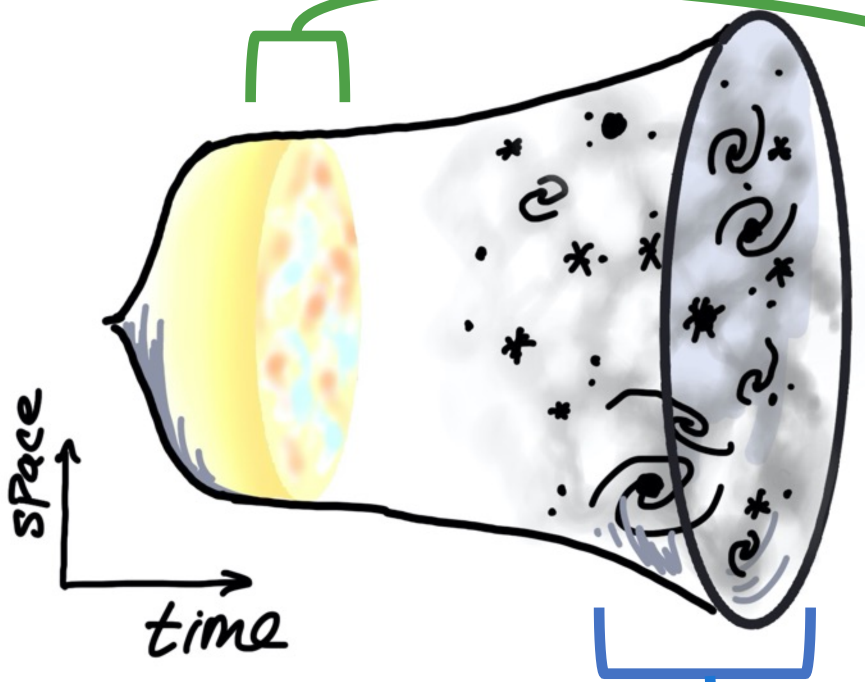


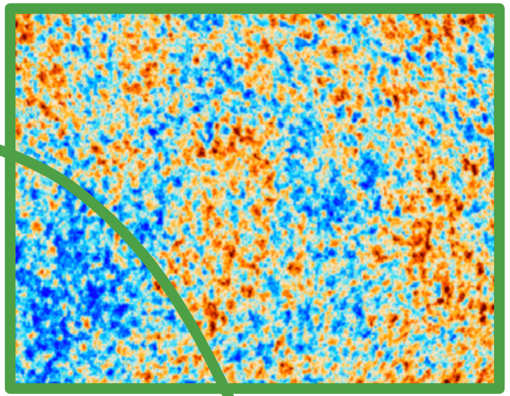
3x2pt:

- **Lens positions x lens position**
- **Lens positions x source shears**
- **source shears x source shears**

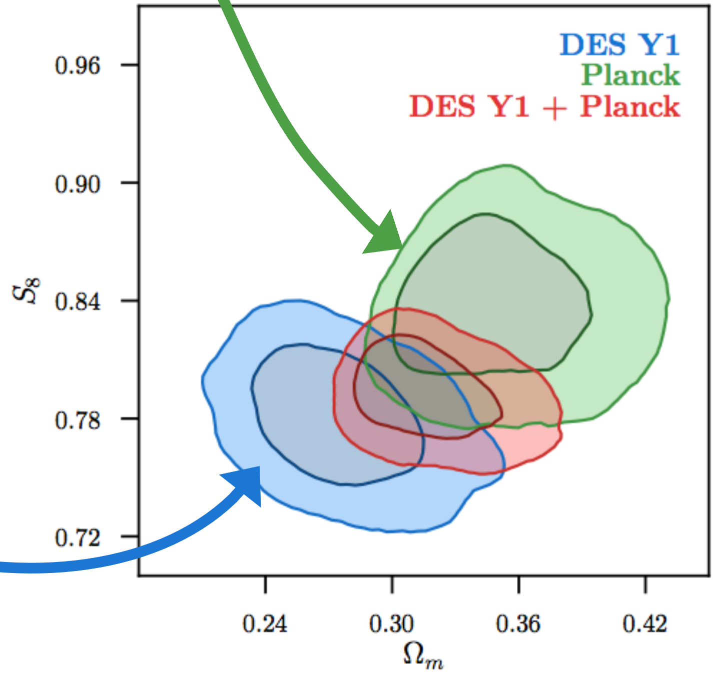
Credit: Yuuki Omori

One way DES can test ΛCDM is by checking for consistency with other measurements.



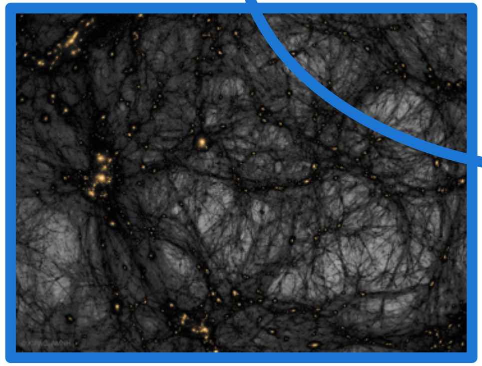


Early-time CMB measurements from Planck



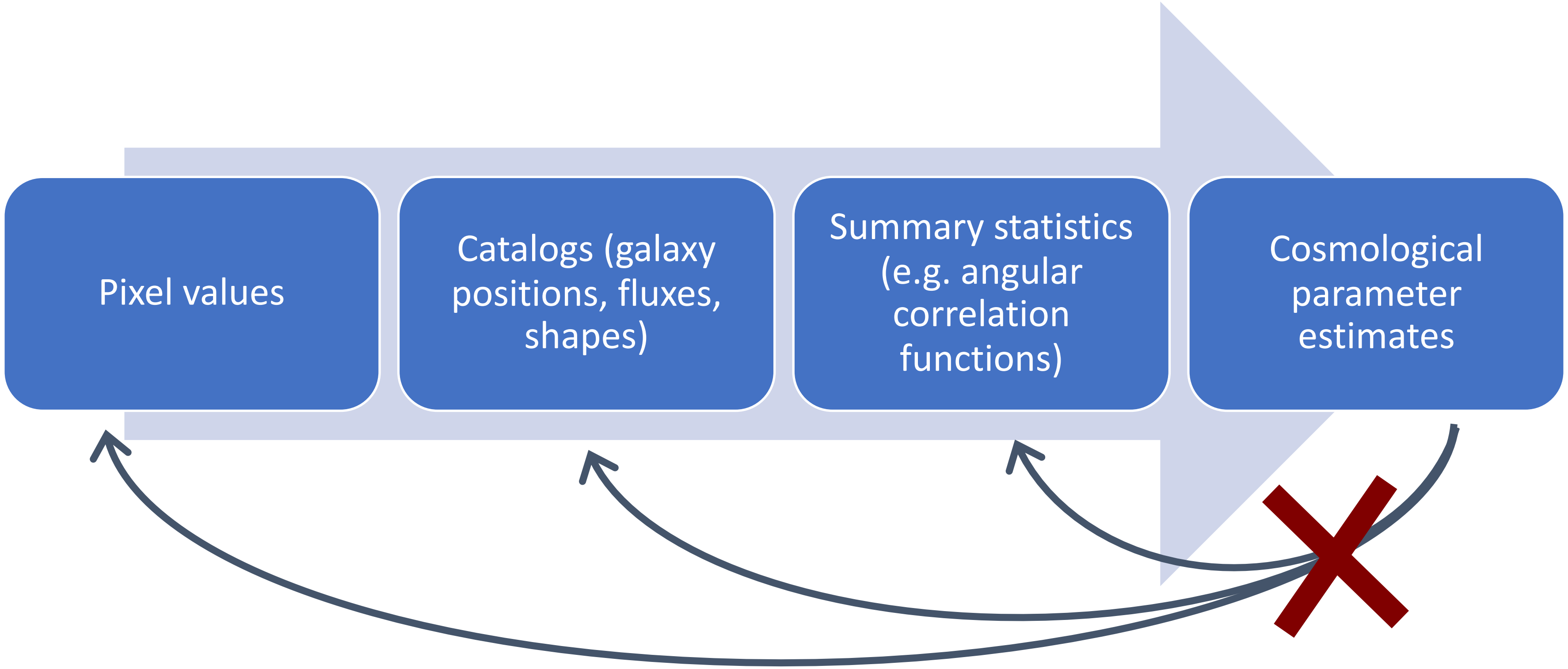
Dark Energy Survey Collaboration 2017, arXiv:1708.01530

Late-time structure measurements from DES



< Credit Tom Abel & Ralf Kaehler (KIPAC, SLAC), AMNH

We don’t want our knowledge of the cosmological parameter values influencing analysis choices.



Blind analysis* = concealed-results analysis

- Framework for experiment/analysis design aimed at protecting against experimenter bias.
- Not one specific method, depends on the details of the experiment.
  - Hidden signal box for rare decay studies (E791, 1990)
  - ”salting”- adding fake signals (LIGO)
  - Catalog-level
    - Shear transformations (KiDS 2015 & since, DES shear-only analysis)
    - Redshift transformations (Brieden et al 2020, 2006.10857, BOSS DR12)
  - Parameter estimation level
    - Parameter value offsets (Conley et al 2006, many SNe analyses, DES Y1 3x2pt)
    - Modify covariance (Sellentin 2019, arXiv:1910.08533, KIDS-450)

*It’s worth noting that while “blinding” is a commonly used and recognized term in our field, by referencing a disability as a metaphor it is an example of ableist language. For clarity (and out of habit) I’ll still be using the term in this talk, but I’m trying to shift towards using alternative phrasing when possible.

Considerations for blinding

1. Concealing the true results
2. Preserving the ability to check for errors
3. Feasible implementation

Goal of analysis:

Given a

$\hat{\mathbf{d}}$ Measured array of observable quantities

We would like to constrain

$\mathbf{\Theta}$ parameters

of a

$\mathbf{d}(\Theta)$ model prediction for the observables.

Blinding involves a transformation

$$\hat{\mathbf{d}} \to \hat{\mathbf{d}}_{\rm bl} = B(\hat{\mathbf{d}})$$

designed to prevent experimenters' analysis decisions from being influenced by how results compare to expectations.

Considerations for blinding

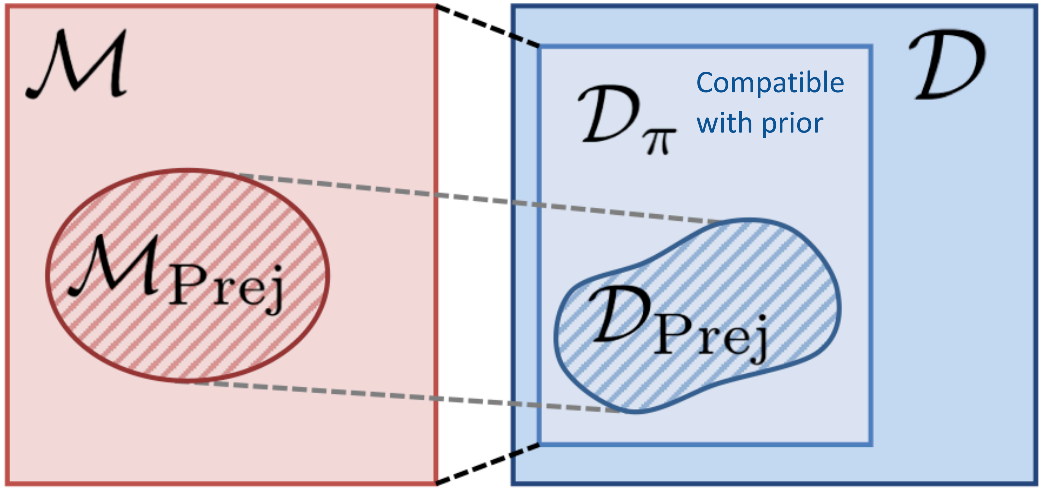
1. Concealing the true results

2. Preserving the ability to check for errors

3. Feasible implementation

Parameter space allowed by priors

Space of possible data realizations



"Prej"=Prejudice, probability distribution describing experimenter expectations

Blinding transformation must be capable of changing data so that it is equally probable that

$$\mathrm{Prej}(\Theta_{\rm unbl}) \gg \mathrm{Prej}(\Theta_{\rm bl})$$

or

$$\mathrm{Prej}(\Theta_{\rm unbl}) \ll \mathrm{Prej}(\Theta_{\rm bl}).$$

Where the arguments are:

$$\Theta_{\rm unbl} = \underset{\Theta}{\mathrm{argmax}} \left\{ P\left(\Theta|\hat{\mathbf{d}}\right)\right\}$$

$$\Theta_{\rm bl} = \underset{\Theta}{\mathrm{argmax}} \left\{ P\left(\Theta|B(\hat{\mathbf{d}})\right)\right\},$$

Considerations for blinding

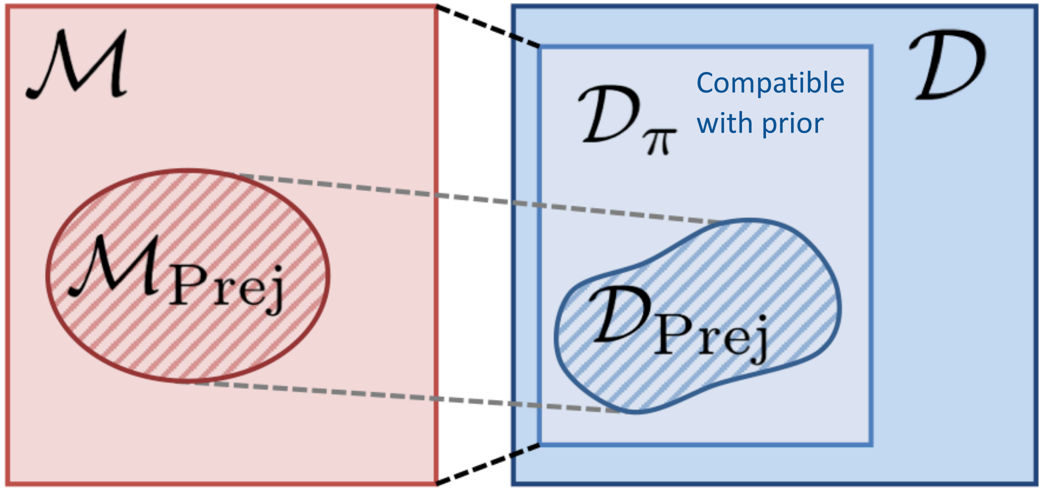
1. Concealing the true results
2. **Preserving the ability to check for errors**
3. Feasible implementation

Transformation should preserve the internal consistency of data.

$$\frac{\mathcal{L}\left(\hat{\mathbf{d}}_{\mathrm{bl}}|\Theta_{\mathrm{bl}}\right)}{\mathcal{L}\left(\hat{\mathbf{d}}_{\mathrm{unbl}}|\Theta_{\mathrm{unbl}}\right)} \approx 1$$

Parameter space allowed by priors

Space of possible data realizations



"Prej"=Prejudice, probability distribution describing experimenter expectations

Caveat: validation tests generally require some implicit modeling assumptions!

Consistently transforming data for multiple observables is challenging.

1. Concealing the true results
2. Preserving the ability to check for errors
3. **Feasible implementation**

DES Y1: Hide axes or introduce unknown offset

What data are we transforming?

Pixel values

Catalogs (galaxy positions, fluxes, shapes)

Summary statistics (e.g. angular correlation functions)

Cosmological parameter estimates

Harder to preserve consistency of multiple observables. More steps to undo.

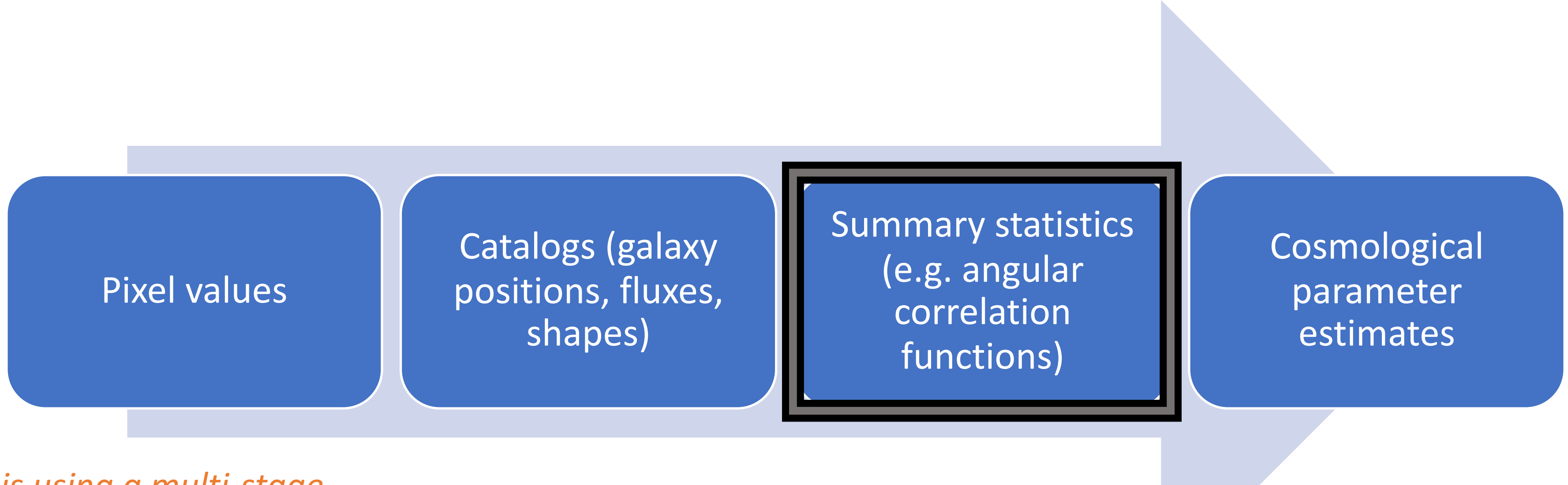
Easier to implement, easier to (on purpose or accidentally) undo.

New multi-probe blinding method: transform summary statistics

$$B(\hat{\mathbf{d}}) = \hat{\mathbf{d}} + \Delta_{\mathrm{bl}}$$

$$\Delta_{\mathrm{bl}} = \mathbf{d}(\Theta_{\mathrm{ref}} + \Delta\Theta) - \mathbf{d}(\Theta_{\mathrm{ref}})$$

$\hat{\mathbf{d}}$	Measured array of observable quantities
$\mathbf{d}(\Theta)$	Model prediction for the observables



In practice, DES Y3 is using a multi-stage strategy that also includes catalog-level shear transformation & parameter offsets.

To reveal true results, need to redo parameter estimation.

Application: DES Y3 3x2pt

$$B(\hat{\mathbf{d}}) = \hat{\mathbf{d}} + \Delta_{\rm bl}$$

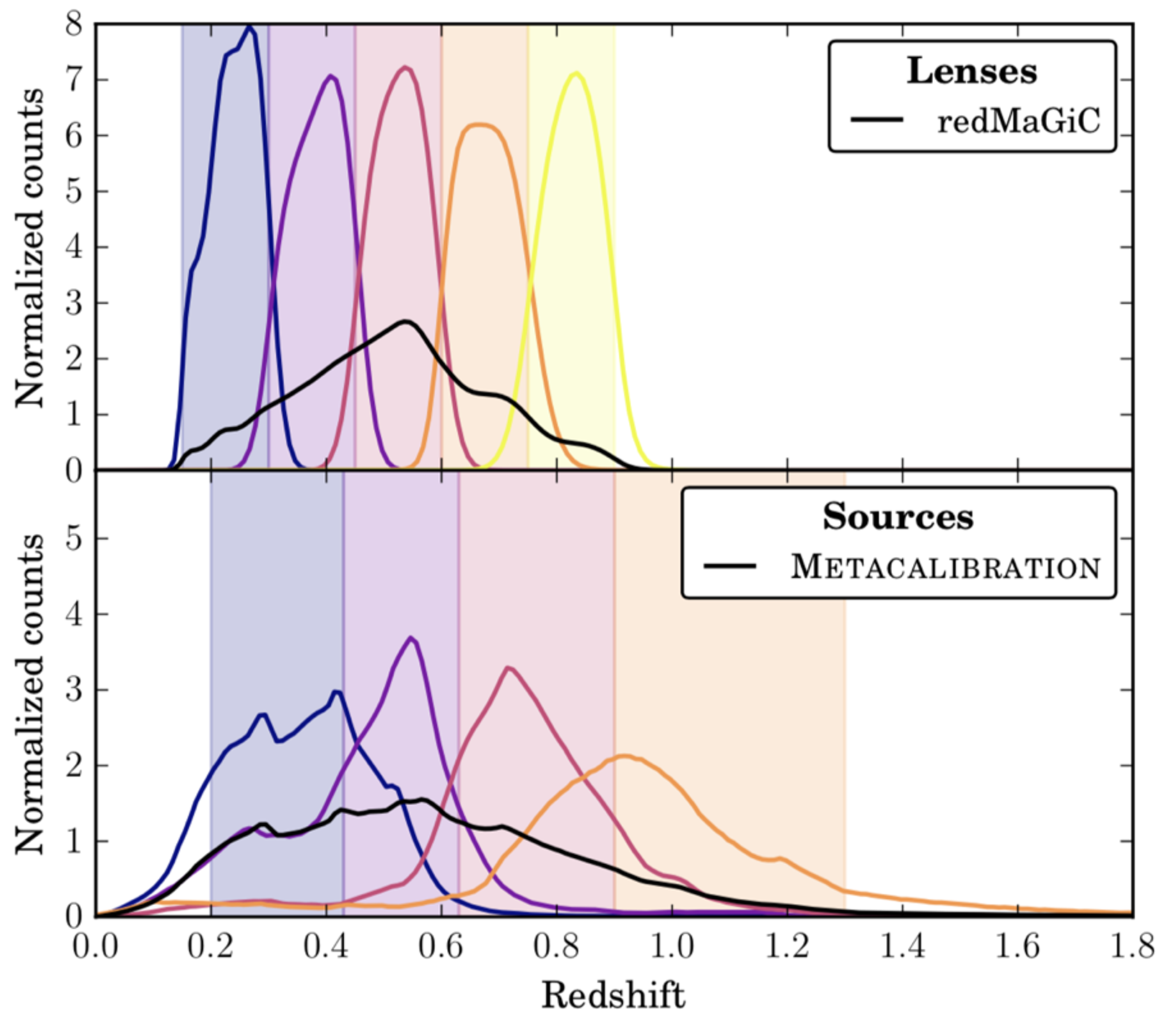
$$\Delta_{\rm bl} = \mathbf{d}(\Theta_{\rm ref} + \Delta\Theta) - \mathbf{d}(\Theta_{\rm ref})$$

Data: 2pt correlations binned in angle and z

- Lens positions x lens positions
- Lens positions x source shears
- source shear x source shears

Model: wCDM

Focused on shifting w and σ_8 results



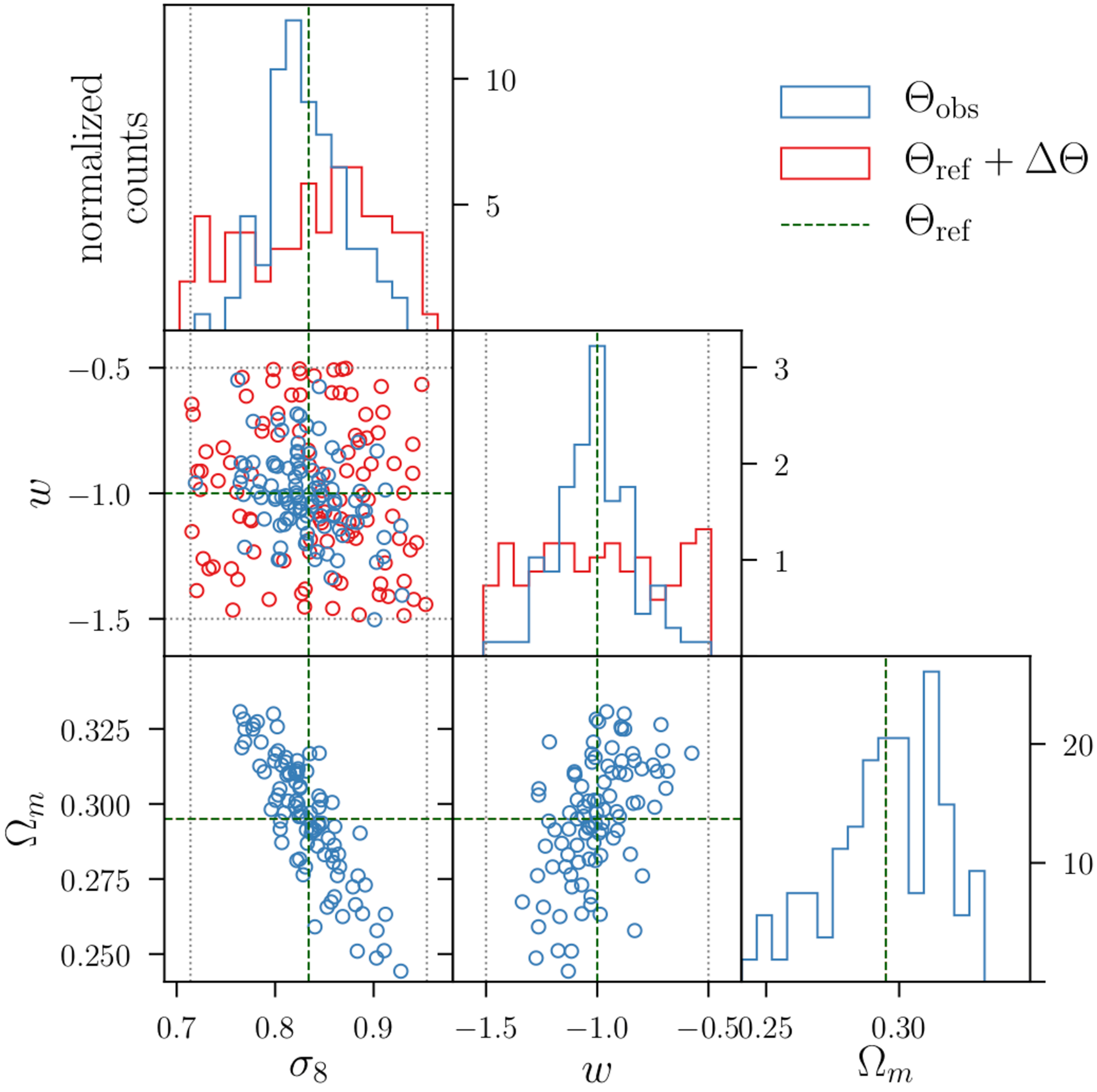
Testing performance

- Using DES Y1 pipeline, with covariance*0.27 for increased sky area.
- Analyze synthetic data generated at Θ_{obs}
  - 100 realizations drawn from Gaussian distribution in subset of cosm parameters
  - Focused on "noiseless data", equal to theory prediction at Θ_{obs}

$$\hat{\mathbf{d}} = \mathbf{d}(\Theta_{\rm obs})$$

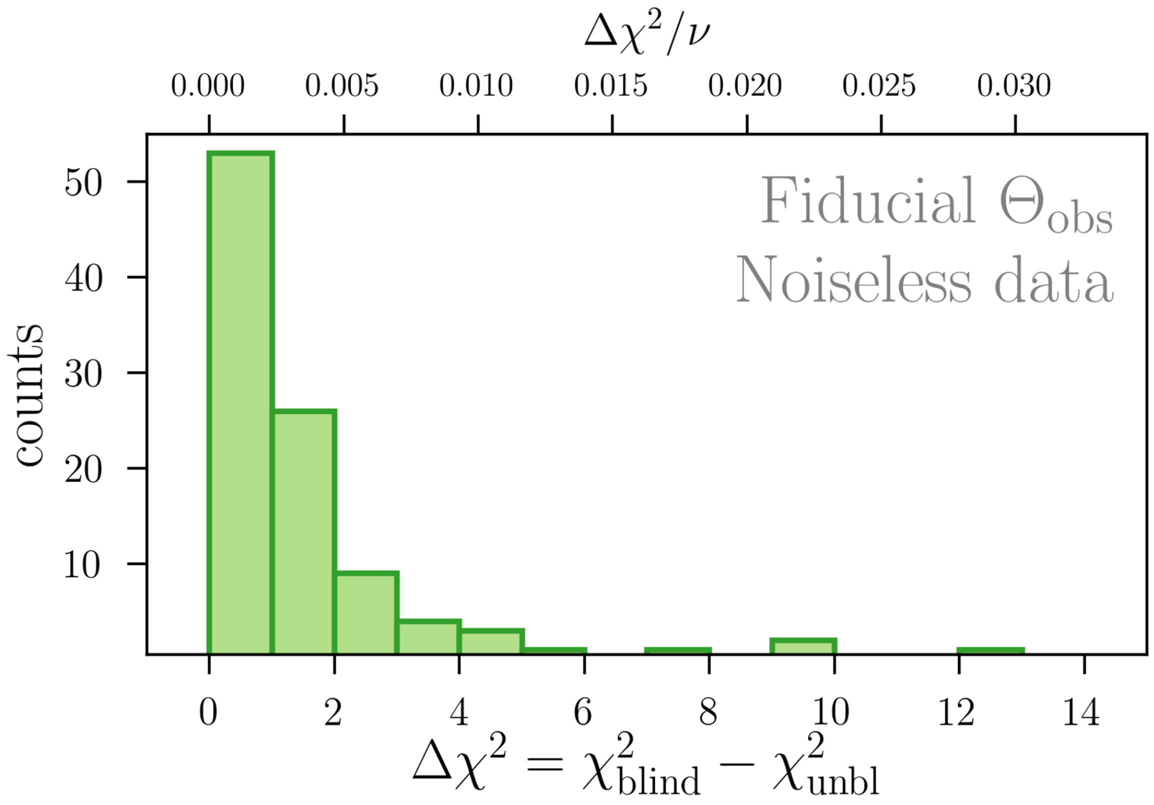
- Transform using predictions at Θ_{ref}, $\Theta_{ref}+\Delta\Theta$
  - Θ_{ref} fixed to fiducial cosmology
  - $\Delta\Theta$ drawn from flat distribution in w and σ_8

$$\Delta_{\rm bl} = \mathbf{d}(\Theta_{\rm ref} + \Delta\Theta) - \mathbf{d}(\Theta_{\rm ref})$$
$$B(\hat{\mathbf{d}}) = \hat{\mathbf{d}} + \Delta_{\rm bl}$$



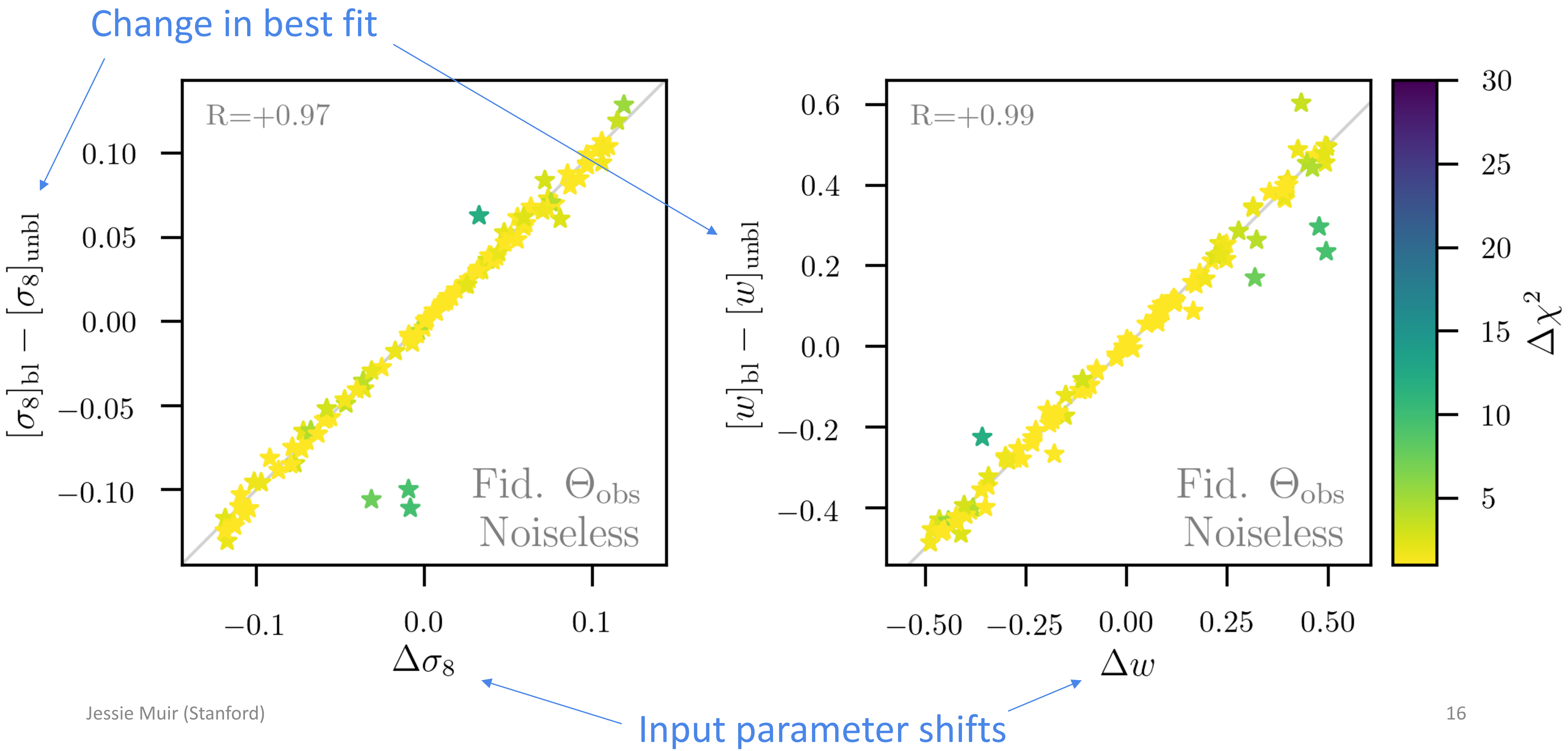
Preservation of 3x2pt's internal consistency

- For reach realization, find change in χ^2 due to blinding.
  - Measures the extent to which the signal part of the blinded data doesn't match any model prediction
- Chisq values obtained from a max-likelihood search.
  - Somewhat tricky given 27D parameter search: if search fails, $\Delta\chi^2$ is overestimated.
  - Ran Multinest chains for selected realizations to improve estimate.
- Desired threshold $\Delta\chi^2$ <30
  - This is 1 sigma in the expected chisq distribution
  - 30=sqrt(2*[457 datapoints - 27 params])



All realizations below threshold, internal consistency is preserved.

This method can significantly shift results.



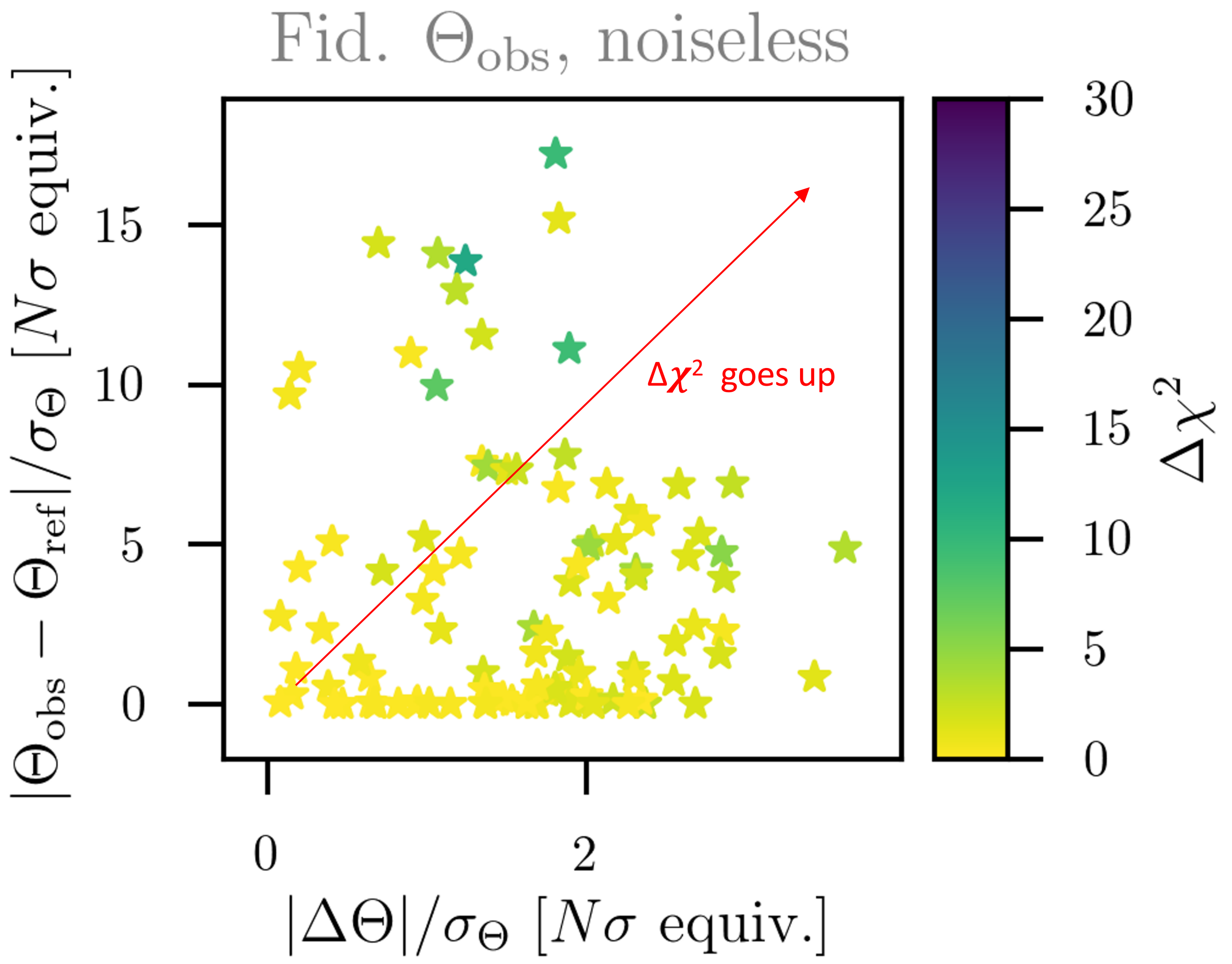
Performance as a function of parameter choices

Recall, for this test:

$$\hat{\mathbf{d}} = \mathbf{d}(\Theta_{\rm obs})$$

$$B(\hat{\mathbf{d}}) = \hat{\mathbf{d}} + \Delta_{\rm bl}$$

$$\Delta_{\rm bl} = \mathbf{d}(\Theta_{\rm ref} + \Delta\Theta) - \mathbf{d}(\Theta_{\rm ref})$$



Will have perfect performance ($\Delta\chi^2$=0) if

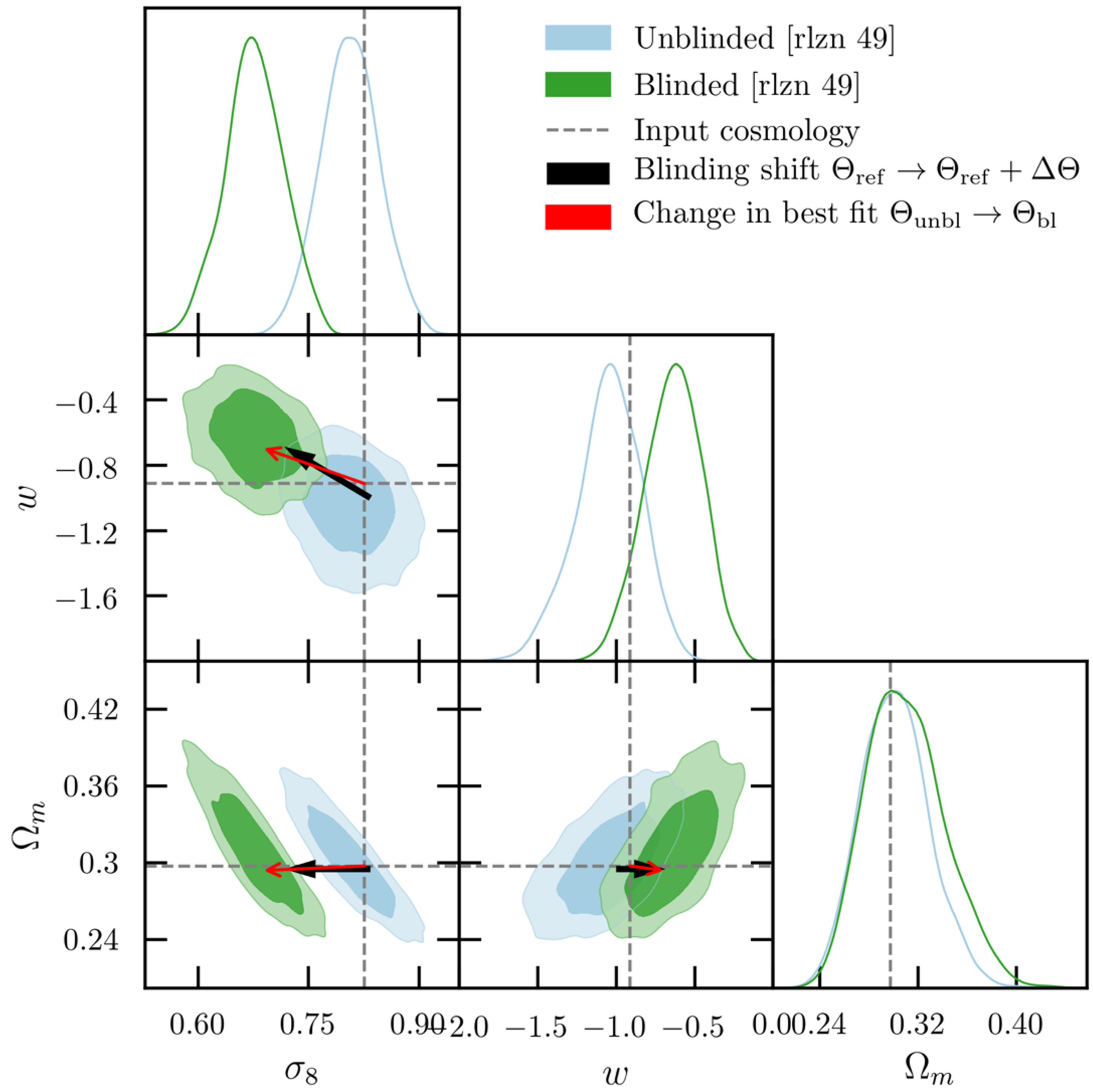
- $\Delta\Theta = 0$ (no transformation)
- $\Theta_{\rm ref} - \Theta_{\rm obs} = 0$ (perfectly subtract and replace signal)

Effect on posteriors

For an example realization with $\Delta\chi^2$=1.9.

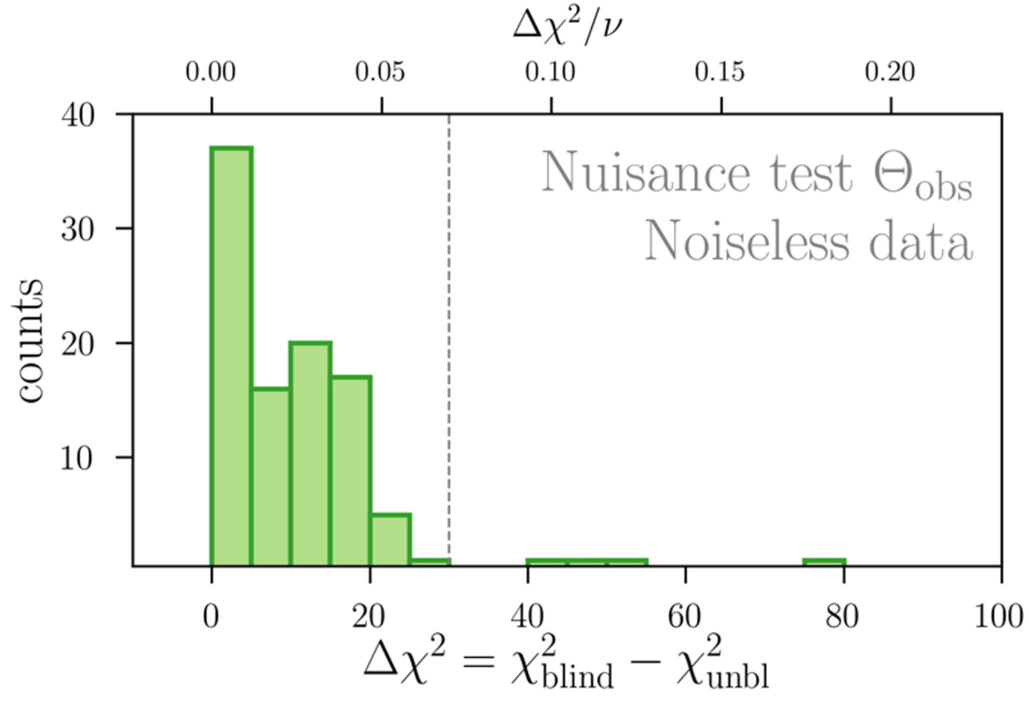
Warning:

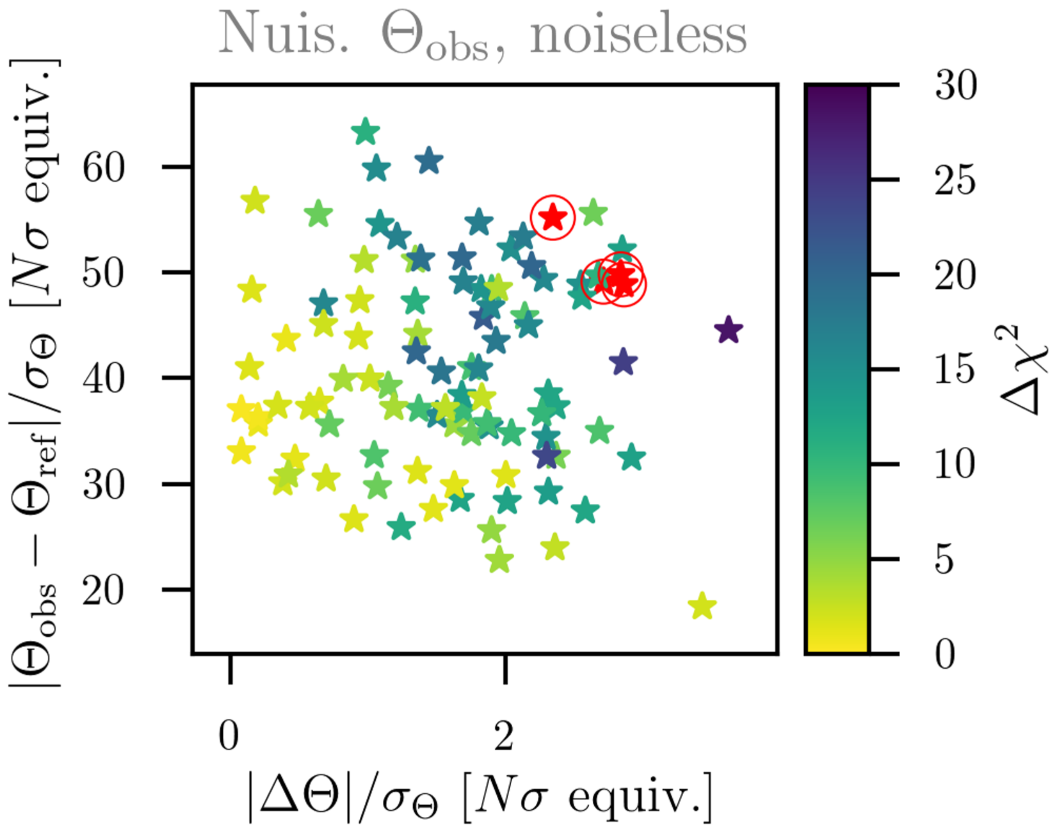
- for some of the realizations with $\Delta\chi^2$~10, contours get pushed into nuisance parameter prior boundary



Where does performance break down? "Nuisance test"

- When generating synthetic data,
  - Use same cosmology parameters as fiducial test
  - Draw nuisance params from wide flat distribution, +/- 3σ from DES Y1 3x2pt posteriors
    - m_nu, 2 parameter IA model, 5 lens galaxy bias, 9 photo-z biases, 4 shear calibration params
- Use same input parameter shift realizations as fiducial test.





Looking forward

$$B(\hat{\mathbf{d}}) = \hat{\mathbf{d}} + \Delta_{\mathrm{bl}}$$

$$\Delta_{\mathrm{bl}} = \mathbf{d}(\Theta_{\mathrm{ref}} + \Delta\Theta) - \mathbf{d}(\Theta_{\mathrm{ref}})$$

$\hat{\mathbf{d}}$	Measured array of observable quantities
$\mathbf{d}(\Theta)$	Model prediction for the observables

- Applicable to any summary statistic that is used as input for parameter estimation.
- In DES Y3, in addition to 3x2pt,
  - Being used for DES X SPT lensing "6x2pt" analysis.
  - Tests planned for application to galaxy cluster analysis

Summary

- Blind/concealed-results analysis is a tool for protecting results from unconscious experimenter bias.

- We introduce a blinding method, now being used in the DES Year 3 analysis, which works by transforming the galaxy clustering and weak lensing two-point correlation functions.
  - Paper link: https://arxiv.org/abs/1911.05929

- This transformation could be applied to any summary statistic that is an input for parameter estimation, and could be a useful tool for future cosmology experiments.

Extra slides

Looking forward

Size of input parameter shifts

Distance between true params and reference assumed for blinding.

Covariance

$$\Delta\chi^2 \sim |\Delta\Theta|^2 |\Theta_{\rm obs} - \Theta_{\rm ref}|^2 / |C|$$

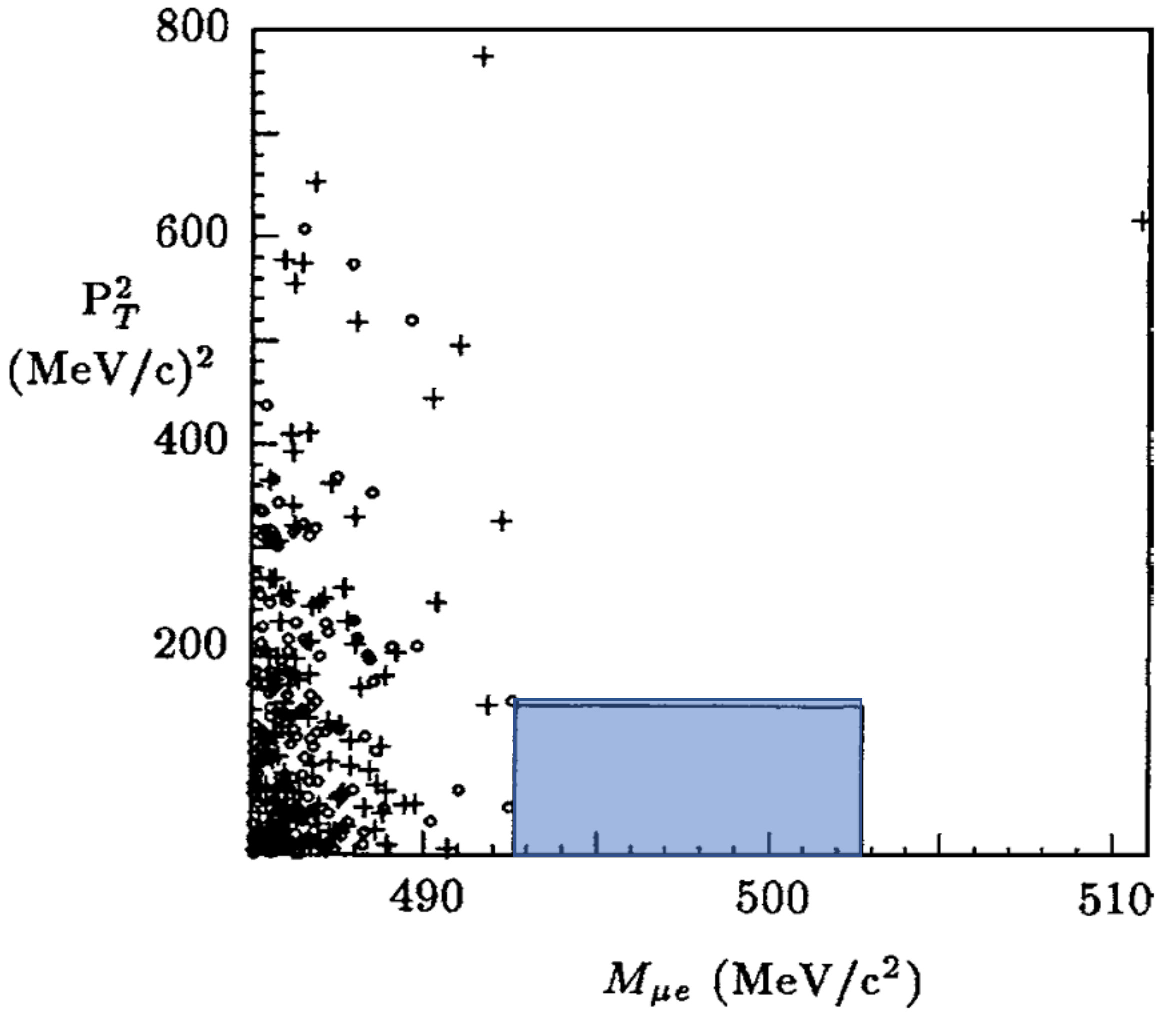
If measurement errors decrease by factor α<1

- $C \rightarrow \alpha^2 C$
- $\Delta\Theta \rightarrow \alpha\Delta\Theta$ - smaller range of shifts needed to obscure agreement with prejudice
- $(\Theta_{obs}-\Theta_{ref}) \rightarrow \sim\alpha(\Theta_{obs}-\Theta_{ref})$ to the extent that max size of difference is set by priors and that reduces with increasing precision

Overall scaling: $\Delta\chi^2 \sim \alpha^2$

Example: E791 experiment (1990, BNL)

- Search for rare Kaon decays
- Result: number events detected in signal region.
- Avoided looking in "hidden signal box" until cuts removing background are fixed.



Klein and Roodman 2005