DESI (edited)

An *Unofficial* DESI Analysis

Turning 5 million galaxies into 3 numbers…

Oliver H. E. Philcox

Junior Fellow @ Simons Foundation

Postdoc @ Columbia

Assistant Prof @ Stanford [from September]

Acknowledgements

Reanalyzing DESI DR1:
1. ΛCDM Constraints from the Power Spectrum & Bispectrum

Anton Chudaykin,[1,*] Mikhail M. Ivanov,[2,3,†] and Oliver H. E. Philcox[4,5,6,7,‡]

(ChatGPT)

arXiv:2507.13433

Additional thanks

- **"East Coast" team**: Marko Simonovic, Matias Zaldarriaga, Giovanni Cabass, Kazu Akitsu, Stephen Chen
- **"West Coast" team:** Guido D'Amico, Leonardo Senatore, Pierre Zhang, Matt Lewandowski

and, of course, the **DESI collaboration!**

Galaxy Clustering Experiments

The Past

- BOSS [2009 — 2014]

The Present

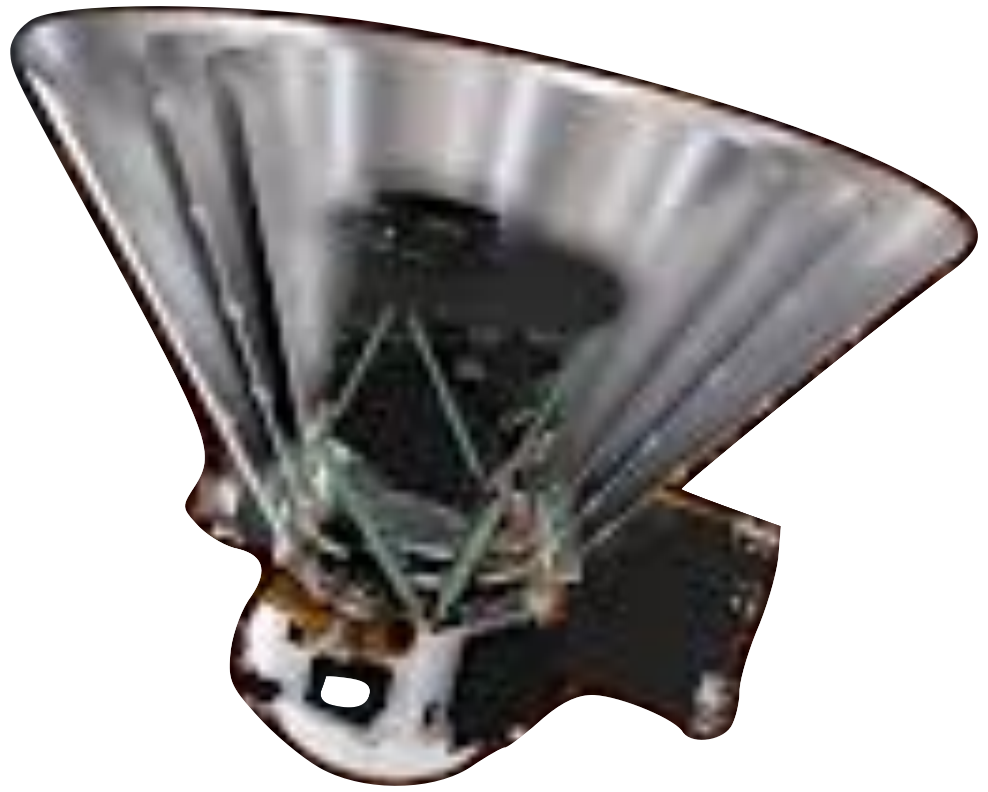
- DESI [2021 — 2026]
- Euclid [2024 — 2030]

The Nearly

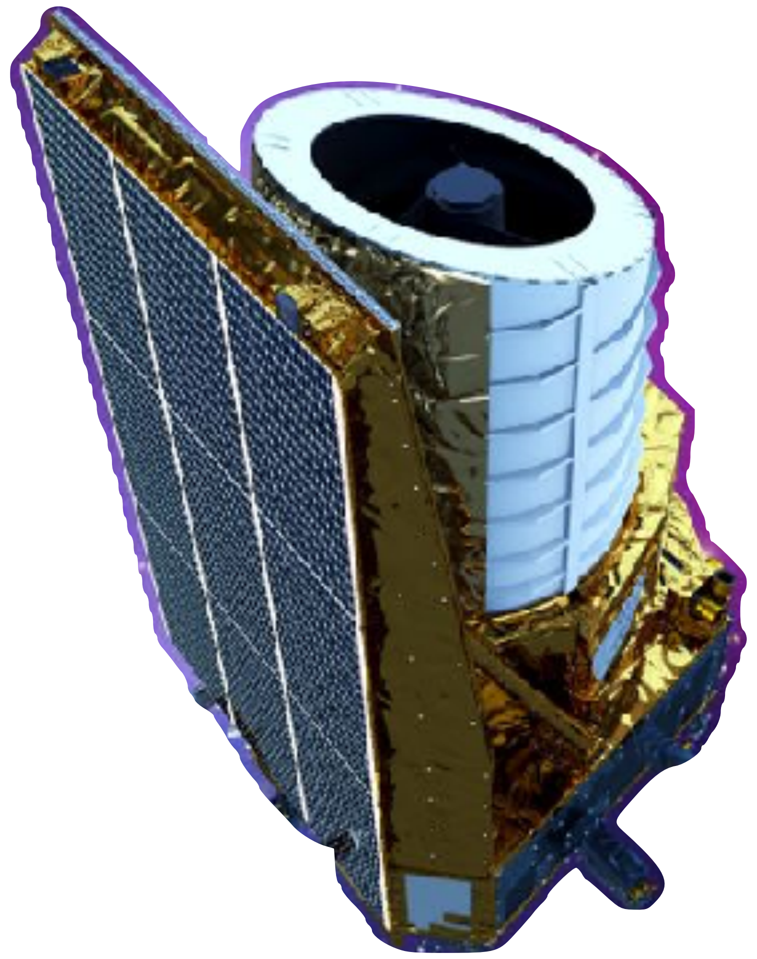
- SphereX [2025 — 2027]
- Rubin [2025 — 2035]
- Roman [launches 2027]

The Future

- Spec-S5? [planning]



SPHEREx



Euclid

DESI

Roman

Rubin

What Do These Probe?

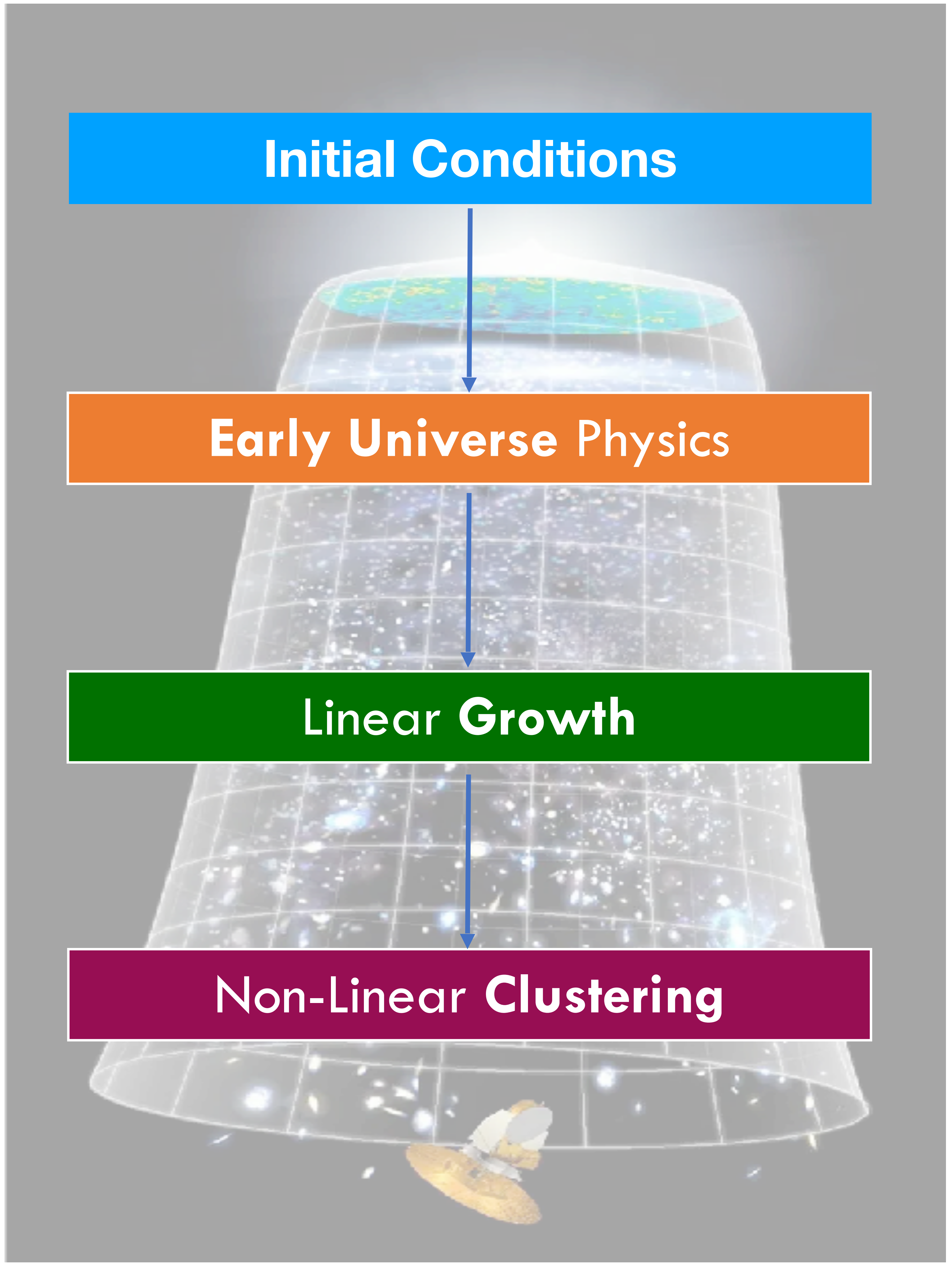
Major Goal: map the distribution of galaxies, $\delta_g(\mathbf{x}, z)$ across space and time

- Surveys range from low-**redshift** ($z \lesssim 0.1$) to high-redshift ($z \lesssim 3$)
    - *Low-z* — magnitude limited
    - *High-z* — large volume
- The surveys range from **ultra-large** to **ultra-deep**
    - *Large* — GR and non-Gaussianity
    - *Deep* — Non-linearities and structure formation

Each blob is a 3D galaxy position!

How to Model A Galaxy Survey

- **Initial Conditions**
    - **Gaussian**, with $\zeta \sim \mathcal{N}(0, P_\zeta)$
    - (Almost) **Scale-invariant**, with $P_\zeta(k) \sim A_s k^{n_s - 4}$
    - **Adiabatic** — all fields have the same initial conditions!
- **Early Universe Physics**
    - Standard expansion history including **matter-radiation** equality $k_{\mathrm{eq}} \sim \Omega_m H_0$
    - Recombination physics, including **sound horizon** $r_d \sim \Omega_b H_0^2, \Omega_m H_0^2, \cdots$
- **Linear Growth**
    - Background — $H(z), D_A(z)$ set by $H_0, \Omega_m, \Omega_r, \Omega_\Lambda$
    - Perturbations — $D(z), f(z) \sim \Omega_m(z)^\gamma$ set by $\Omega_m, \Omega_r, \Omega_\Lambda$ + GR
- **Non-Linear Clustering**
    - Collapse of **dark matter** into bound structures
    - Formation of **galaxies** around dark matter halos ← *This is the hard bit!*



How to Model A Galaxy Survey — beyond ΛCDM

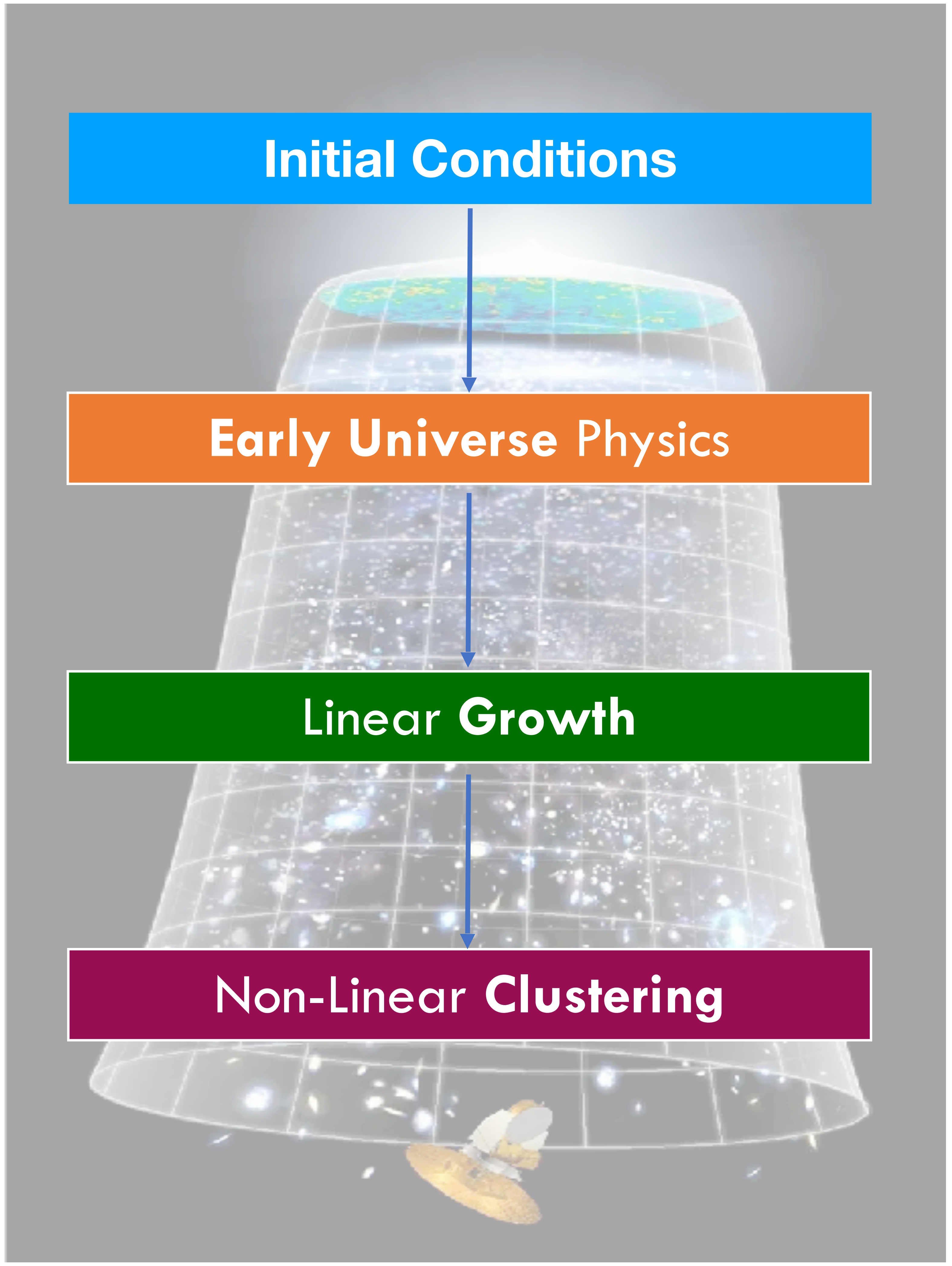
- **Initial Conditions**
  - **Gaussian**, with $\zeta \sim \mathcal{N}(0, P_\zeta)$
  - (Almost) **Scale-invariant**, with $P_\zeta(k) \sim A_s k^{n_s - 4}$
  - **Adiabatic** — all fields have the same initial conditions!
- **Early Universe Physics**
  - Standard expansion history including **matter-radiation** equality $k_{\rm eq} \sim \Omega_m H_0$
  - Recombination physics, including **sound horizon** $r_d \sim \Omega_b H_0^2, \Omega_m H_0^2, \cdots$
- **Linear Growth**
  - Background — $H(z), D_A(z)$ set by $H_0, \Omega_m, \Omega_r, \Omega_\Lambda$
  - Perturbations — $D(z), f(z) \sim \Omega_m(z)^\gamma$ set by $\Omega_m, \Omega_r, \Omega_\Lambda$ + GR
- **Non-Linear Clustering**
  - Collapse of **dark matter** into bound structures
  - Formation of **galaxies** around dark matter halos

Modified Inflation

Modified Early Universe

Modified Expansion History

Modified Structure Formation



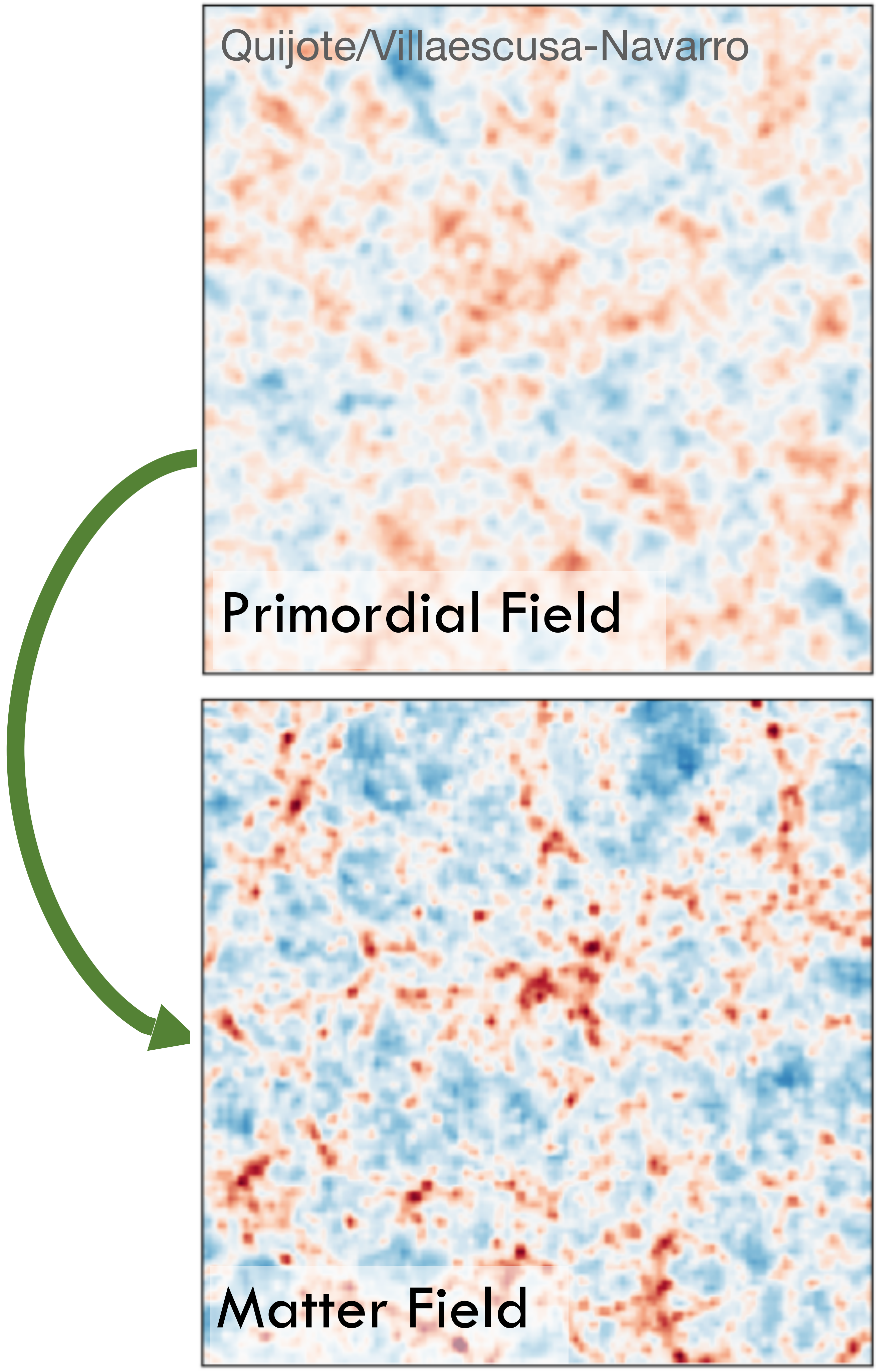
Modeling Non-Linearities

Step 1: predict the distribution of dark matter $\delta_m(\mathbf{x}, z)$

- This can be done either **analytically** or **numerically**
- Most modern analyses use the "Effective Field Theory of Large-Scale Structure"
  - This *smoothes* ("coarse-grains") the dark matter on some scale $R \gtrsim 10\,\mathrm{Mpc}$
    $\delta_m \to \delta_m \star \text{smoothing}[R]$
  - We can compute δ_m as a **perturbation series** in the initial conditions

$$\delta_m \sim K_1 \zeta + \int K_2 \zeta^2 + \int\int K_3 \zeta^3 + \cdots$$

  - The coupling kernels are set by **gravity**
  - Free **counterterms** account for **small-scale** physics ("renormalization")
- **Alternative** — run **N-body simulations** and **emulate** the statistics of interest



Cabass+, **Philcox**+, Chen+, d'Amico+, Assassi, Zaldarriaga, Simonovic, Senatore, White, Castorina, Vlah,…

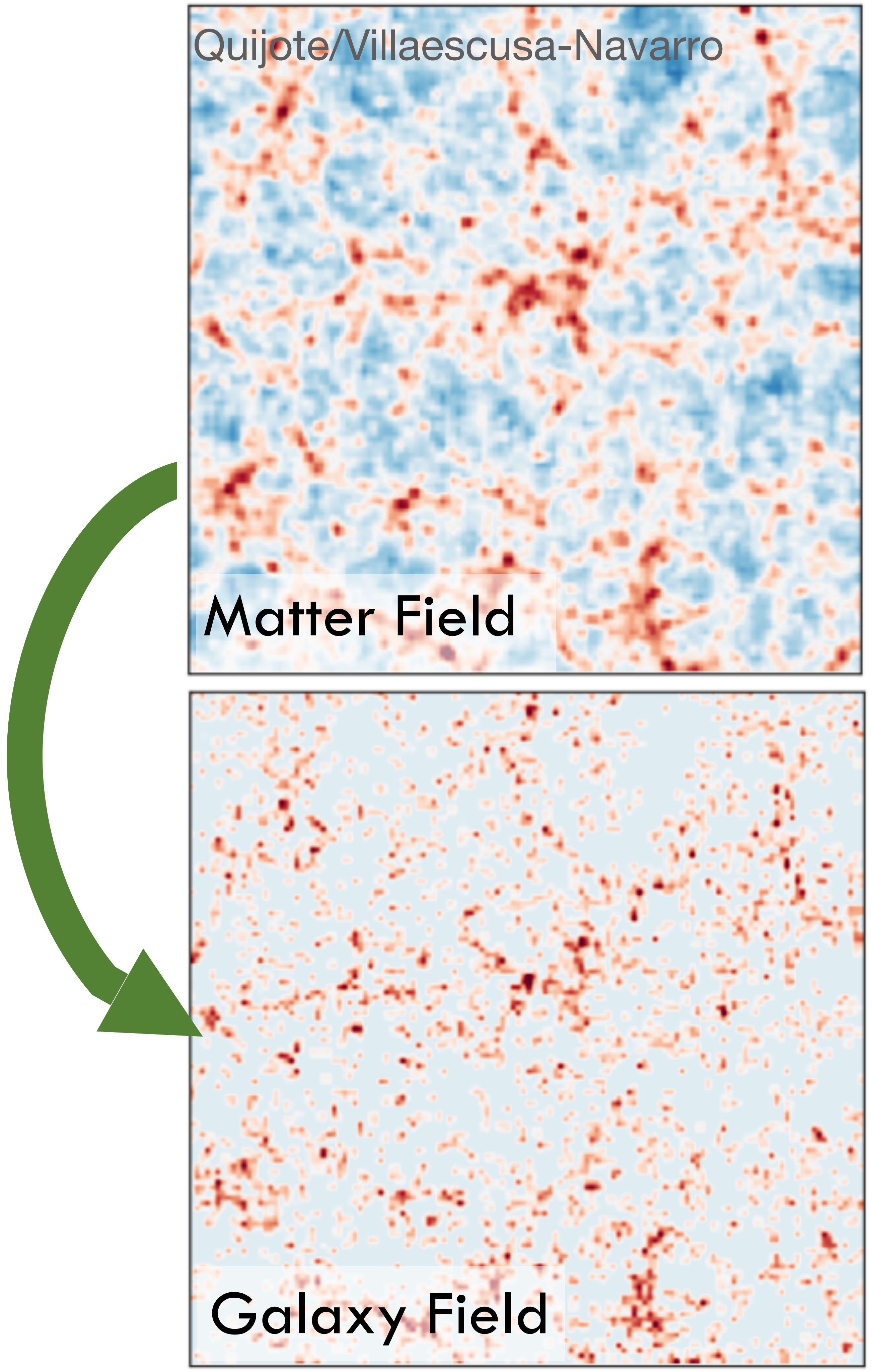
Modeling Non-Linearities

Step 2: predict the distribution of galaxies $\delta_g[\delta_m](\mathbf{x}, z)$

- This can be done **analytically**, **semi-analytically** or **numerically**
- **Effective Field Theory** approach:
  - Compute δ_g as a **perturbation series** in the dark matter density, δ_m
  - Use **symmetries** to account for **any galaxy formation** effects [e.g., homogeneity, isotropy, Galilean invariance]:

$$\delta_g = b_1\delta + b_2\delta^2 + b_s s_{ij}s^{ij} + b_\nabla \nabla^2\delta + \cdots$$

  - The **bias parameters** encode galaxy formation physics
  - This is **robust** but limited to large-scales ⇒ *loss of information*
  - Can model δ_m either from **theory** or **simulations** (Hybrid EFT)
- **Alternatives:**
  - Post-process **simulations** to add galaxies, with an HOD or Semi-Analytic Model
  - Perform a full hydrodynamical simulation!

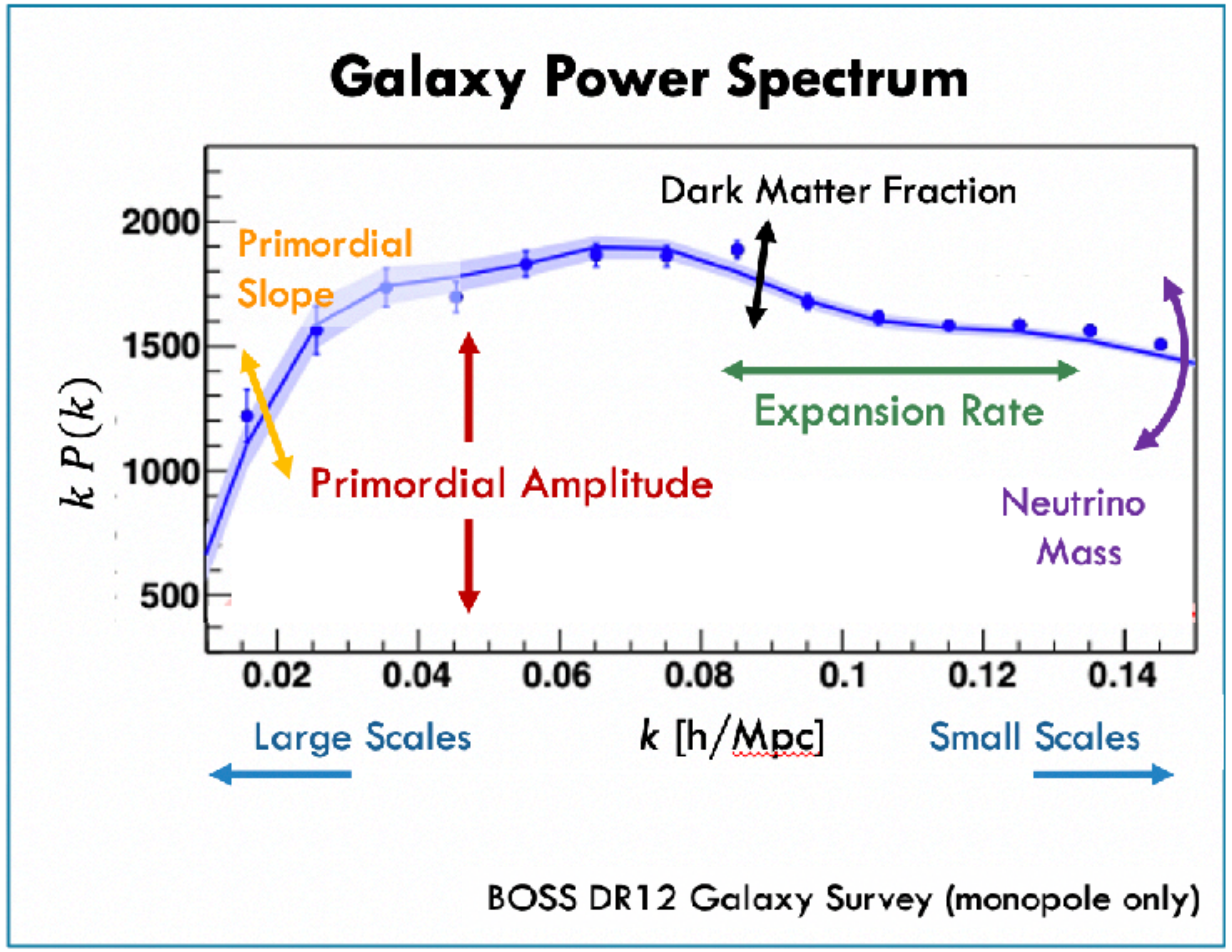


Cabass+, **Philcox**+, Chen+, d'Amico+, Assassi, Zaldarriaga, Simonovic, Senatore, White, Castorina, Vlah, Kokron, Modi, Angulo…

What Statistics Should We Use?

Most analyses focus on the simplest statistics

- The **B**aryon **A**coustic **O**scillations (BAO)
  - Good for tracing the **expansion history** across time
- The **Galaxy Power Spectrum** (or correlation function)
  - This is a **powerful** probe of ΛCDM!



(Many **many** references here!)

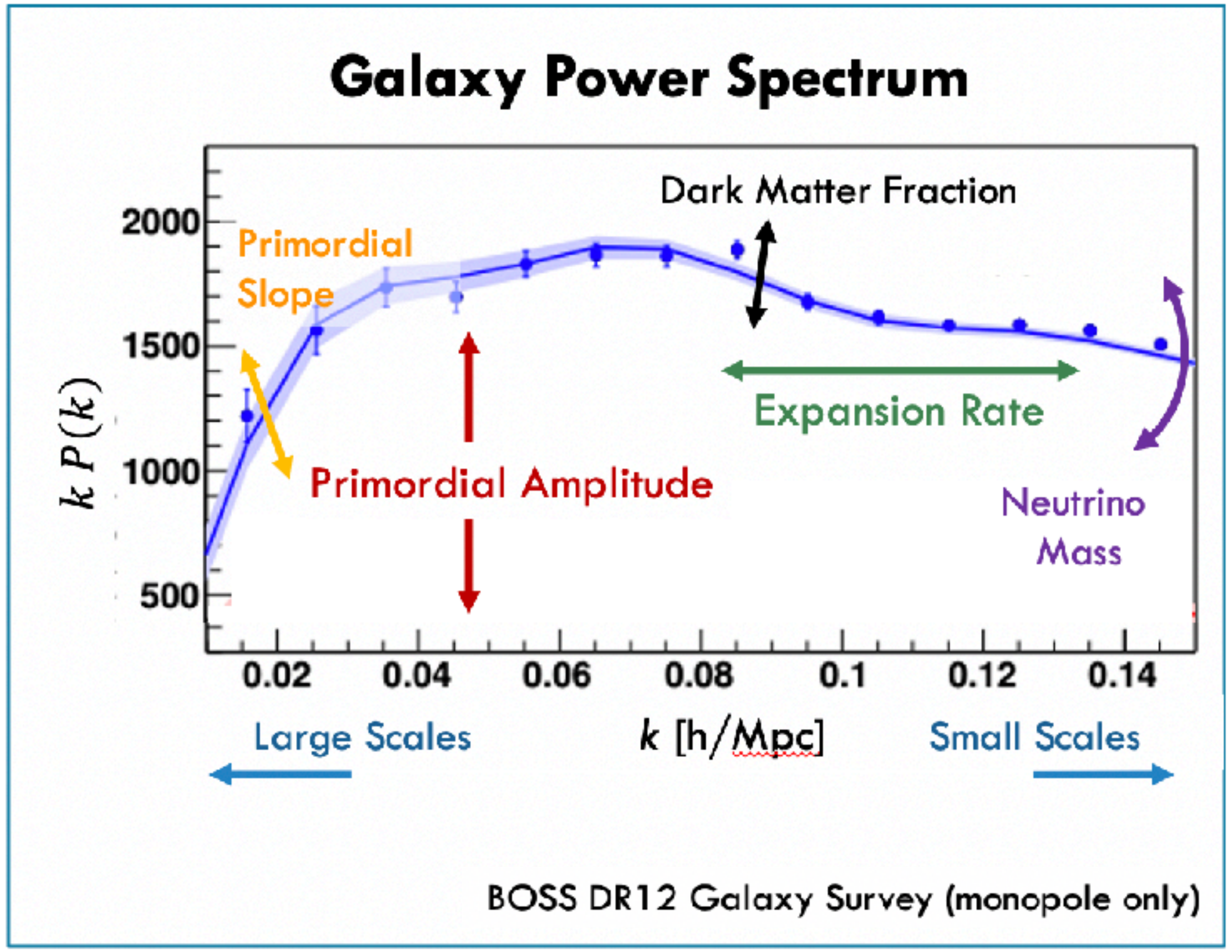
What Statistics Should We Use?

Most analyses focus on the simplest statistics

- The **B**aryon **A**coustic **O**scillations (BAO)
- The **Galaxy Power Spectrum** (or correlation function)

Other options include:

- ***Bispectra*** and ***Trispectra*** (3-Point and 4-Point Functions)
  - Traces non-linear information and **inflation** *but* more expensive!
- ***Galaxy bias*** and ***halo mass functions*** $n_{\rm gal}(M_{\rm halo})$
  - These probe halo-scale physics ($R \lesssim 10h^{-1}{\rm Mpc}$) but are difficult to model!
- ***Wavelets, kNNs, CNNs, marked statistics, density-split statistics, …***
  - These are **non-linear** probes that must be modeled numerically!
- Cross-Correlations with **weak lensing**
  - Strong probes of **gravity**
- Galaxy **spins**, galaxy **shapes**
  - Higher-order physics including **tensors**



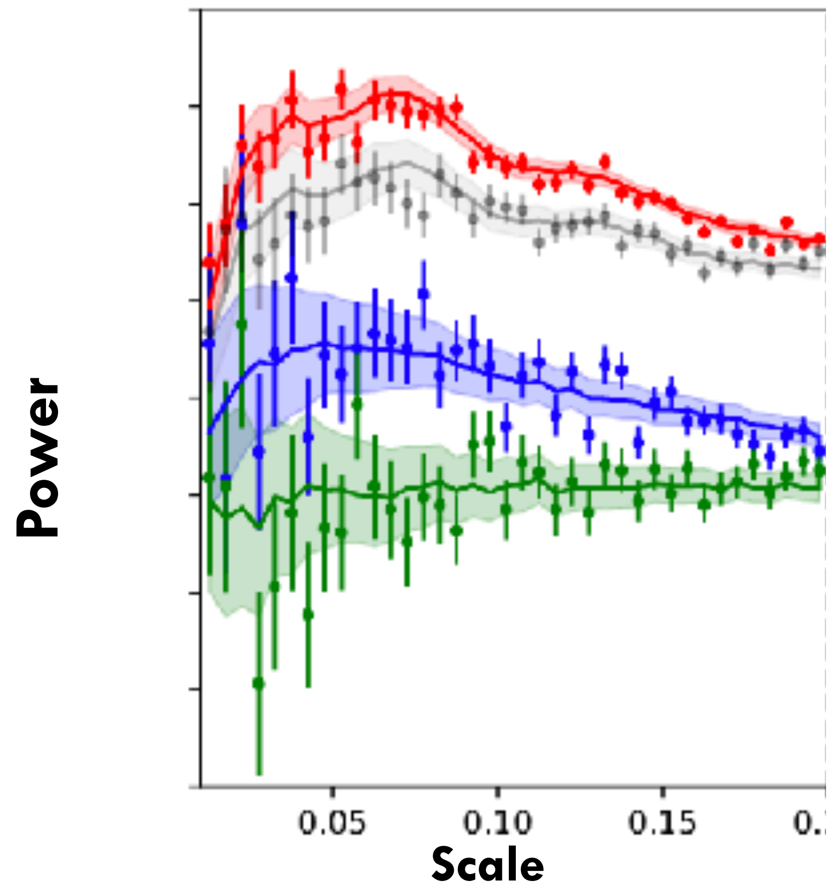
(Many **many** references here!)

Theory In Practice

- There are **fast codes** implementing the theory predictions:
  - CLASS-PT, Velocileptors, PyBird, PBJ, Class OneLoop, FOLPS-ν,…
- These predict the **power spectrum** or **bispectrum** of galaxies
- By combining with an **observed dataset** and a **Gaussian likelihood** we can constrain any ΛCDM parameter entering the model!

Cosmology Parameters

Perturbation Theory Code



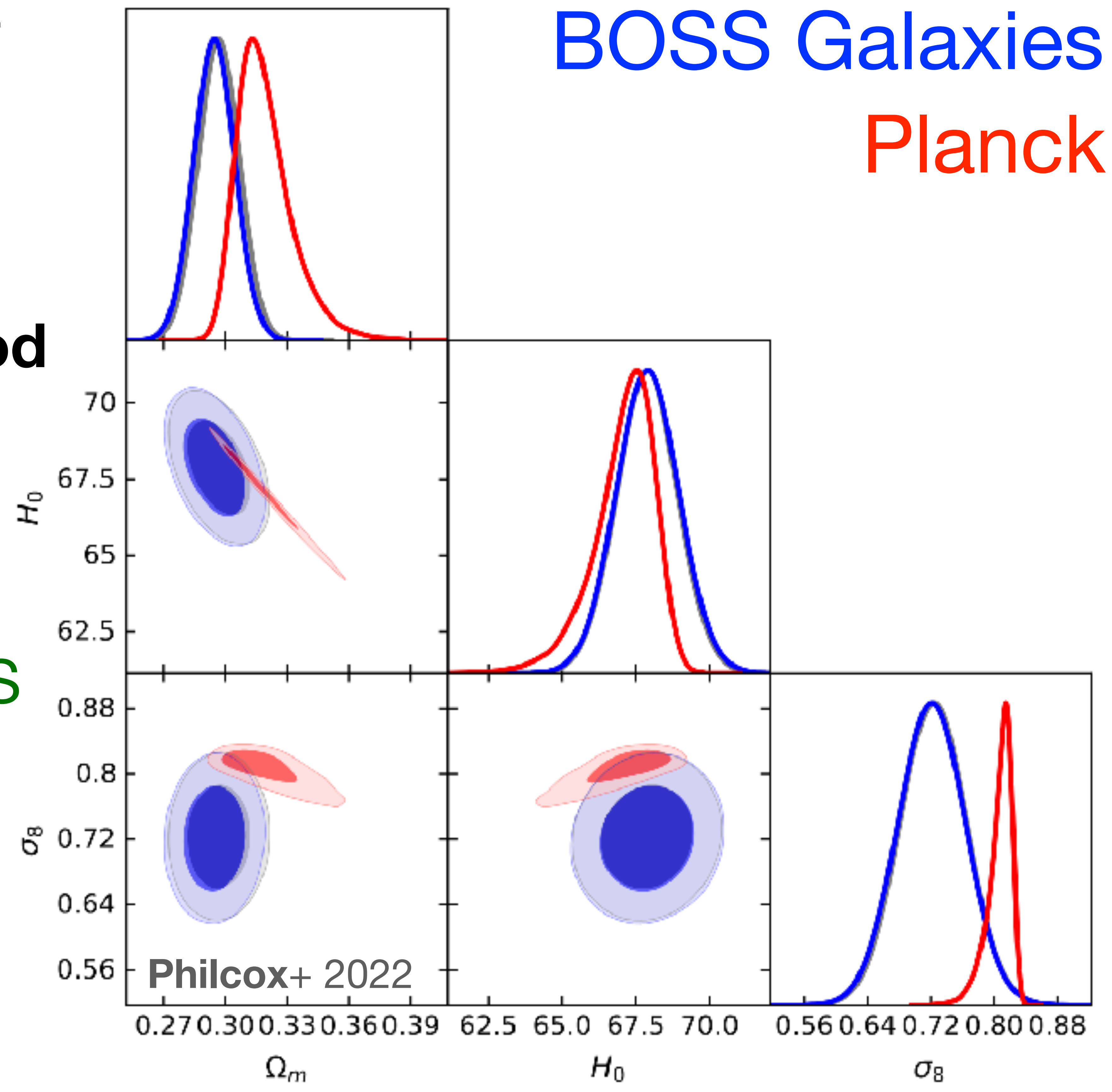
Predictions for Statistics

Ivanov, **Philcox**, Cabass, Chen, Vlah, Zhang, d'Amico, Simonovic, Zaldarriaga, Senatore, Moretti, Lesgourges, Werth, Aviles, Gil-Marin, Beutler, …

Theory In Practice

- There are **fast codes** implementing the theory predictions:
  - CLASS-PT, Velocileptors, PyBird, PBJ, Class OneLoop, FOLPS-ν,…
- These predict the **power spectrum** or **bispectrum** of galaxies
- By combining with an **observed dataset** and a **Gaussian likelihood** we can constrain any ΛCDM parameter entering the model!
- This has been used to measure $\Omega_m, H_0, \sigma_8, \cdots$ from BOSS / eBOSS in **full-shape / direct modeling** analyses
  - *(See also ShapeFit and $f\sigma_8$ measurements)*
- Recently, it has been applied to the **Year 1 DESI dataset**

Can we reproduce this?

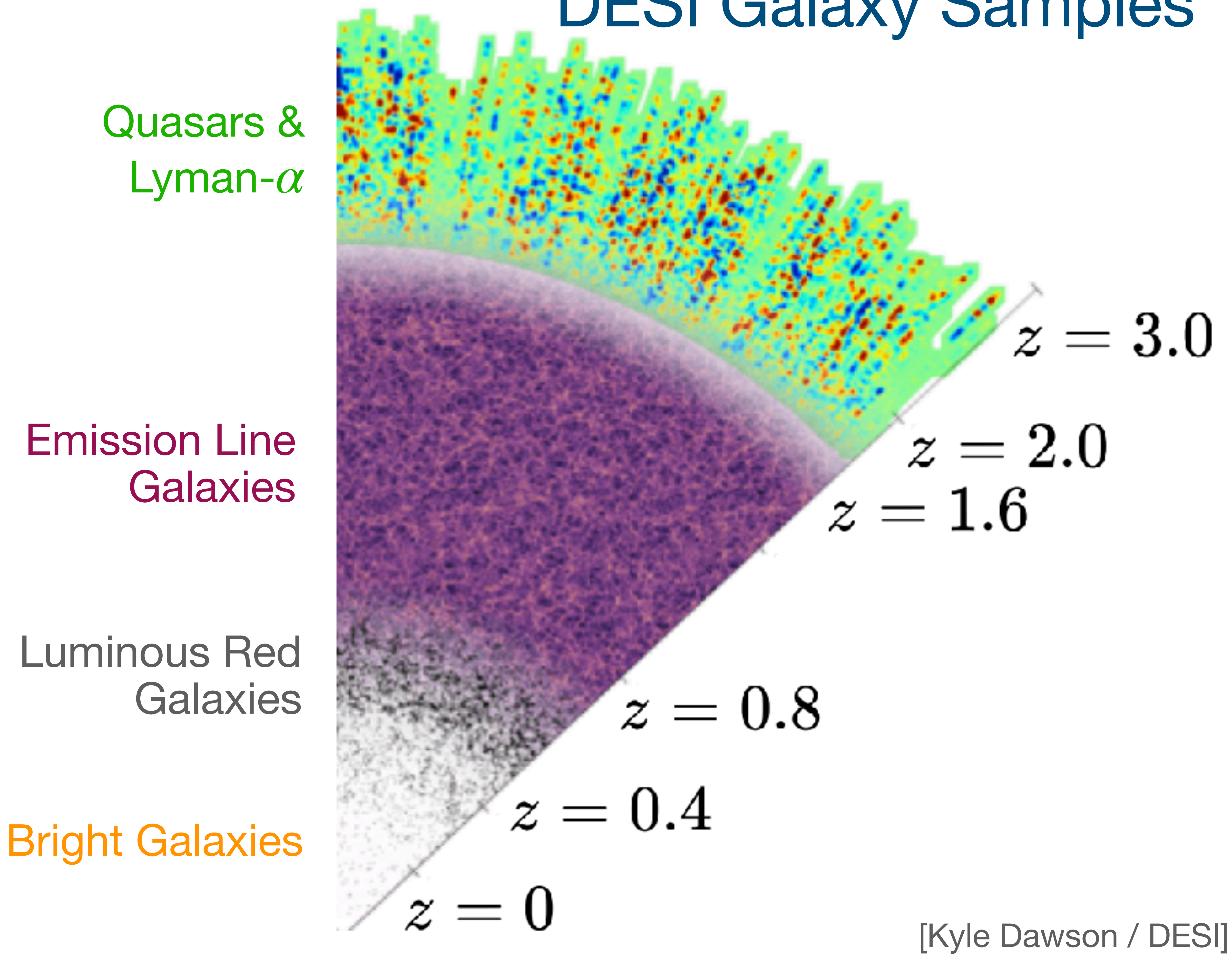


Ivanov, **Philcox**, Cabass, Chen, Vlah, Zhang, d'Amico, Simonovic, Zaldarriaga, Senatore, Moretti, Lesgourges, Werth, Aviles, Gil-Marin, Beutler, …

The DESI Universe

- We analyze **six** DESI chunks:
  - BGS: Bright Galaxy Sample ($0.1 < z < 0.4$)
    - Low redshift, magnitude limited
  - LRG: Luminous Red Galaxies ($0.4 < z < 1.1$)
    - Similar to previous surveys!
  - ELG: Emission Line Galaxies ($1.1 < z < 1.6$)
    - Low-redshift tail dropped due to systematic contamination
  - QSO: Quasars ($0.8 < z < 2.1$)
    - High redshift, **large** shot-noise
  - Lyman-alpha Emission
    - Not included in DR1
- Each is split into **north** and **south** regions

DESI Galaxy Samples

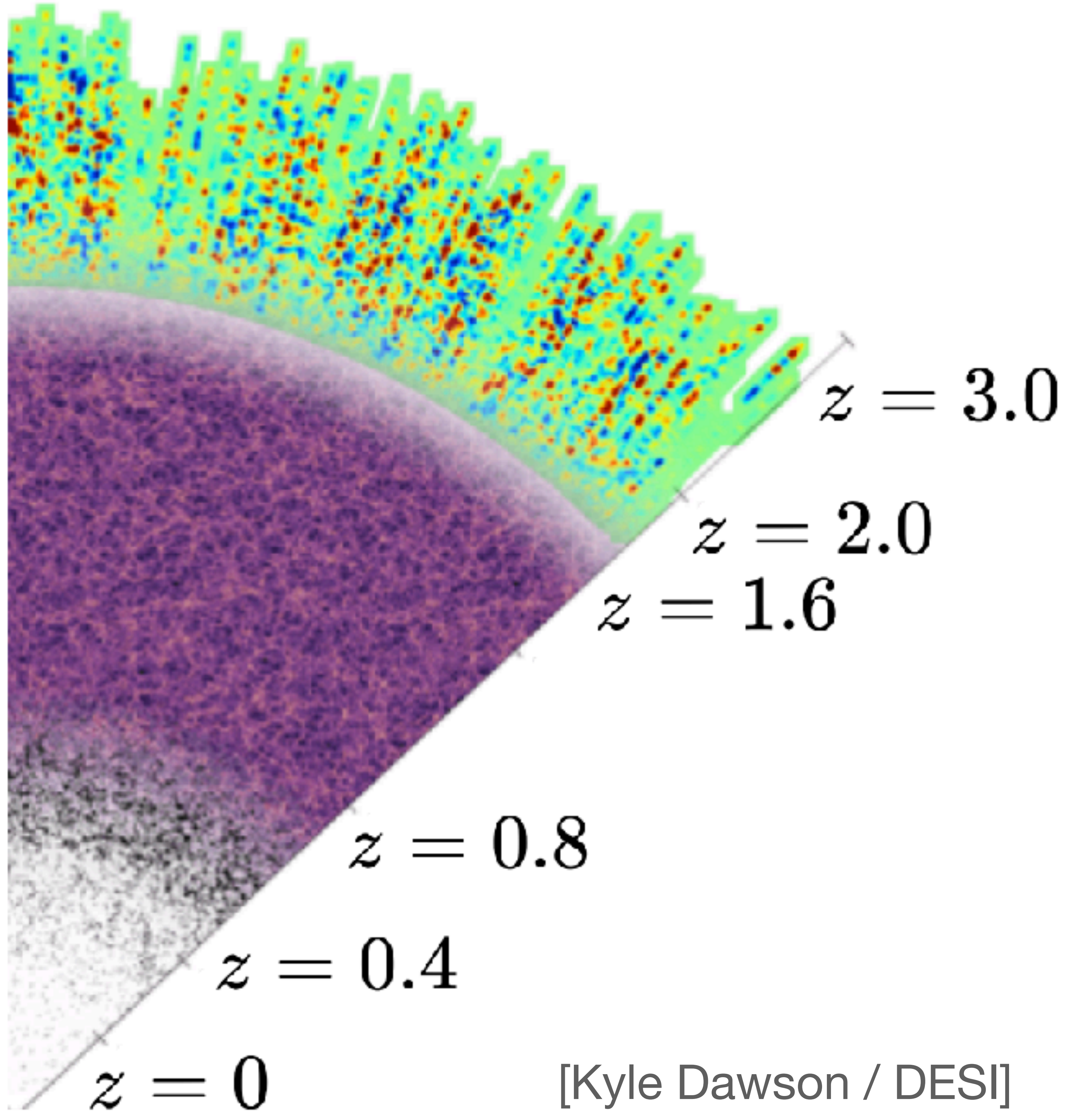


[Kyle Dawson / DESI]

DESI 2024 II, V, DESI Data Release 1

Reanalyzing DESI

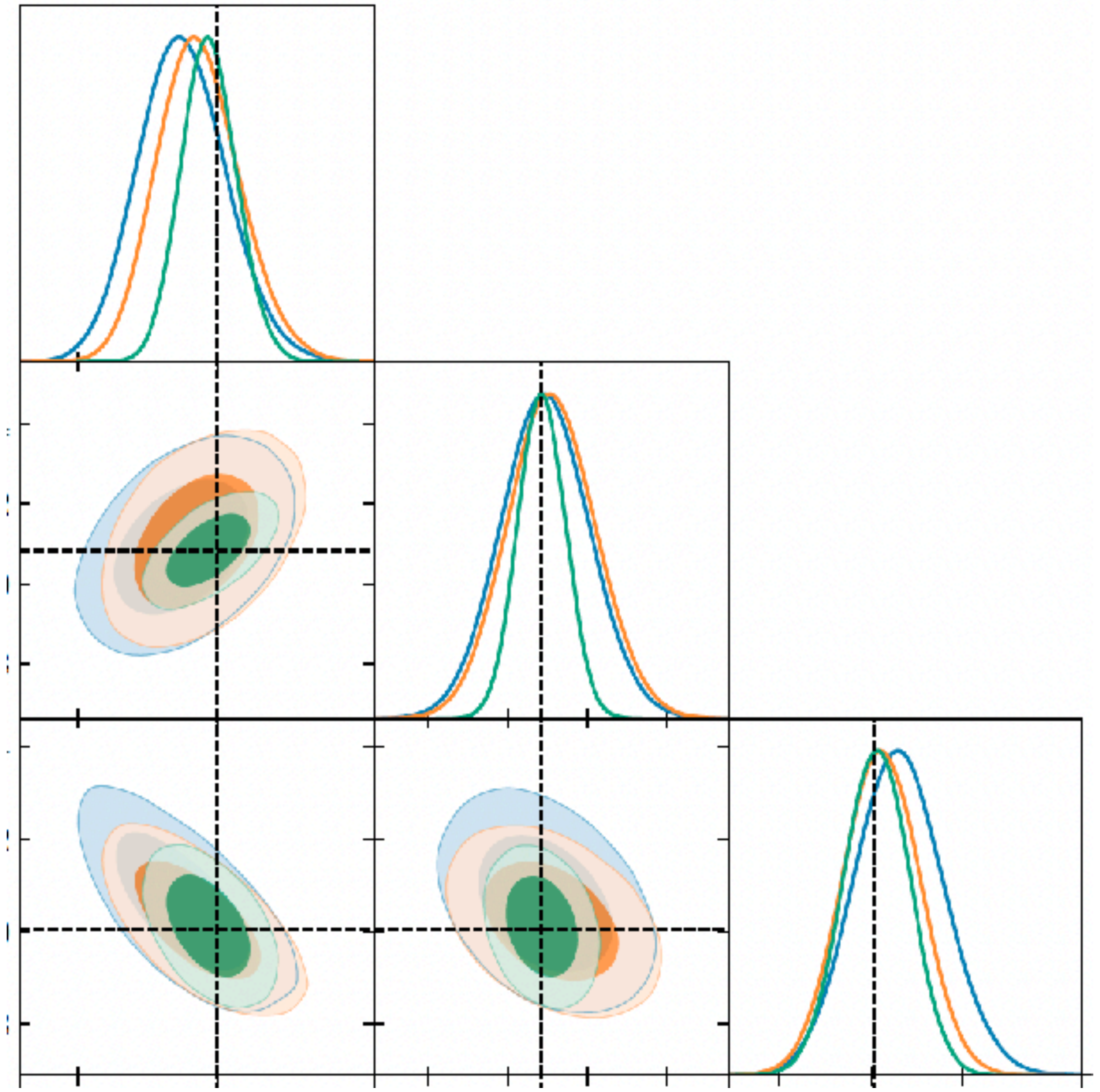
DESI Galaxies



[Kyle Dawson / DESI]

???

Cosmological Parameters



DESI 2024 II, V, DESI Data Release 1, Chudaykin, Ivanov, **Philcox** 2025

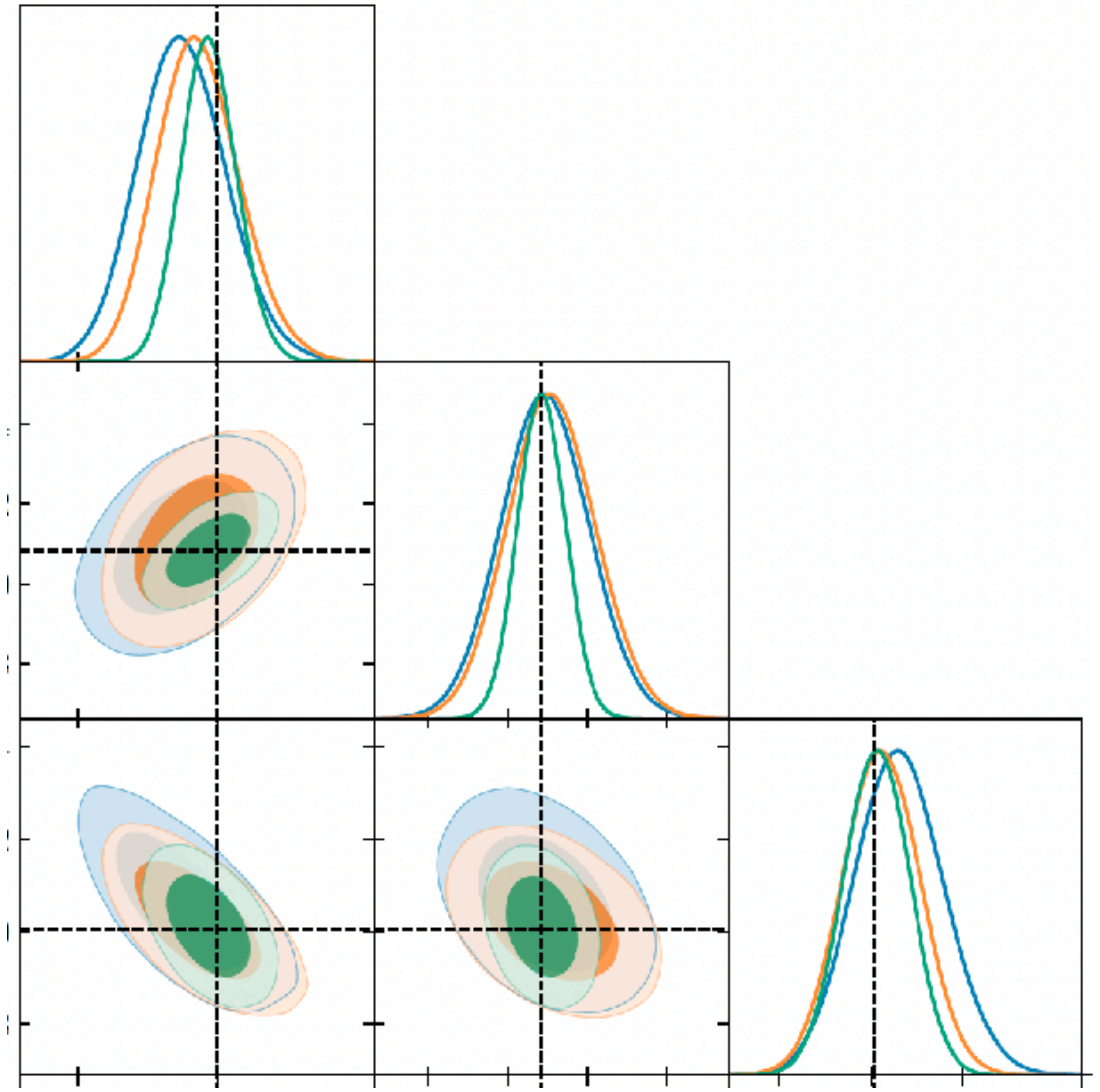
Reanalyzing DESI

DESI Data Release 1 (LRGs)

TARGETID	Z	NTILE	RA	DEC	...
int64	float64	int64	float64	float64	...
39627540901396844	0.42060841162467566	1	159.30684159361635	-10.155757636765902	...
39627546836338876	0.8668980715716706	1	158.44667596279407	-9.962760066342906	...
39627546840531340	0.9348172077800124	1	158.4799294702238	-9.880343166939232	...
39627546840533707	0.7646678553759423	1	158.65071160360105	-9.900898173028425	...
39627546840534067	0.88129590000311	1	158.67878216902403	-9.91791308567385	...
39627546840534396	0.6646155566176719	1	158.70027052890555	-9.885818986284596	...
39627546844725593	0.7619120932610688	1	158.72751630870823	-10.011383569041937	...
39627546844726132	0.8129116729090922	1	158.76343950179967	-9.912671320450734	...
39627546844726593	0.835471640017949	1	158.79898500886574	-9.952788127324665	...
39627546848921194	0.8148312339778753	1	159.052157885943	-9.992428612452807	...
39627546848922139	0.7200341373651288	1	159.10202657806508	-9.938566366253678	...
39627546848922621	0.7606337242857438	1	159.1309146297404	-10.02377942401391	...
39627546848922874	0.7198972751282844	1	159.1462785833043	-9.950181865635432	...
39627546848923188	0.7210857282186207	1	159.16399100631358	-9.912947332242044	...
39627546848923381	0.569430729151765	1	159.17802210549974	-9.97892860399317	...
39627546848923415	0.8891288789150124	1	159.18008439182032	-10.072752528866118	...
39627546848923493	0.9513285375888253	1	159.1840389390485	-9.910321824120278	...
39627546848923519	0.7212784017696859	1	159.1860701777553	-9.944737378735352	...
39627546853114634	0.8131126675553368	1	159.25137421856687	-10.058275905081851	...
39627546853115304	0.5559672054059013	1	159.28855963426028	-9.955979493106813	...
39627546853115470	0.7147216867384578	1	159.2970230990033	-10.012836906791499	...
39627546853115682	0.9274570688680336	1	159.30835543527493	-10.106935803496164	...
...	...	...	...	...	...

???

Cosmological Parameters



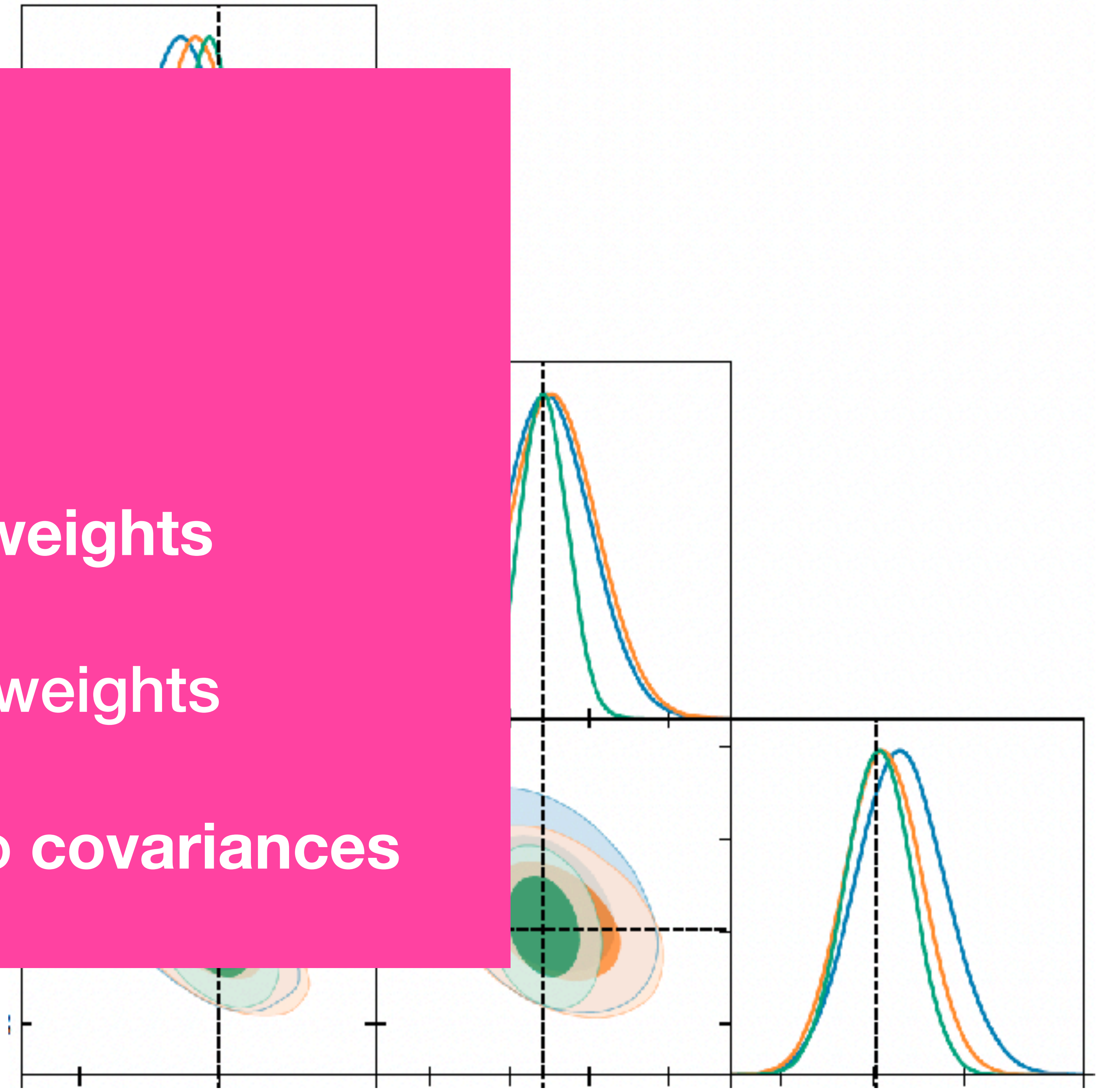
DESI 2024 II, V, DESI Data Release 1, Chudaykin, Ivanov, **Philcox** 2025

Reanalyzing DESI

DESI Data Release 1 (LRGs)

TARGETID int64	Z float64	NTIL int64
39627540901396844	0.42060841162467566	
39627546836338876	0.8668980715716706	
39627546840531340	0.9348172077800124	
39627546840533707	0.7646678553759423	
39627546840534067	0.88129590000311	
39627546840534396	0.6646155566176719	
39627546844725593	0.7619120932610688	
39627546844726132	0.8129116729090922	
39627546844726593	0.835471640017949	
39627546848921194	0.8148312339778753	
39627546848922139	0.7200341373651288	
39627546848922621	0.7606337242857438	
39627546848922874	0.7198972751282844	
39627546848923188	0.7210857282186207	
39627546848923381	0.569430729151765	
39627546848923415	0.8891288789150124	
39627546848923493	0.9513285375888253	
39627546848923519	0.7212784017696859	
39627546853114634	0.8131126675553368	
39627546853115304	0.5559672054059013	
39627546853115470	0.7147216867384578	
39627546853115682	0.9274570688680336	
...	...	...

Cosmological Parameters



This is hard

The data release only contains:

- **Galaxy positions, redshifts and systematic weights**
- Random positions, redshifts and systematic weights

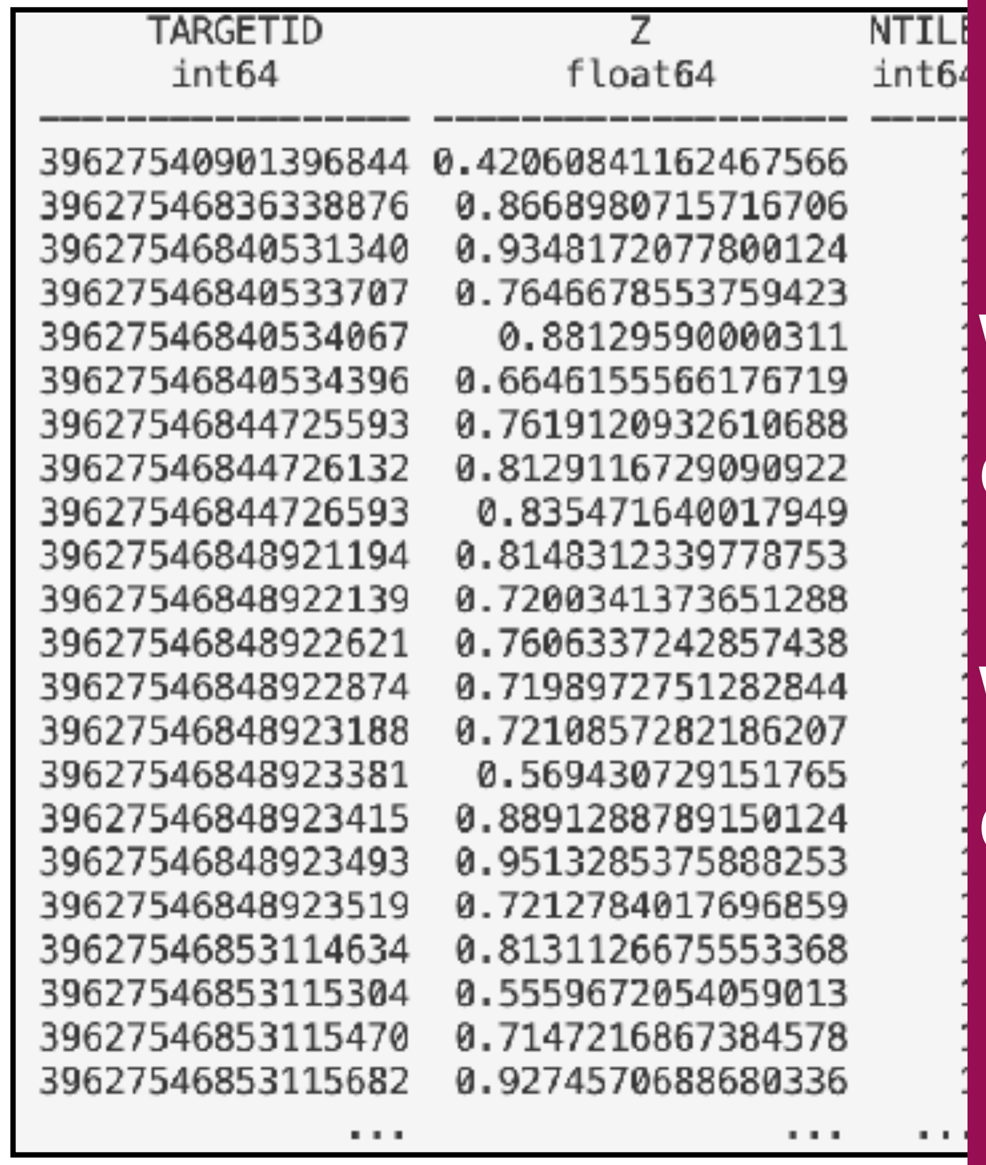
There are no **simulations**, no **power spectra** and no **covariances**

DESI 2024 II, V, DESI Data Release 1, Chudaykin, Ivanov, **Philcox** 2025

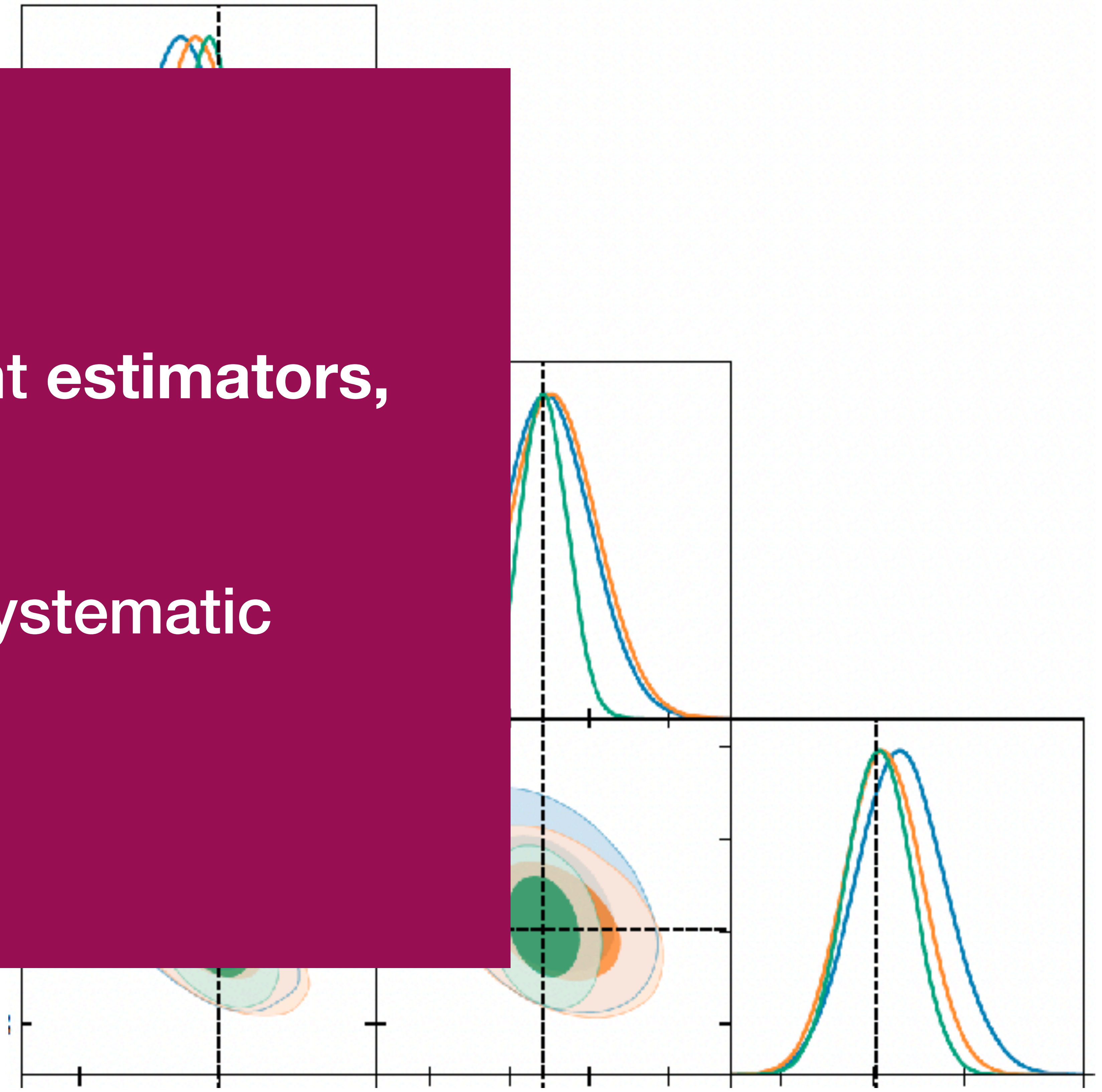
Reanalyzing DESI

DESI Data Release 1 (LRGs)

Cosmological Parameters



TARGETID int64	Z float64	NTIL int6
39627540901396844	0.42060841162467566	
39627546836338876	0.8668980715716706	
39627546840531340	0.9348172077800124	
39627546840533707	0.7646678553759423	
39627546840534067	0.88129590000311	
39627546840534396	0.6646155566176719	
39627546844725593	0.7619120932610688	
39627546844726132	0.8129116729090922	
39627546844726593	0.835471640017949	
39627546848921194	0.8148312339778753	
39627546848922139	0.7200341373651288	
39627546848922621	0.7606337242857438	
39627546848922874	0.7198972751282844	
39627546848923188	0.7210857282186207	
39627546848923381	0.569430729151765	
39627546848923415	0.8891288789150124	
39627546848923493	0.9513285375888253	
39627546848923519	0.7212784017696859	
39627546853114634	0.8131126675553368	
39627546853115304	0.5559672054059013	
39627546853115470	0.7147216867384578	
39627546853115682	0.9274570688680336	
...	...	...



This is important

We develop an **independent** pipeline, using different **estimators**, **covariance estimates**, and **theory codes**

We can include **more data with new methods for systematic** corrections

DESI 2024 II, V, DESI Data Release 1, Chudaykin, Ivanov, **Philcox** 2025

Two-Point Estimators

- The DESI data is a **point cloud** of positions and weights for **galaxies** and **randoms**

$$n_g(\mathbf{x}) \sim \sum_{i=1}^{N_g} w_{g,i}\delta_{\rm D}(\mathbf{x}-\mathbf{x}_{g,i}), \quad n_r(\mathbf{x}) \sim \sum_{i=1}^{N_r} w_{r,i}\delta_{\rm D}(\mathbf{x}-\mathbf{x}_{r,i})$$

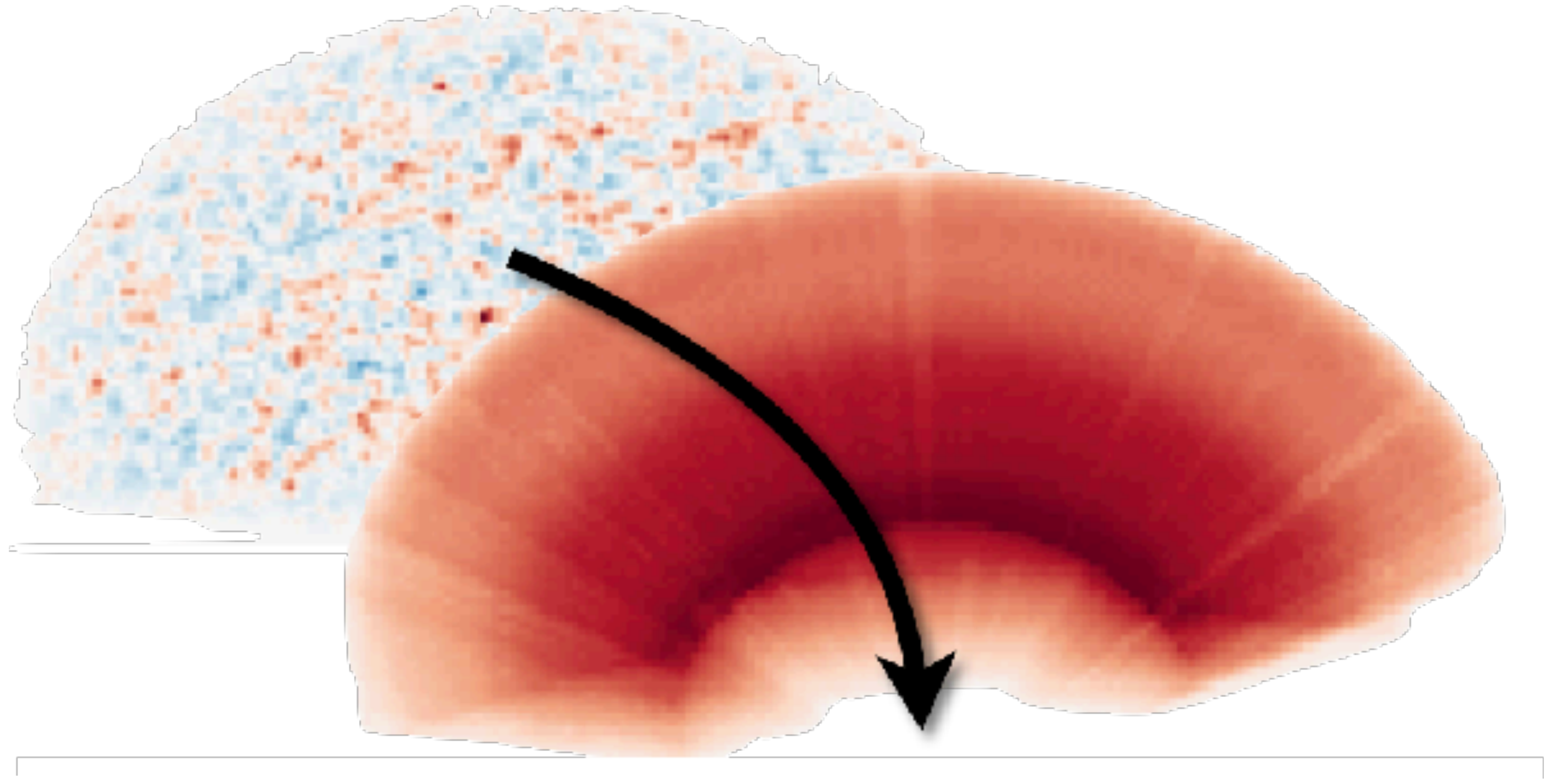
- We want to turn this into a **power spectrum**

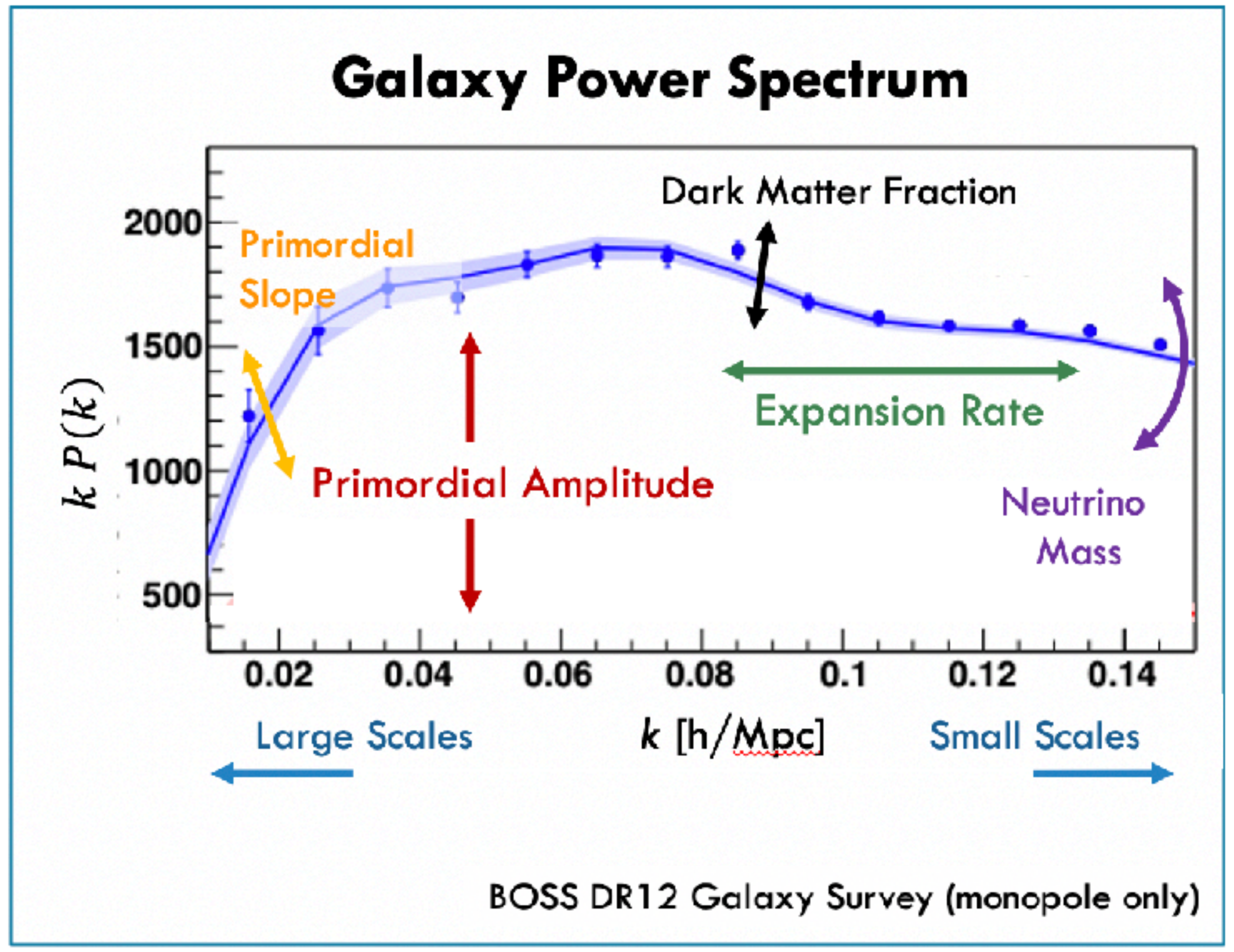
1. **FKP** estimator ← Used by DESI

$$P_{\rm FKP}(\mathbf{k}) \sim \left| n_g(\mathbf{k}) - \frac{N_g}{N_r} n_r(\mathbf{k}) \right|^2 / \langle n^2 \rangle$$

- This is (almost) optimal
- The output is convolved with the mask:

$$P_{\rm FKP}(\mathbf{k}) \sim \int d\mathbf{q}\ \left| n(\mathbf{k}-\mathbf{q}) \right|^2 P_{\rm true}(\mathbf{q}) / \langle n^2 \rangle$$



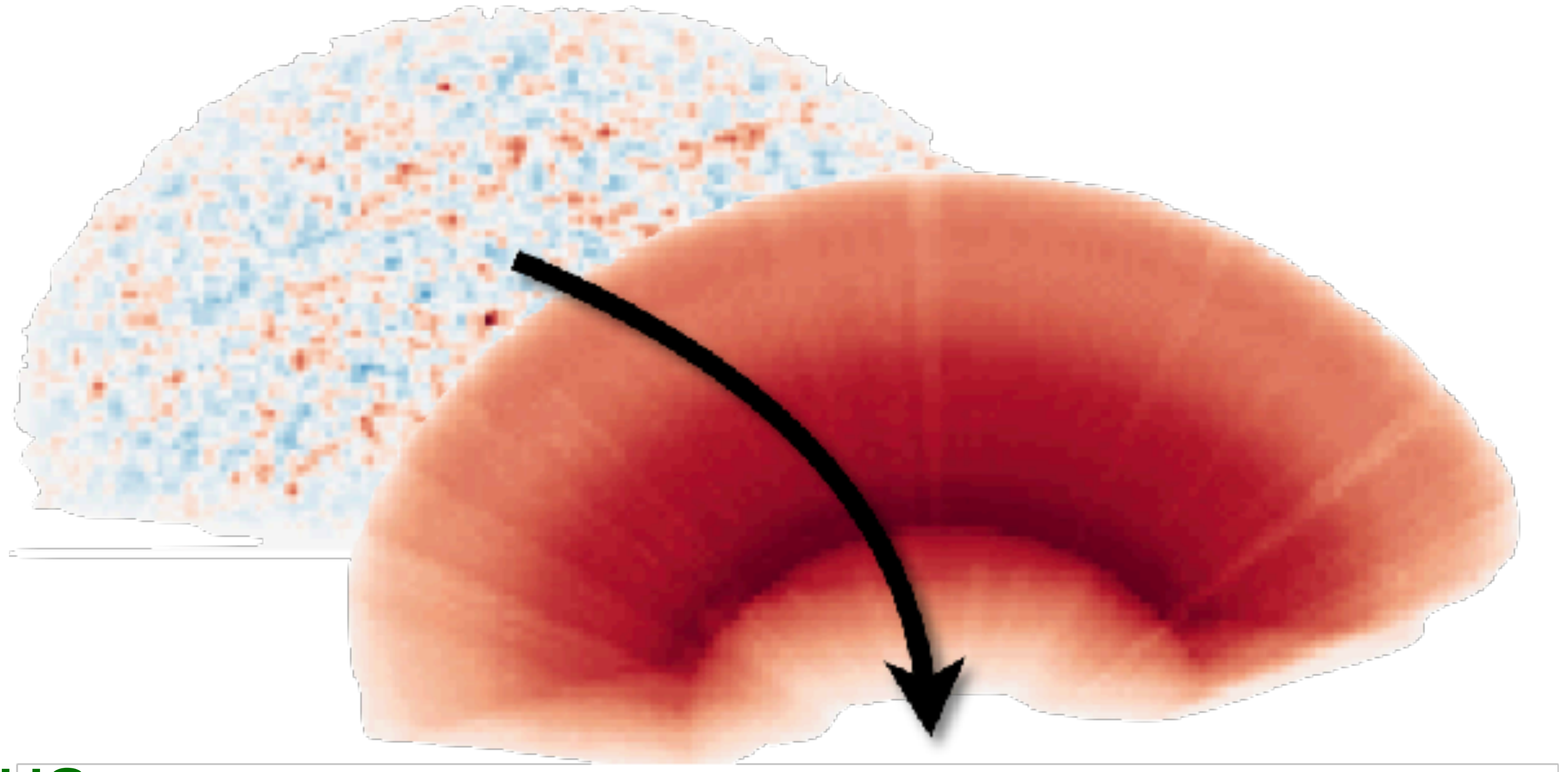


Feldman, Kaiser, Peacock 1980s, BOSS DR12, Beutler+, Gil-Marin+, McDonald+, Ivanov+19, **Philcox**+20,21, **Philcox** & Floss 2025

Two-Point Estimators

- The DESI data is a **point cloud** of positions and weights for **galaxies** and **randoms**

$$n_g(\mathbf{x}) \sim \sum_{i=1}^{N_g} w_{g,i}\delta_{\mathrm{D}}(\mathbf{x}-\mathbf{x}_{g,i}), \quad n_r(\mathbf{x}) \sim \sum_{i=1}^{N_r} w_{r,i}\delta_{\mathrm{D}}(\mathbf{x}-\mathbf{x}_{r,i})$$



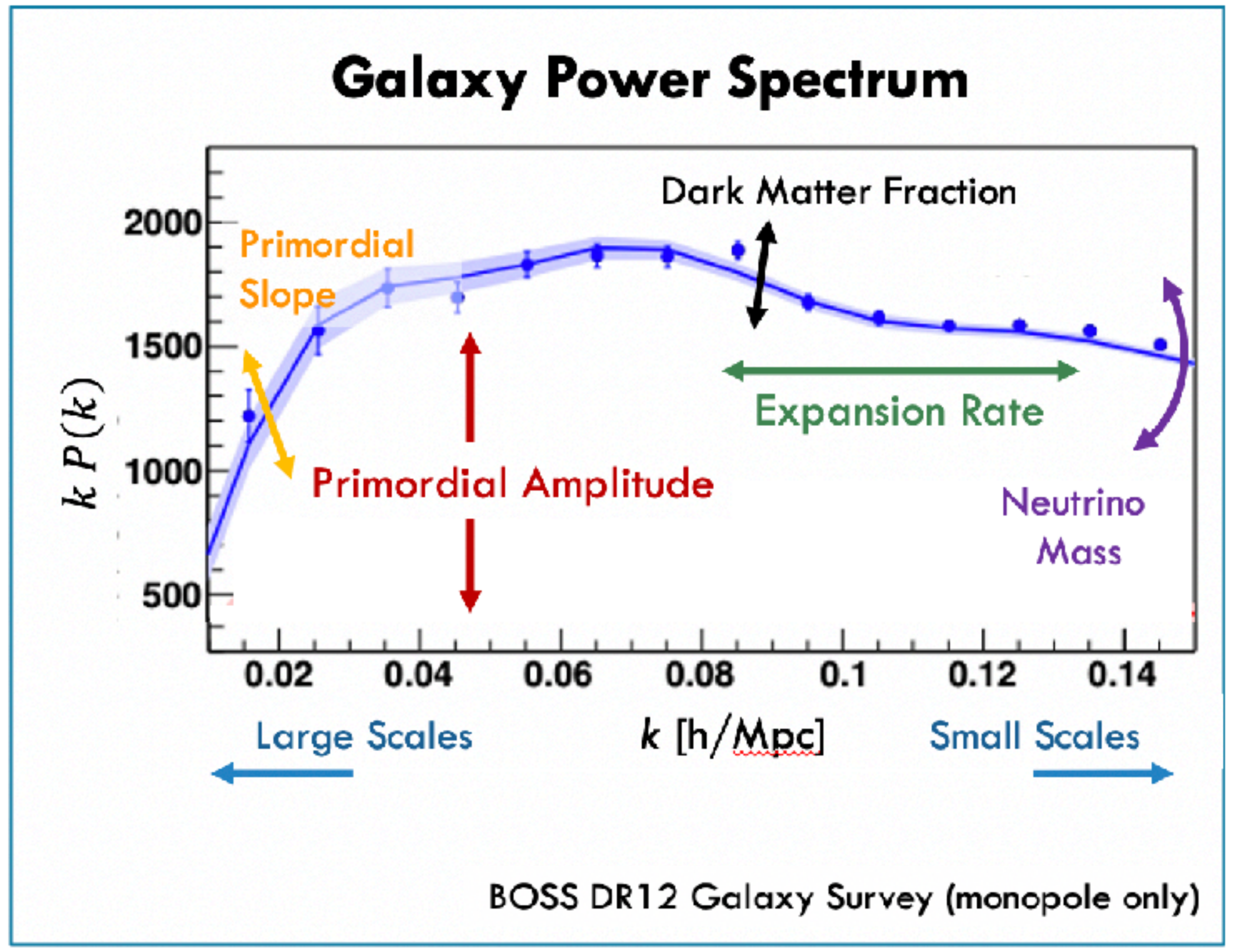
- We want to turn this into a **power spectrum**

2. Quasi-Optimal “**Unwindowed**” estimator ⟵ Used by us

$$P_{\mathrm{unwin}}(\mathbf{k}) \sim \int d\mathbf{q}\, \mathbf{W}^{-1}(\mathbf{k},\mathbf{q}) \left| n_g(\mathbf{q}) - \frac{N_g}{N_r} n_r(\mathbf{q}) \right|^2 / \langle n^2 \rangle$$

- At leading order, the output is **not** convolved with the mask:

$$P_{\mathrm{unwin}}(\mathbf{k}) \sim P_{\mathrm{true}}(\mathbf{k})$$



Feldman, Kaiser, Peacock 1980s, BOSS DR12, Beutler+, Gil-Marin+, Ivanov+19, **Philcox**+20,21, **Philcox** & Floss 2025

Two-Point Estimators

- In practice, we compute the **binned power spectrum** in a set of k-bins and redshift-space **multipoles**, plus a (square) **2D normalization matrix**

$$P_\ell^{\mathrm{unwin}}(k) \sim \sum_{\ell' k'} \mathscr{W}^{-1}_{\ell\ell'}(k,k') \left| n_g(\mathbf{k}') - \frac{N_g}{N_r} n_r(\mathbf{k}') \right|^2_{\ell'}$$

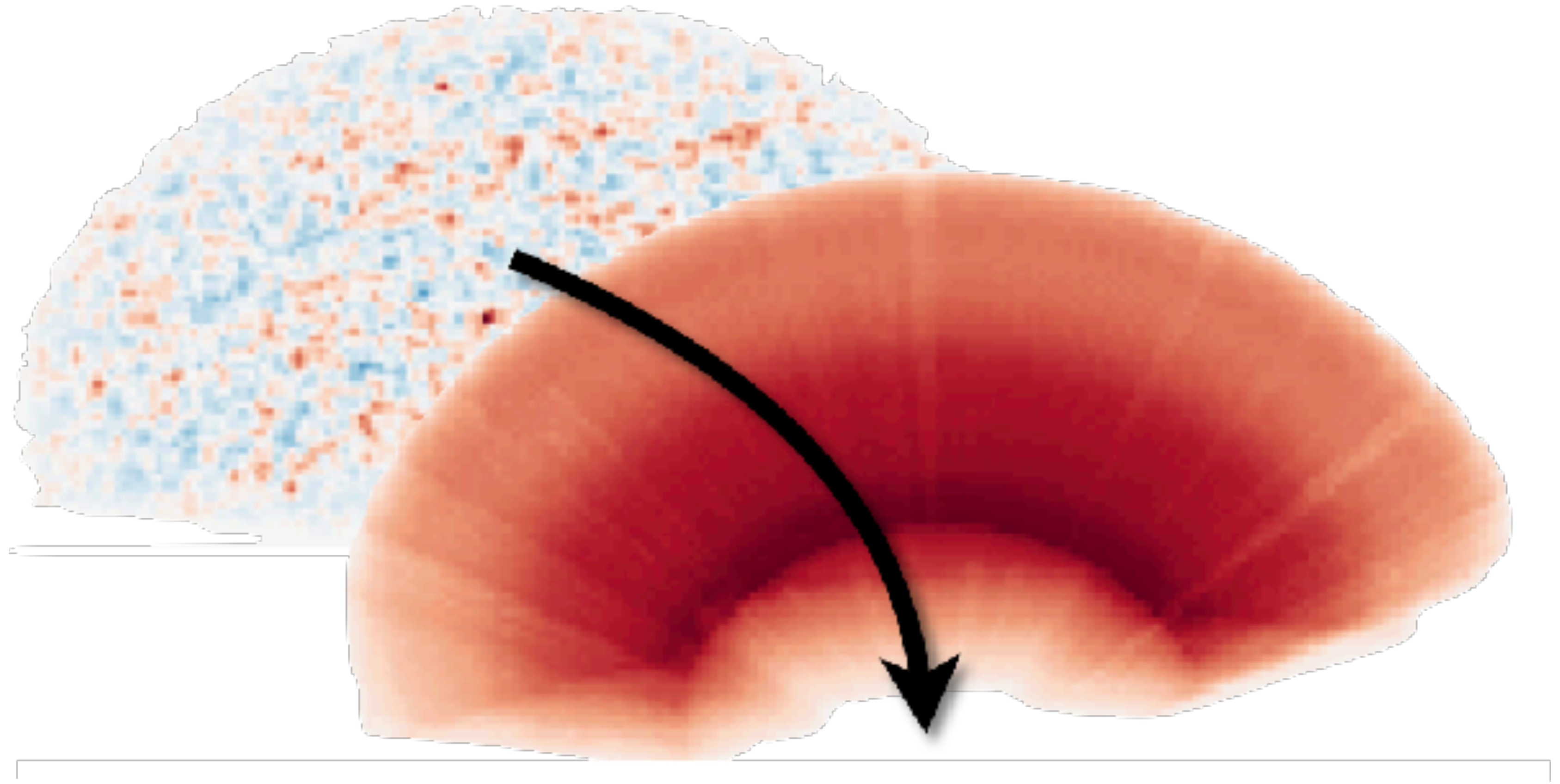
- The **numerator** is the standard FKP numerator (up to an user-defined weight)

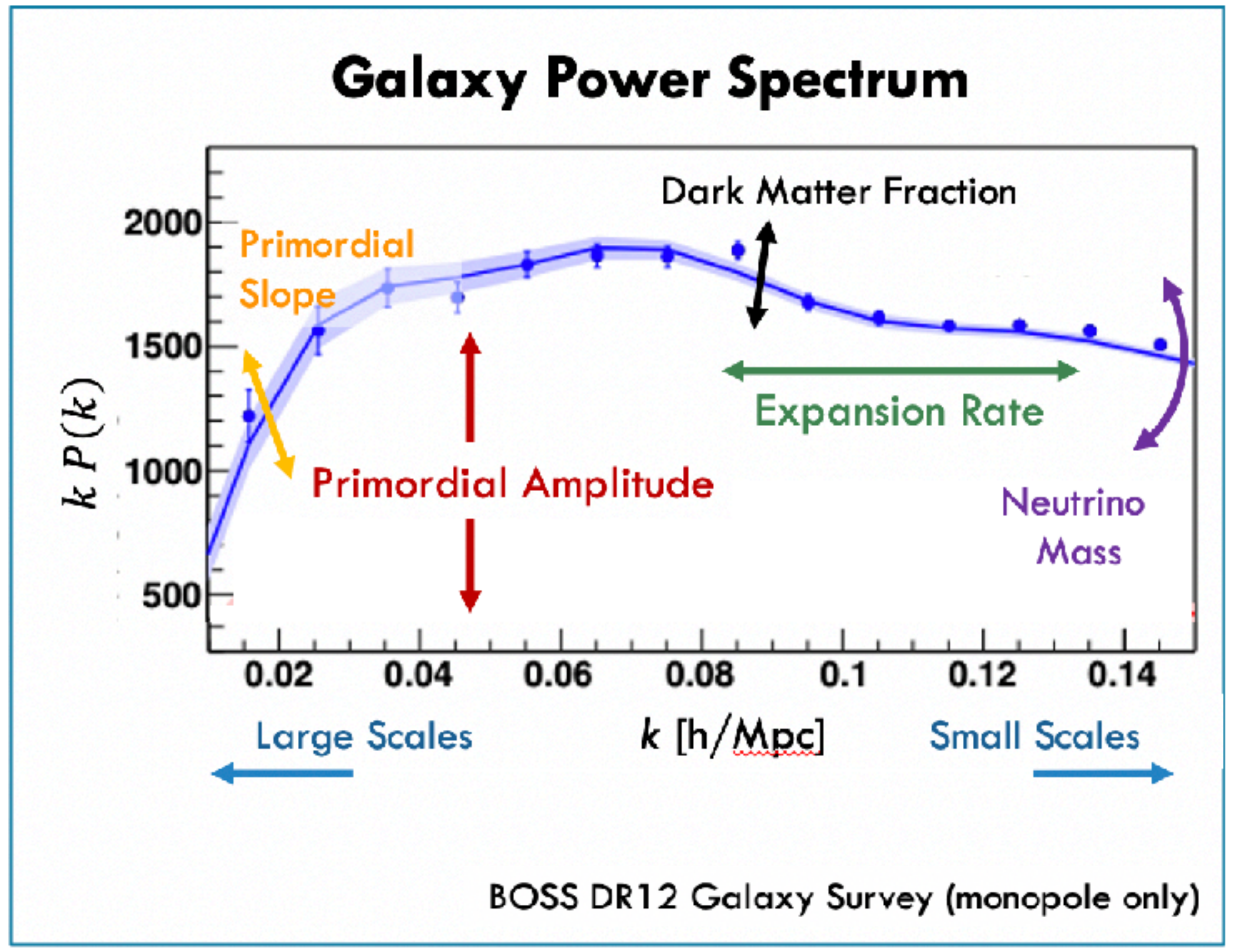
- The **normalization** can be computed using **Monte Carlo** methods and **FFTs**
(Stochastic Trace Estimation; See **Philcox**, Floss 2025)

- We account for **residual corrections** using a rectangular ***theory matrix*** (computed stochastically)

- This is equivalent to the **pseudo-C_ℓ scheme** used in the CMB!

- Up to scale-cuts, it is **equivalent** to the standard estimators





Feldman, Kaiser, Peacock 1980s, BOSS DR12, Beutler+, Gil-Marin+, McDonald+, Ivanov+19, **Philcox**+20,21, **Philcox** & Floss 2025

Three-Point Estimators

- We also want to compute **bispectra**

- These are usually computed using **FKP-like** estimators:

$$B_{\mathrm{FKP}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \sim \left\langle \prod_{i=1}^{3}\left(n_g(\mathbf{k}_i) - \frac{N_g}{N_r}n_r(\mathbf{k}_i)\right)\right\rangle \Big/ \langle n^3\rangle$$

- The theory needs to be convolved with the **mask**

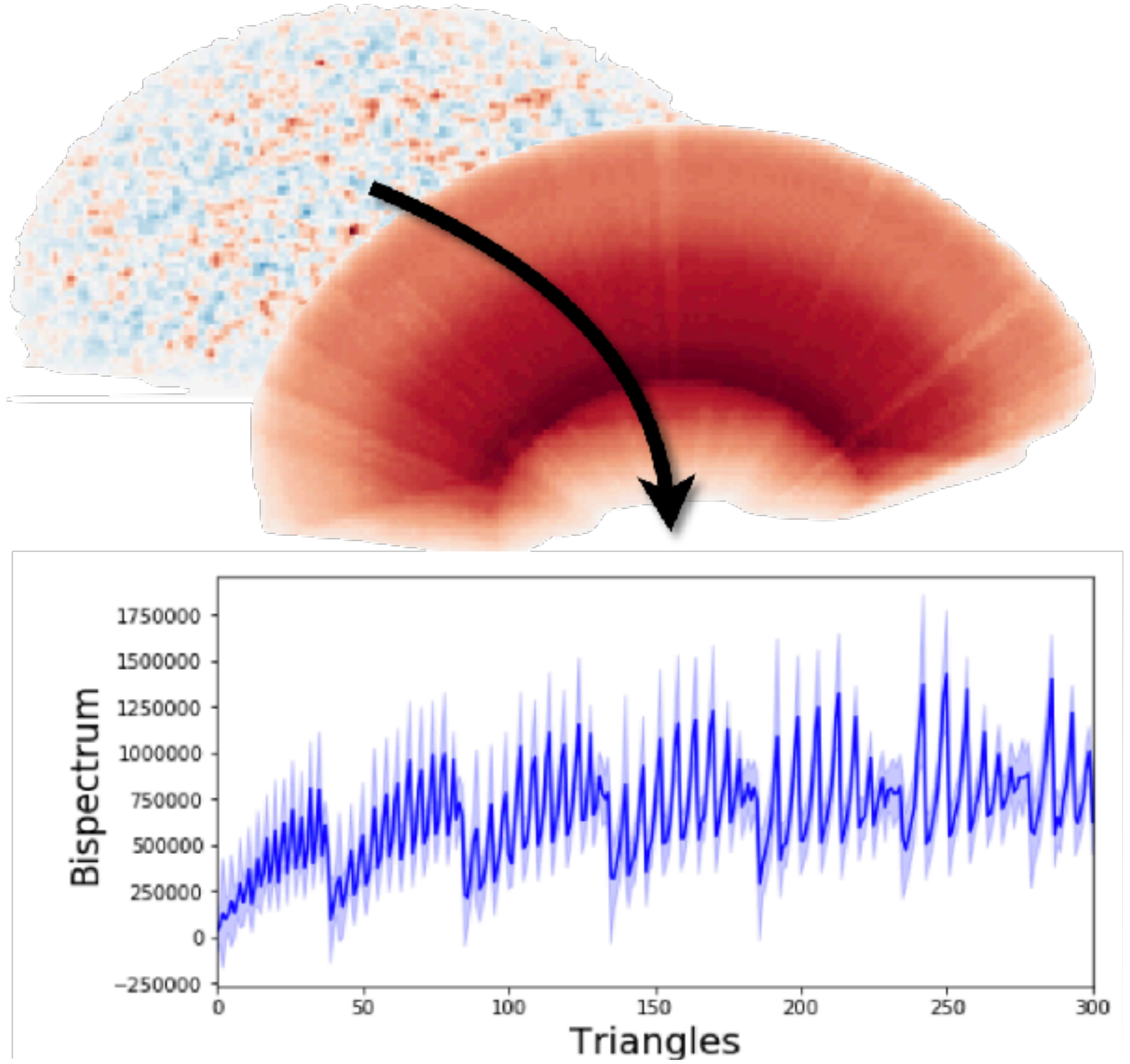
$$B_{\mathrm{FKP}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) \sim \int_{\mathbf{q}_1+\mathbf{q}_2+\mathbf{q}_3=\mathbf{0}} n(\mathbf{k}_1-\mathbf{q}_1)n(\mathbf{k}_2-\mathbf{q}_2)n(\mathbf{k}_3-\mathbf{q}_3)B_{\mathrm{true}}(\mathbf{q}_1,\mathbf{q}_2,\mathbf{q}_3)$$

- This is a **difficult** 6-dimensional integral to compute at **every** step of the MCMC chain

(See Pardede++)

- Various approximations exist, but they can **break down**

(See Gil-Marin+, Chen+)



Feldman, Kaiser, Peacock 1980s, BOSS DR12, Beutler+, Gil-Marin+, McDonald+, Ivanov+19, **Philcox**+20,21, **Philcox** & Floss 2025

Three-Point Estimators

- We compute bispectra using quasi-optimal **“unwindowed”** estimators:

$$\color{blue}{B^{\rm unwin}_{{\rm bin}\,a}} \sim \sum_{b \in bins} \mathscr{W}^{-1}_{ab} \left\langle \prod_{i=1}^{3} \left(n_g(\mathbf{k}_i) - \frac{N_g}{N_r} n_r(\mathbf{k}_i) \right) \right\rangle_{{\rm bin}\,b} \quad \text{(+ linear term)}$$

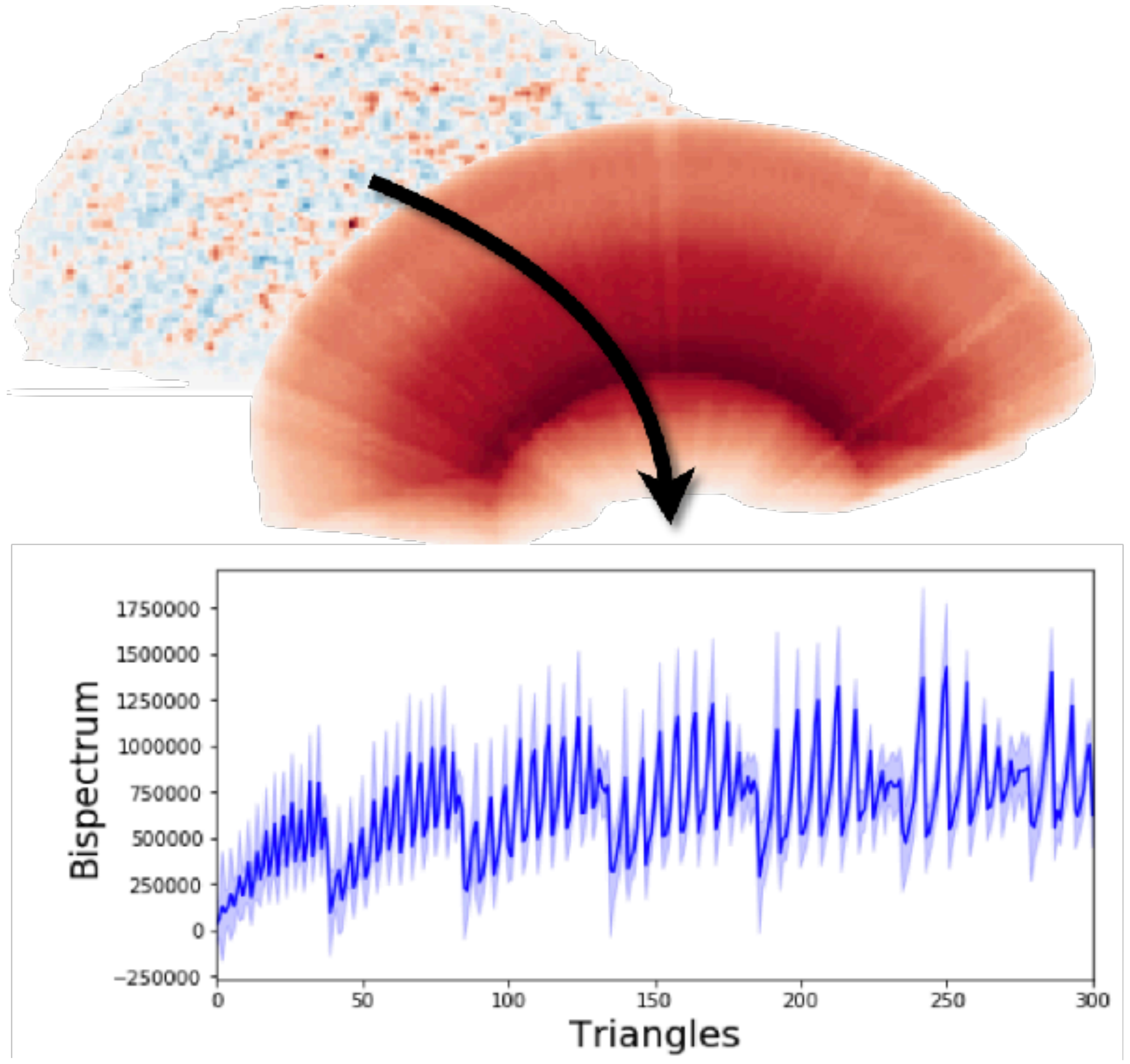
- At leading-order (a good approximation), this does **not** need to be convolved with the mask

$$\color{blue}{B^{\rm unwin}_{{\rm bin}\,a}} \sim B^{\rm true}_{{\rm bin}\,a}$$

- ***Plus**,* it’s almost optimal, even on **large-scales**

- Instead of **mask-convolving** the theory, we **mask deconvolve** the data!

- The **normalization** can be efficiently computed using **Monte Carlo** methods and FFTs
  (**Philcox**, Floss 2025)



Both $P + B$ are computed using the **PolyBin3D** code
(**Philcox**, Floss 2025)

Feldman, Kaiser, Peacock 1980s, BOSS DR12, Beutler+, Gil-Marin+, McDonald+, Ivanov+19, **Philcox**+20,21, **Philcox** & Floss 2025

Accounting for Systematics

- **Radial integral constraint**
  - [missing line-of-sight fluctuations]
  - Included in **normalization** and **theory matrix**
- **Imaging systematics**
  - [the Galaxy contaminates angular modes]
  - Included in **weights** (as in DESI)
  - Template marginalized out over for ELG2 and QSO
- **Stochasticity**
  - [the sample is discrete]
  - Subtract **Poisson** shot-noise

$$P_{\rm shot} \sim \bar{n}_g^{-1}, B_{\rm shot} \sim \bar{n}_g^{-2} + \bar{n}_g^{-1} P(k)$$

- **Wide-angle effects**
  - Included in power spectrum **theory matrix**

$$[\mathrm{Window} \star P^{\rm WA}_{\ell=1,3}] \sim [\mathrm{Window}' \star P^{\rm true}_{\ell}]$$

Chudaykin, Ivanov, **Philcox** 2025

Fiber Collisions

- **Small** angular scales in DESI are **contaminated** by observational systematics, *i.e.* **fiber collisions**

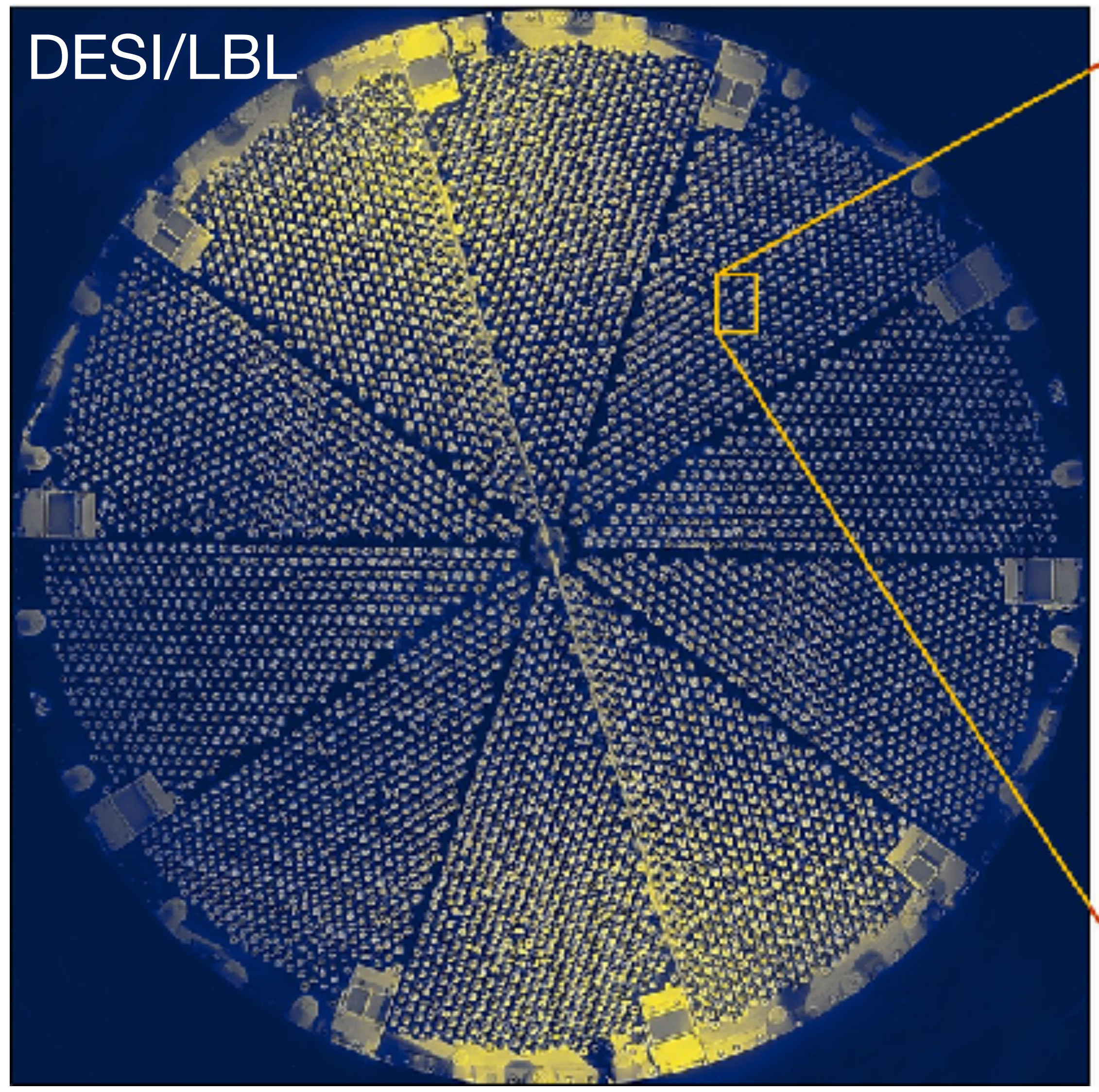
- DESI accounted for this by **removing** pairs of galaxies with small angular separations

$$P(\mathbf{k}) \sim \sum_{\text{galaxy } i} \sum_{\text{galaxy } j} w_i w_j e^{i\mathbf{k}\cdot(\mathbf{x}_i - \mathbf{x}_j)} \times \begin{cases} 1 & \text{if } \theta_{ij} > 0.05^\circ \\ 0 & \text{else} \end{cases}$$

- The correction can be computed by explicit **pair-counting**

- We do the same, correcting the **numerator**, the **normalization,** and the **theory matrix** [*i.e.* residual window]

- DESI applied a **rotation** since the new window is strongly off-diagonal
  - This is **automatically** accounted for in our normalization matrix!



DESI Focal Plane

DESI Fibers

Chudaykin, Ivanov, **Philcox** 2025

Fiber Collisions

- Fiber collisions are **harder** for the bispectrum

- To remove all close **pairs** (or triplets), we'd need to count all **triplets** of galaxies, which is $\sim 10^6$ times more expensive!

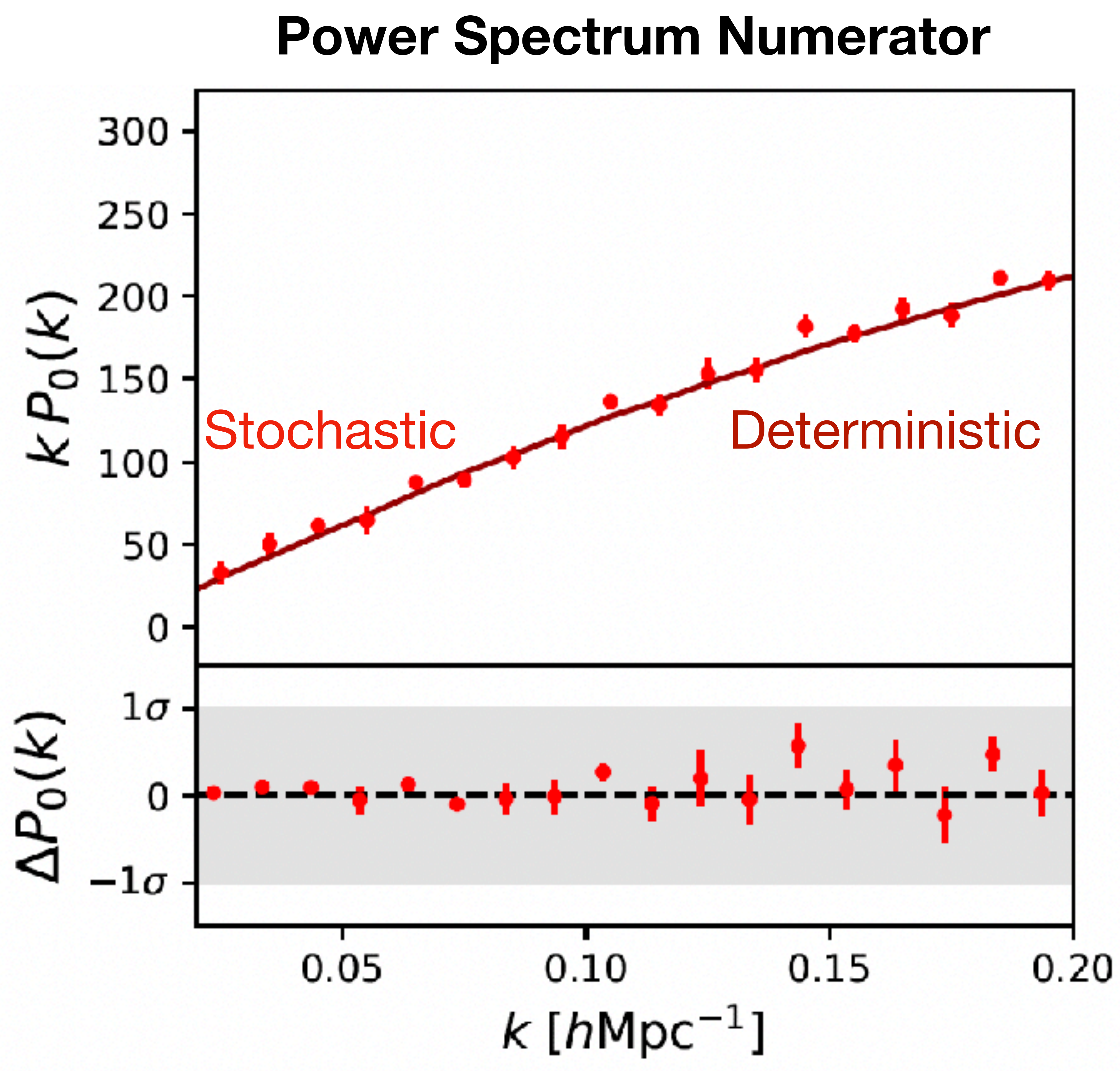
$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \sim \sum_{\text{galaxy } i} \sum_{\text{galaxy } j} \sum_{\text{galaxy } k} w_i w_j w_k (\cdots) \times \begin{cases} 1 & \text{if } \left(\theta_{ij} \text{ and } \theta_{jk} \text{ and } \theta_{ki}\right) > 0.05^\circ \\ 0 & \text{else} \end{cases}$$

- We introduce a novel **stochastic** method for removing these, involving **cross-bispectra**

Power spectrum example:

$$\sum_i w_i \epsilon_i \sum_j w_j \sum_k \epsilon_k \begin{cases} 1 & \text{if } \theta_{jk} > 0.05^\circ \\ 0 & \text{else.} \end{cases} \Rightarrow \sum_i w_i \sum_j w_j \begin{cases} 1 & \text{if } \theta_{ji} > 0.05^\circ \\ 0 & \text{else.} \end{cases}$$

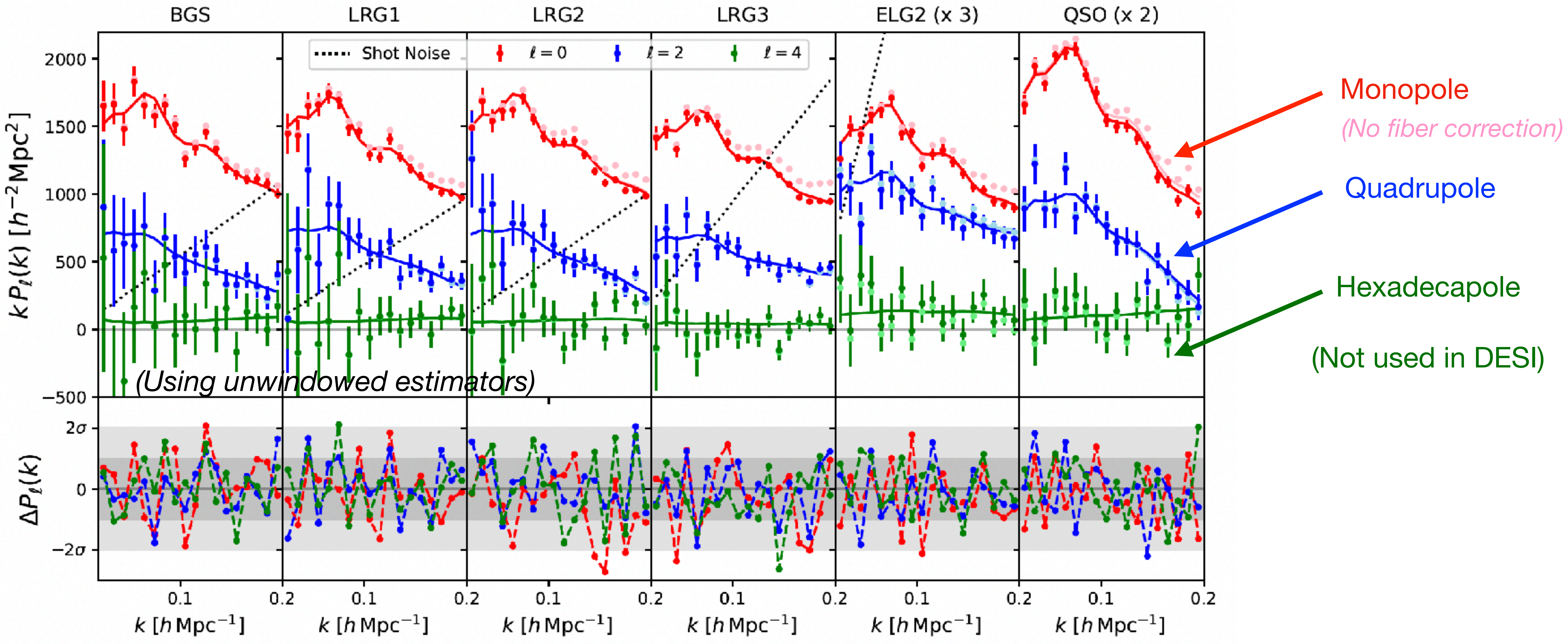
average over $\epsilon_i \sim \mathcal{N}(0,1)$



Chudaykin, Ivanov, **Philcox** 2025

Power Spectrum Data-Set

$P_\ell(k)$ for $\ell = 0{,}2{,}4$ and $0.02\,h\mathrm{Mpc}^{-1} \le k \le 0.20\,h\mathrm{Mpc}^{-1}$

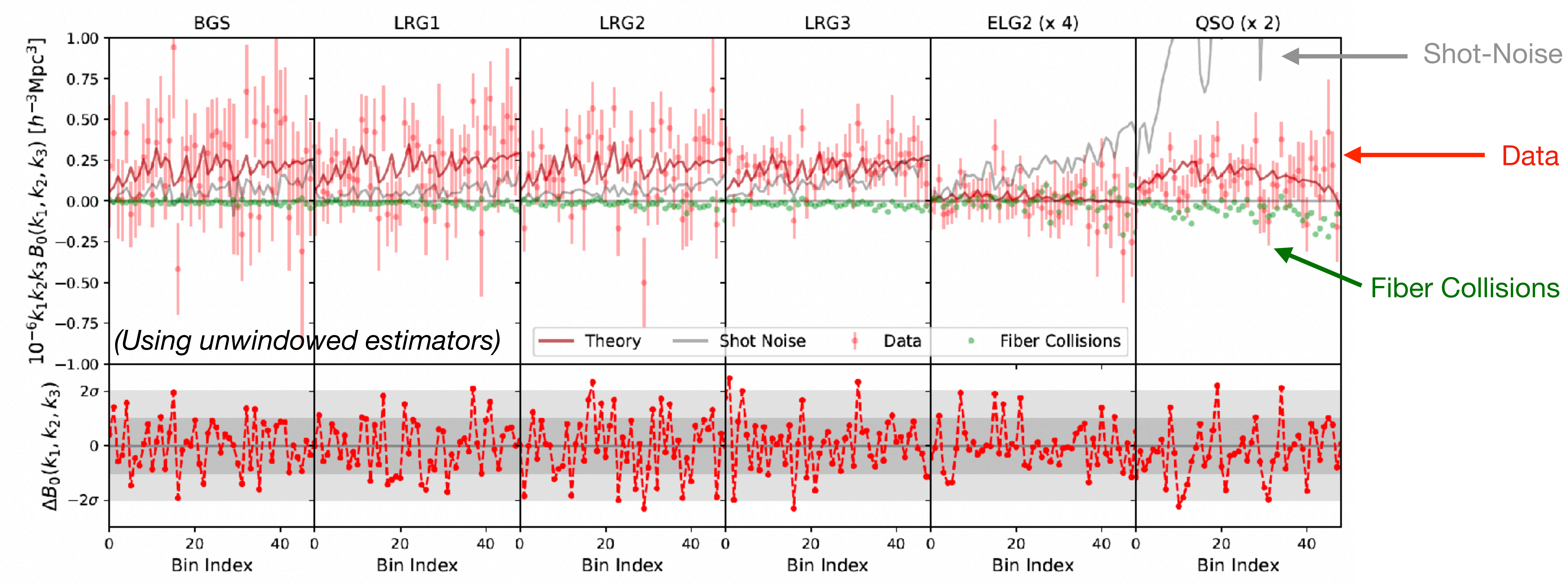


(Maximum SNR: 230σ)

Chudaykin, Ivanov, **Philcox** 2025

Bispectrum Data-Set

$B_0(k)$ for $0.02\,h\mathrm{Mpc}^{-1} \leq k \leq 0.08\,h\mathrm{Mpc}^{-1}$



(Maximum SNR: 9σ)

Chudaykin, Ivanov, **Philcox** 2025

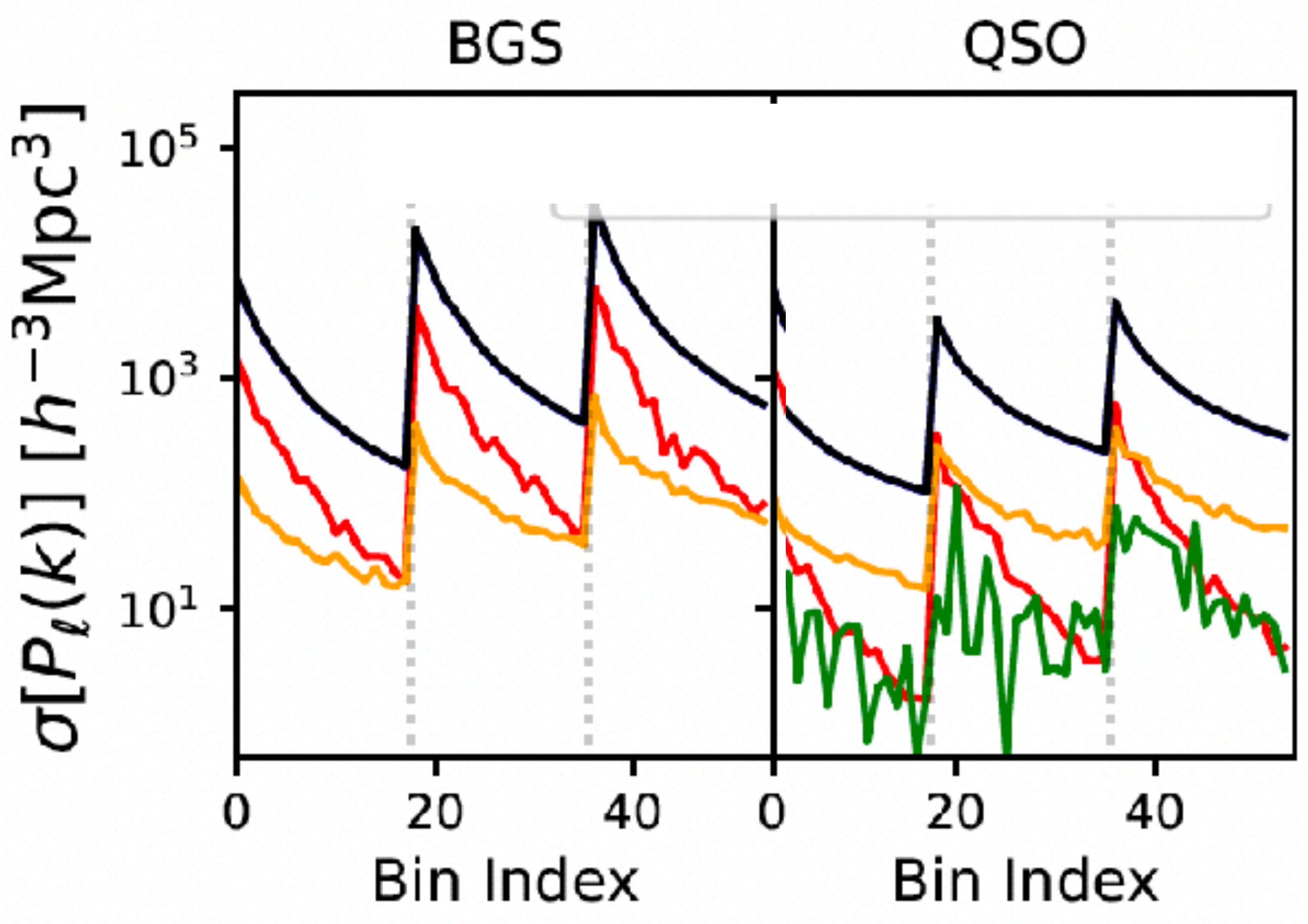
Covariance Matrix

- DESI computed the covariances using **simulations**
- We compute covariance **analytically**

$$\mathrm{cov}[P] \sim P^2(k) \star \mathrm{mask}^4, \qquad \mathrm{cov}[B] \sim P^3(k) \star \mathrm{mask}^6$$

  - This is computed on a **grid**, accounting for the **mask**
  - **PolyBin3D** measures this similarly to the **normalization** and **theory matrix**
  - This depends on a power spectrum **model** fit from the **data**
- **Note:** we do not include non-Gaussian terms…
- DESI rescaled the **simulation** covariance to match a **data-calibrated** theory covariance — we do not need to do this! (It also does not make a difference)
- We **inflate** covariance to include various sources of noise

DESI $P(k)$ Covariance

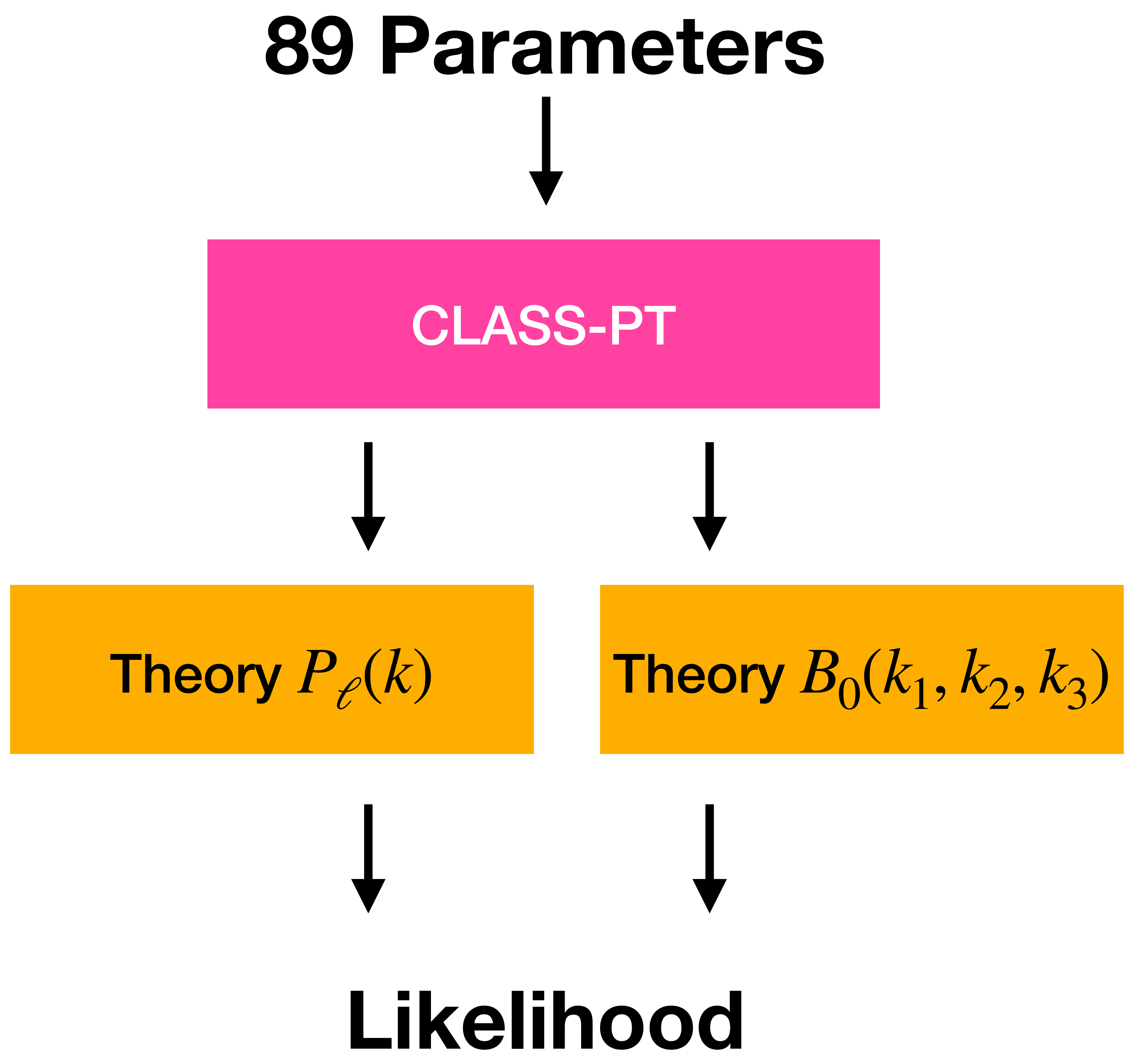


Total *vs* Theory Matrix *vs* Shot *vs* Systematics

DESI 2024 V, Rashkovetskyi+, Forero-Sanchez+, Chudaykin, Ivanov, **Philcox** 2025

Theoretical Model

- We fit the data with the **Effective Field Theory** of **Large Scale Structure** at **one-loop** for P and **tree-level** for B
- **Parameters** (assuming ΛCDM):
  - Cosmology: free $(H_0, \omega_{\rm cdm}, \log A_s)$, plus priors on (ω_b, n_s)
  - Bias: free $(b_1, b_2, b_{\mathcal{G}_2}, b_{\Gamma_3})$
  - Stochasticity: free $P_{\rm shot}, B_{\rm shot}, A_{\rm shot}, a_0, a_2$
  - Counterterms: free $c_0, c_2, c_4, \tilde{c}, c_1$
- Most parameters are **analytically marginalized**
- We rescale parameters by σ_8 according to degeneracies



DESI 2024 V, Maus+, Chudaykin, Ivanov, **Philcox** 2025

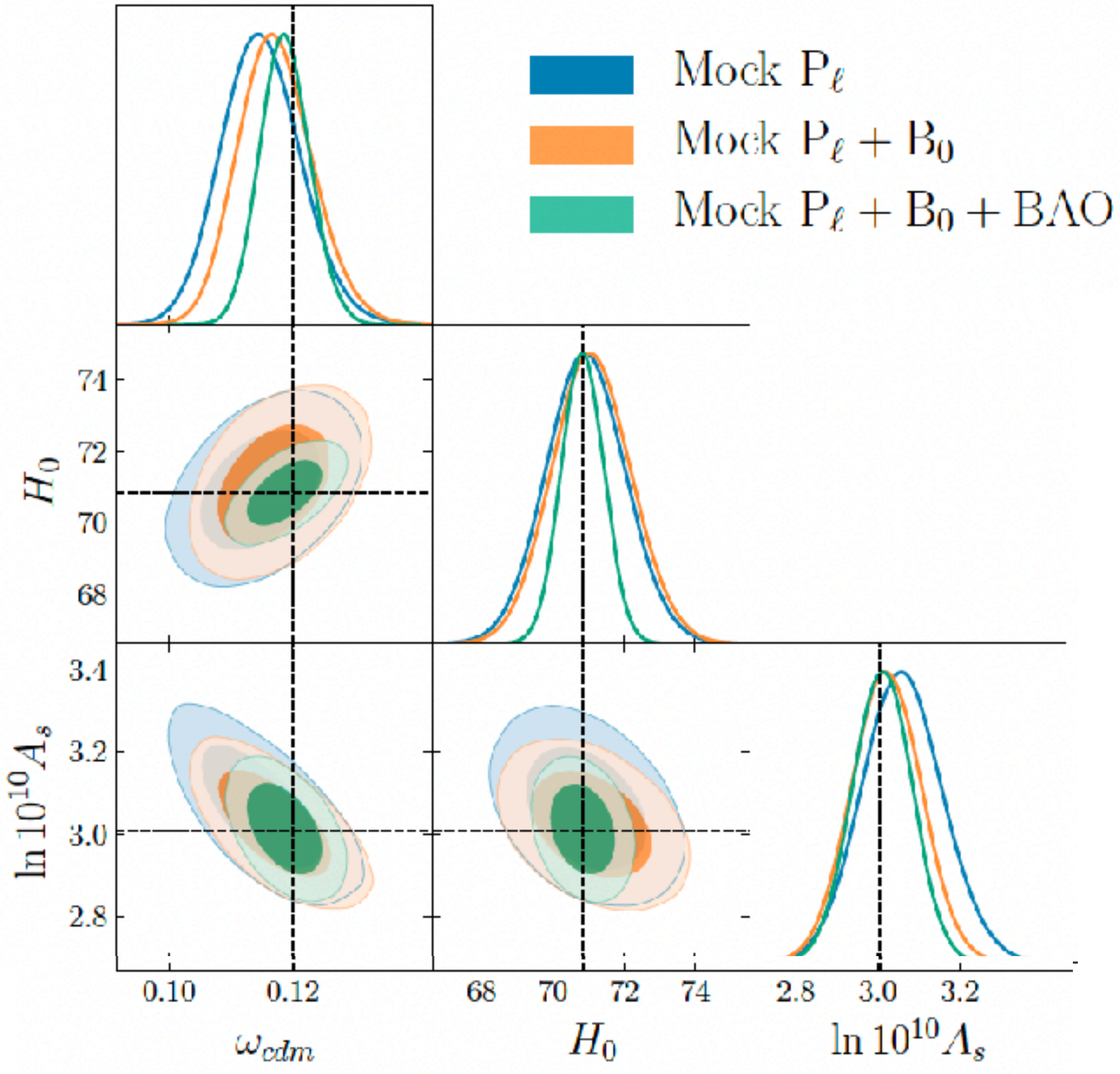
Projection Effects

- When analyzing synthetic data, do we recover the input cosmology?
  - The best-fit can be shifted due to **non-Gaussian** posteriors and **parameter degeneracies**
  - This is a consequence of **Bayes' theorem**, but minimizing these helps to interpret **posteriors**

- We find **good consistency** with inputs, particularly when **extra data** is added ($< 0.4\sigma$)

- The **rescaled biases** (e.g., $b_1 \to b_1\sigma_8$) help a lot!

  e.g. Maus+24



DESI 2024 V, Maus+, Chudaykin, Ivanov, **Philcox** 2025

Constraints on ΛCDM

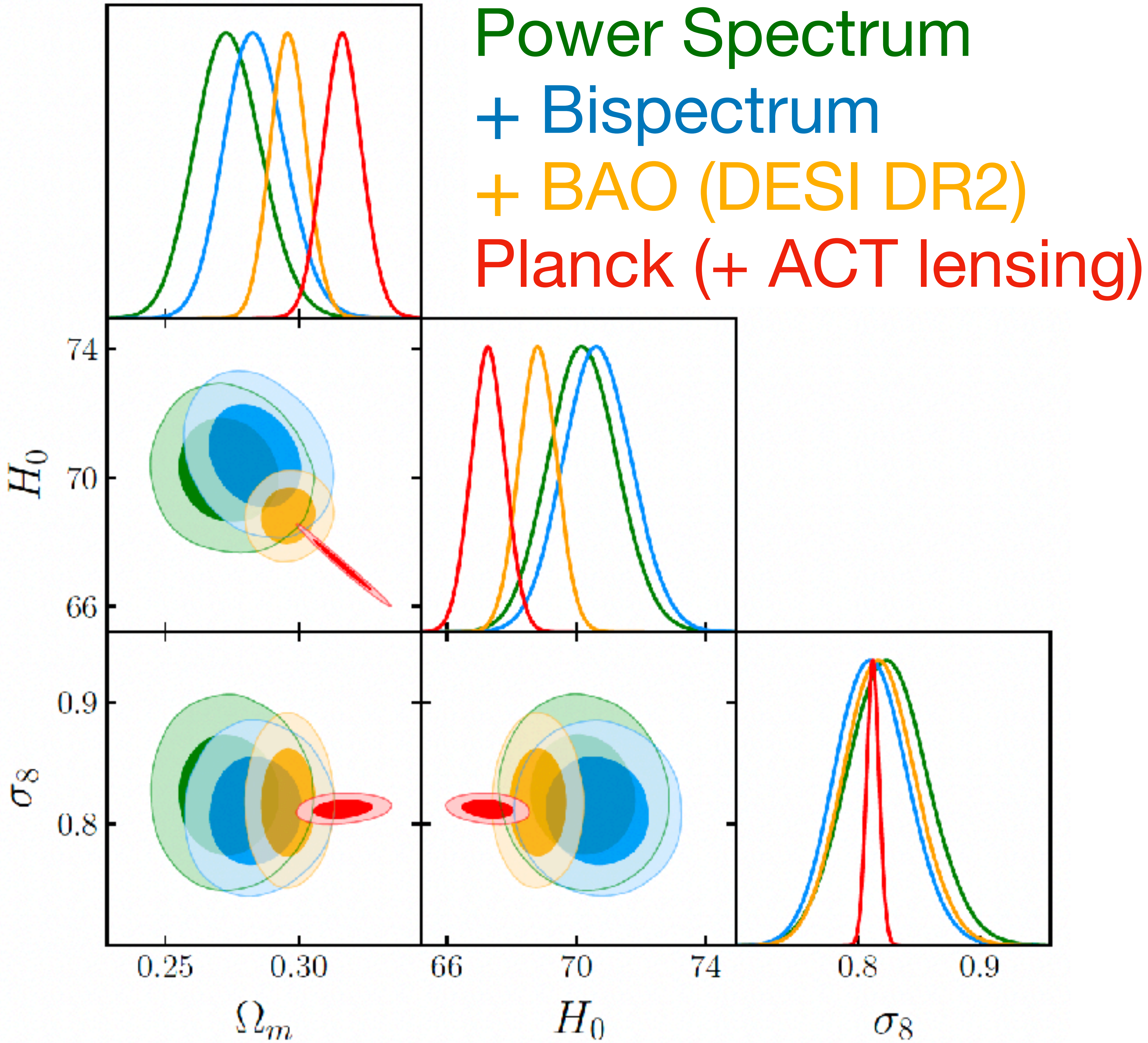
DESI alone finds **strong constraints** on Ω_m, H_0, σ_8

- Adding the **bispectrum** improves constraints by $\sim$10%
- Adding the (official) DESI DR2 BAO gives **significant** improvements in Ω_m, H_0
- No evidence for H_0 tension or S_8 tension $(S_8 = 0.813 \pm 0.031)$

Our constraints are **broadly consistent** with Planck

- $P + B + \mathrm{BAO}$ dataset matches CMB to 2σ (1.8σ with PR4)

Dataset	Ω_m	H_0	σ_8
$P_\ell(k)$	$0.274^{+0.012}_{-0.013}$	$70.22^{+1.06}_{-1.06}$	$0.825^{+0.033}_{-0.033}$
$P_\ell(k) + B_0(k)$	$0.284^{+0.010}_{-0.012}$	$70.67^{+1.05}_{-1.05}$	$0.811^{+0.028}_{-0.031}$
$P_\ell(k) + B_0(k) + \mathrm{BAO}$	$0.296^{+0.007}_{-0.007}$	$68.82^{+0.58}_{-0.58}$	$0.818^{+0.029}_{-0.029}$
CMB	$0.316^{+0.007}_{-0.007}$	$67.28^{+0.53}_{-0.53}$	$0.812^{+0.005}_{-0.005}$

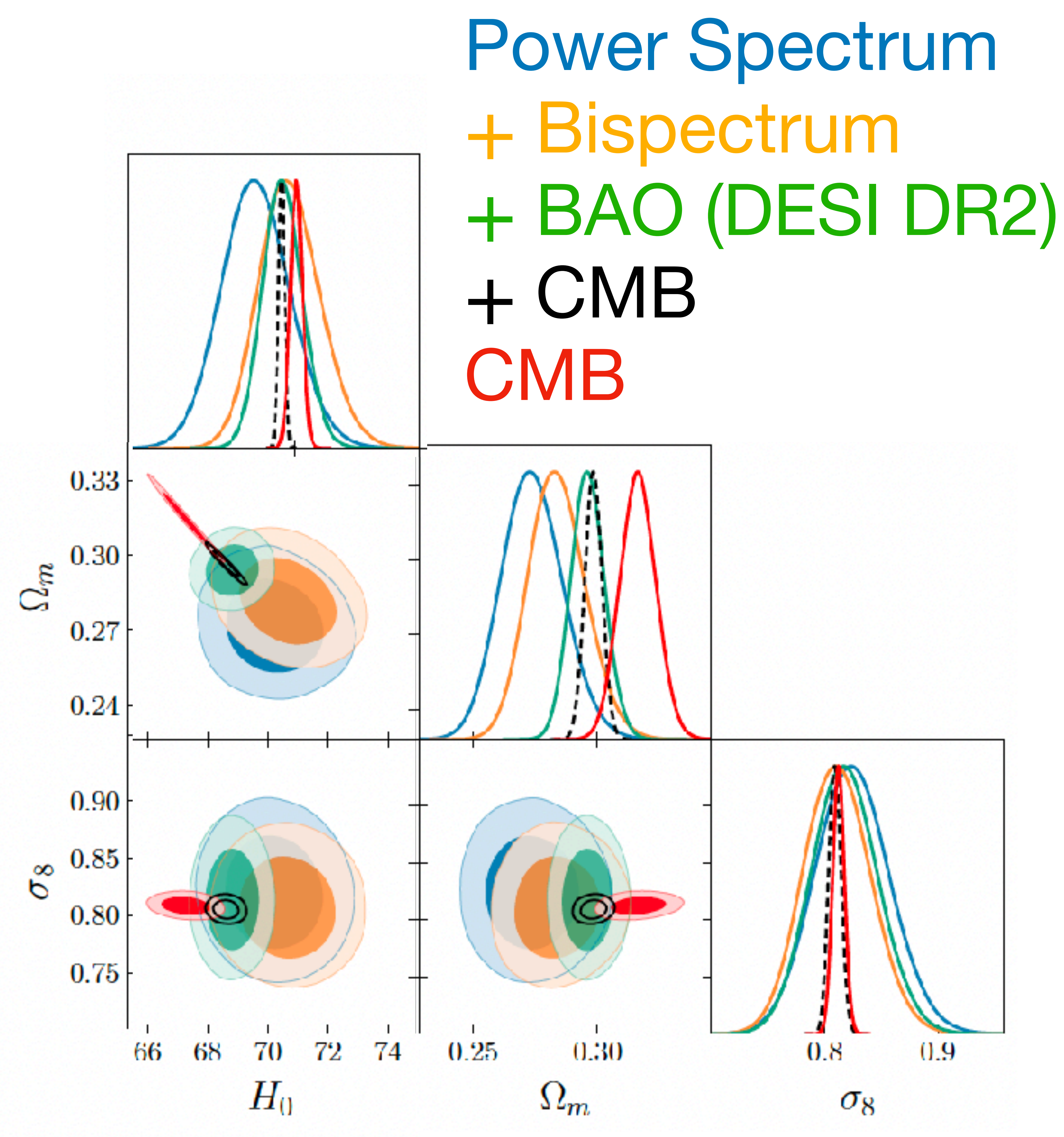


Chudaykin, Ivanov, **Philcox** 2025

Constraints on ΛCDM

We find even **stronger** constraints combining with the CMB:

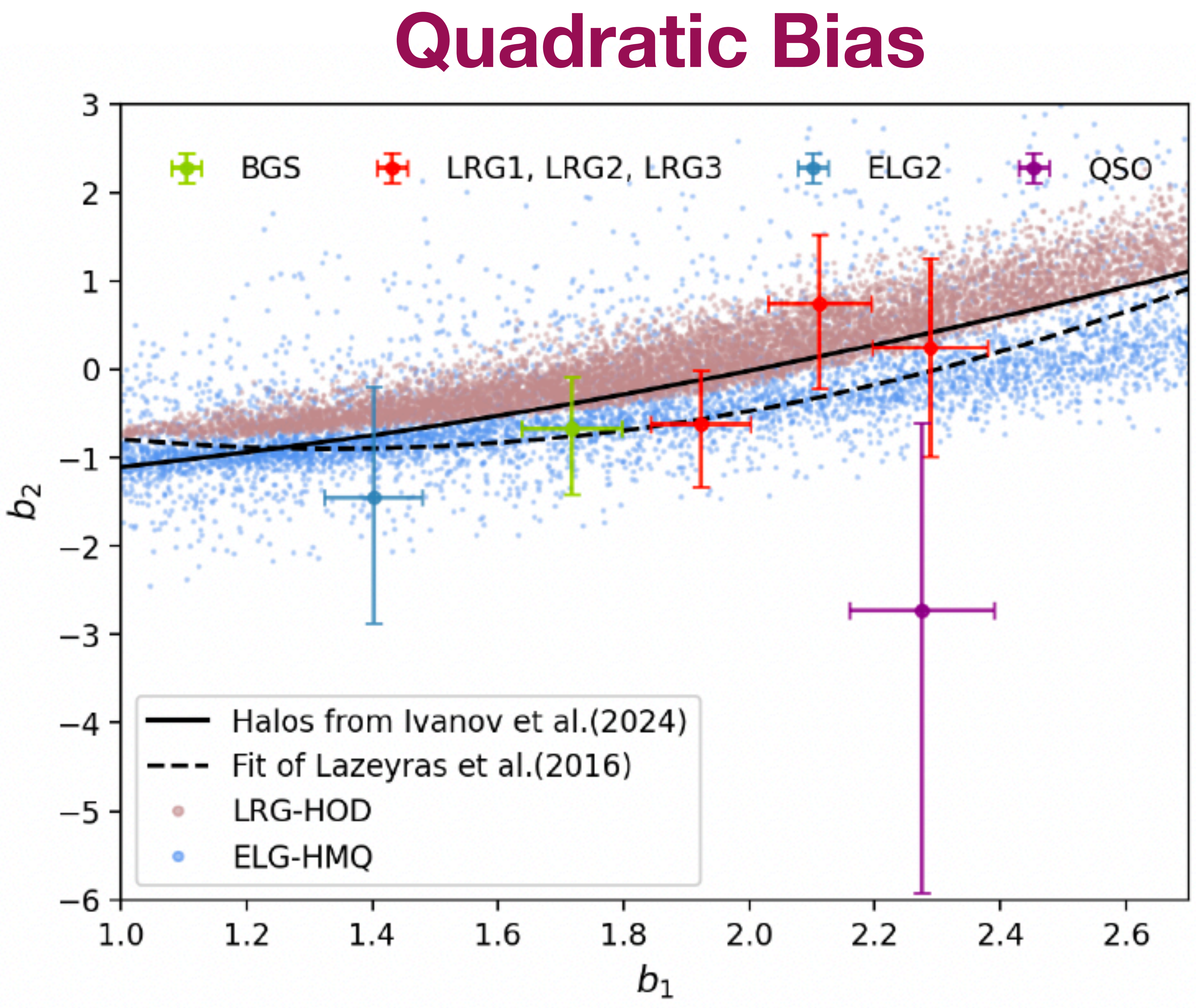
- DESI **enhances** Planck constraints up to $2\times$
  - $\Omega_m = 0.298 \pm 0.003$
  - $H_0 = 68.61 \pm 0.28$
  - $\sigma_8 = 0.809 \pm 0.005$
- Ω_m is still a **bit low** *but* it shifts towards Planck as more data is added

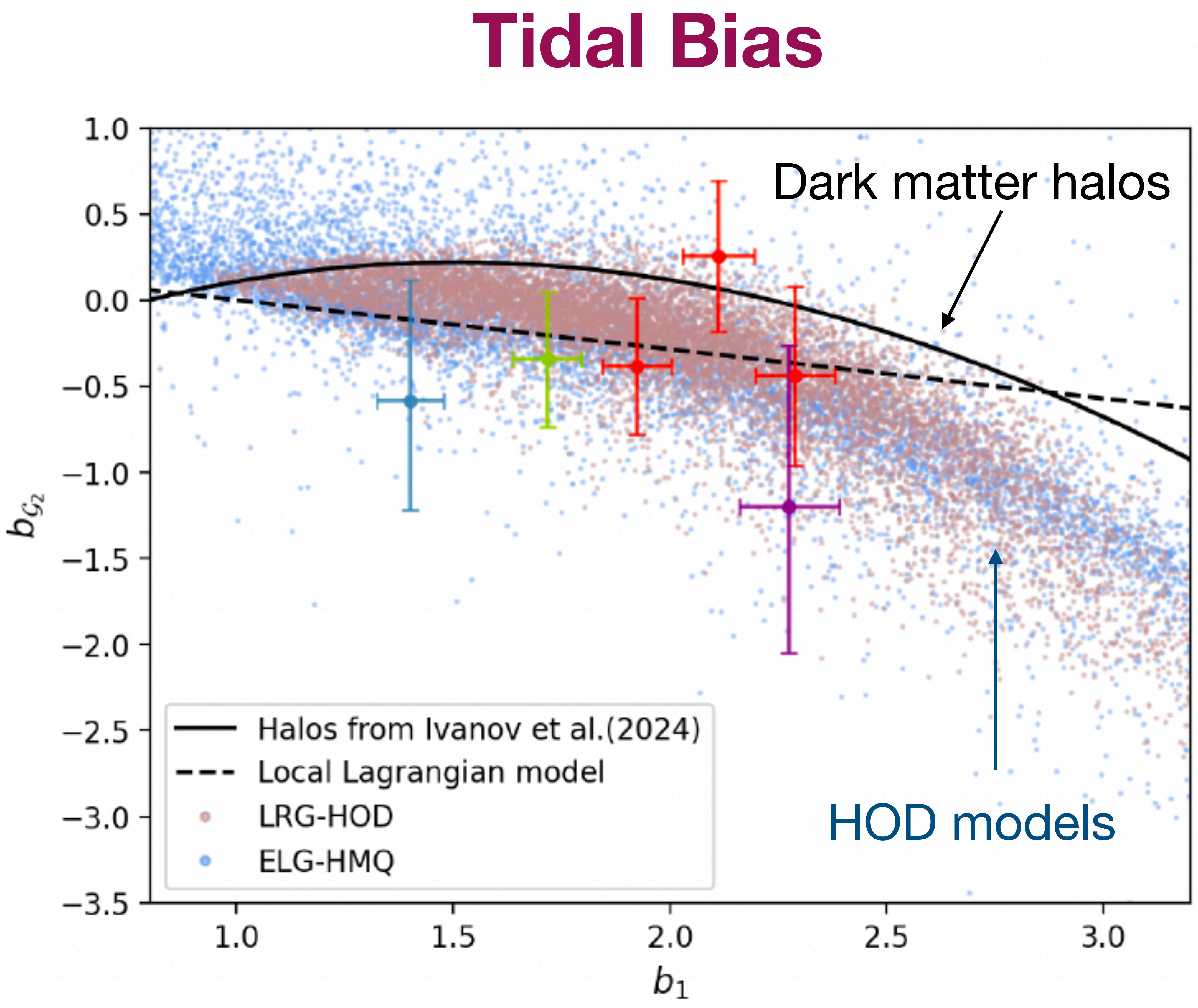


Chudaykin, Ivanov, **Philcox** 2025

Constraints on Bias Parameters

- Adding the **bispectrum** leads to strong constraints on **bias parameters**
- These agree with HOD predictions, but not with **dark matter** predictions

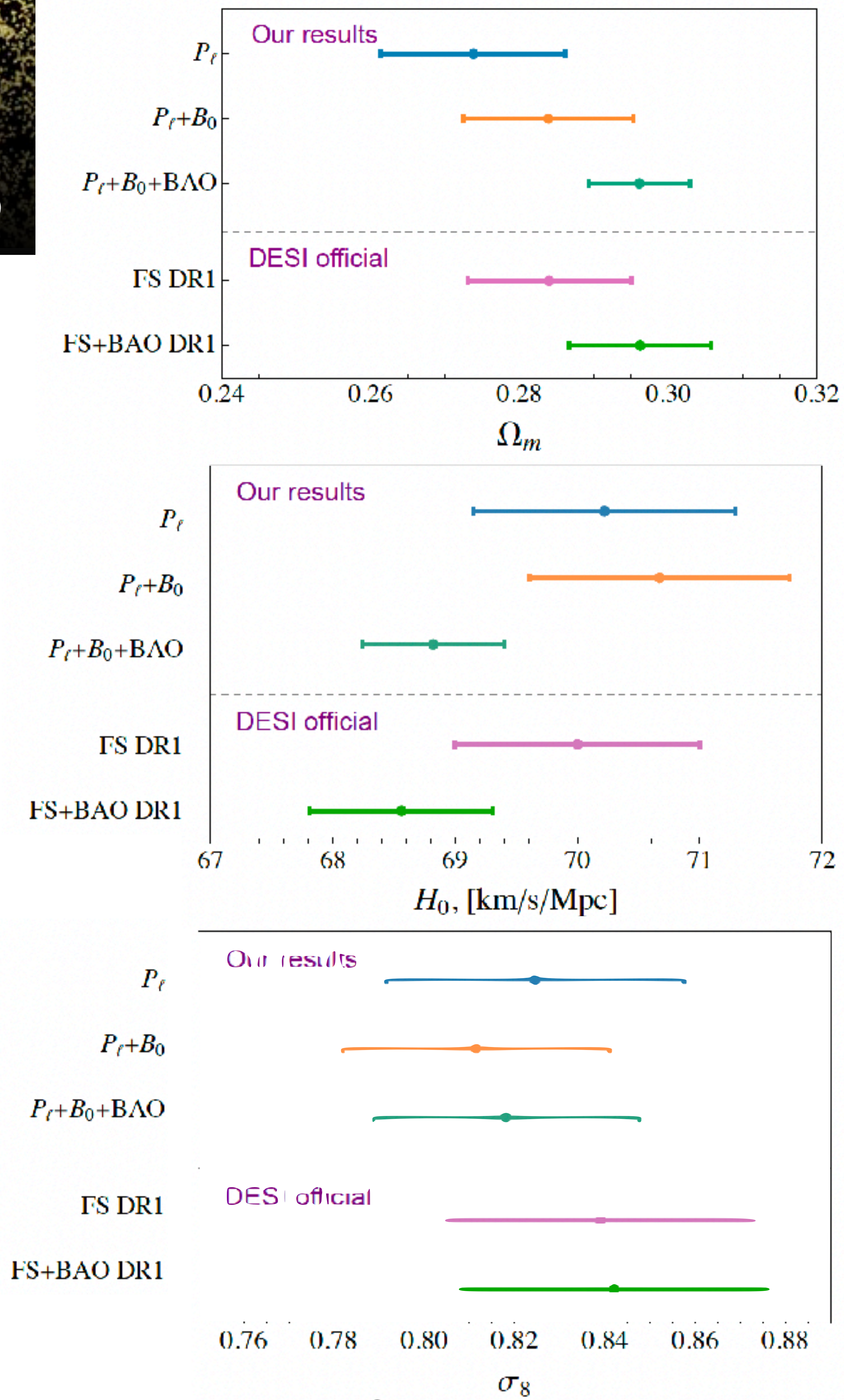




Chudaykin, Ivanov, **Philcox** 2025

Comparison to Official Results

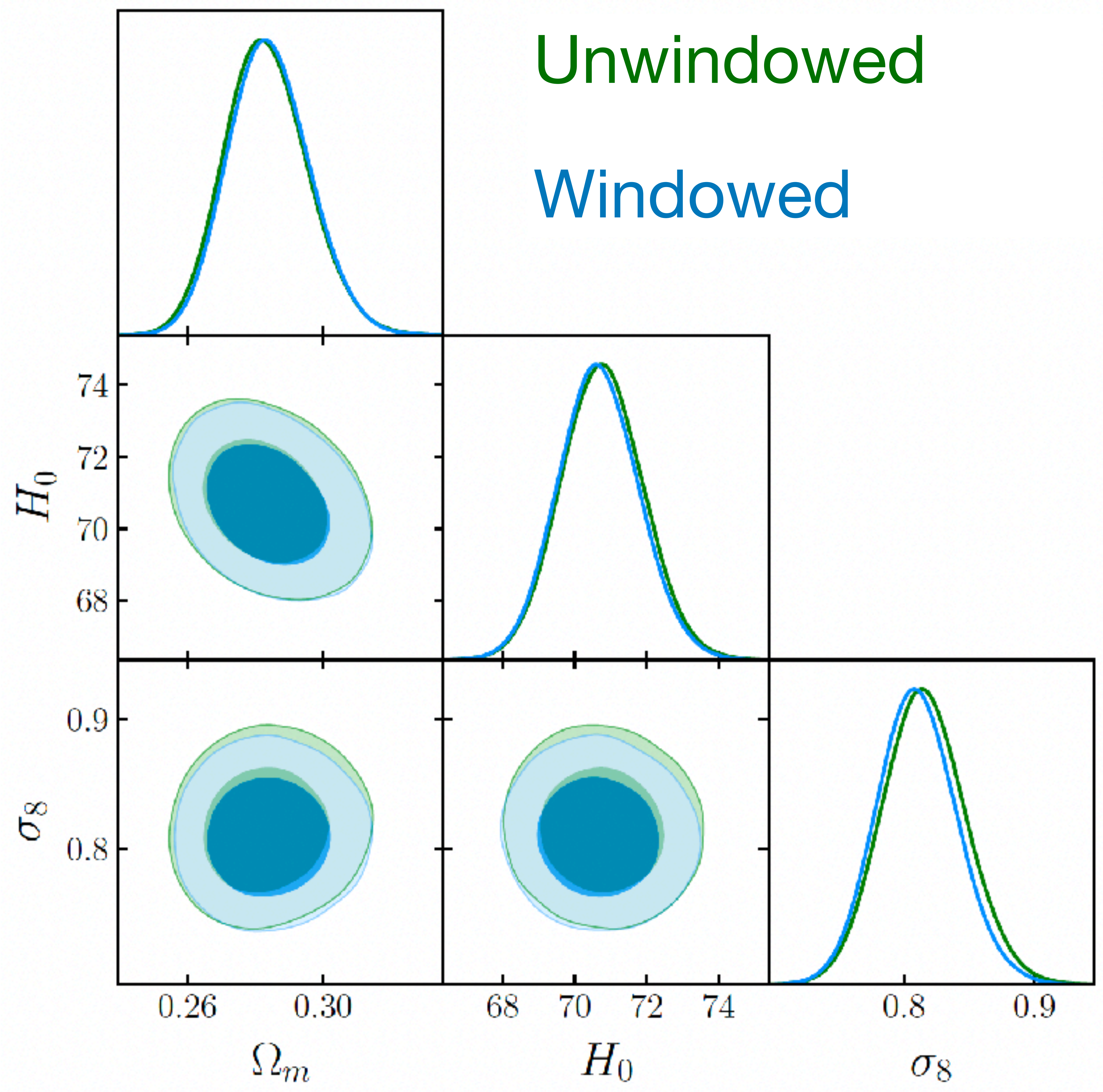
- We find **fairly good** agreement with DESI with $\Delta\Omega_m = -0.8\sigma, \Delta H_0 = +0.2\sigma, \Delta\sigma_8 = -0.4\sigma$
- The main differences are due to:
  - Addition of the **hexadecapole** (P_4)
  - Free **scale-dependent** shot-noise ($P \supset \frac{1}{\bar{n}}\left[1+a_0k^2\right]$)
  - Free higher-order **fingers-of-God** counterterm ($\tilde{c}$)
  - Free cubic bias (b_{Γ_3})
  - Analytic covariance
- And of course, the addition of the bispectrum and DR2 BAO



Chudaykin, Ivanov, **Philcox** 2025

Other Systematic Checks

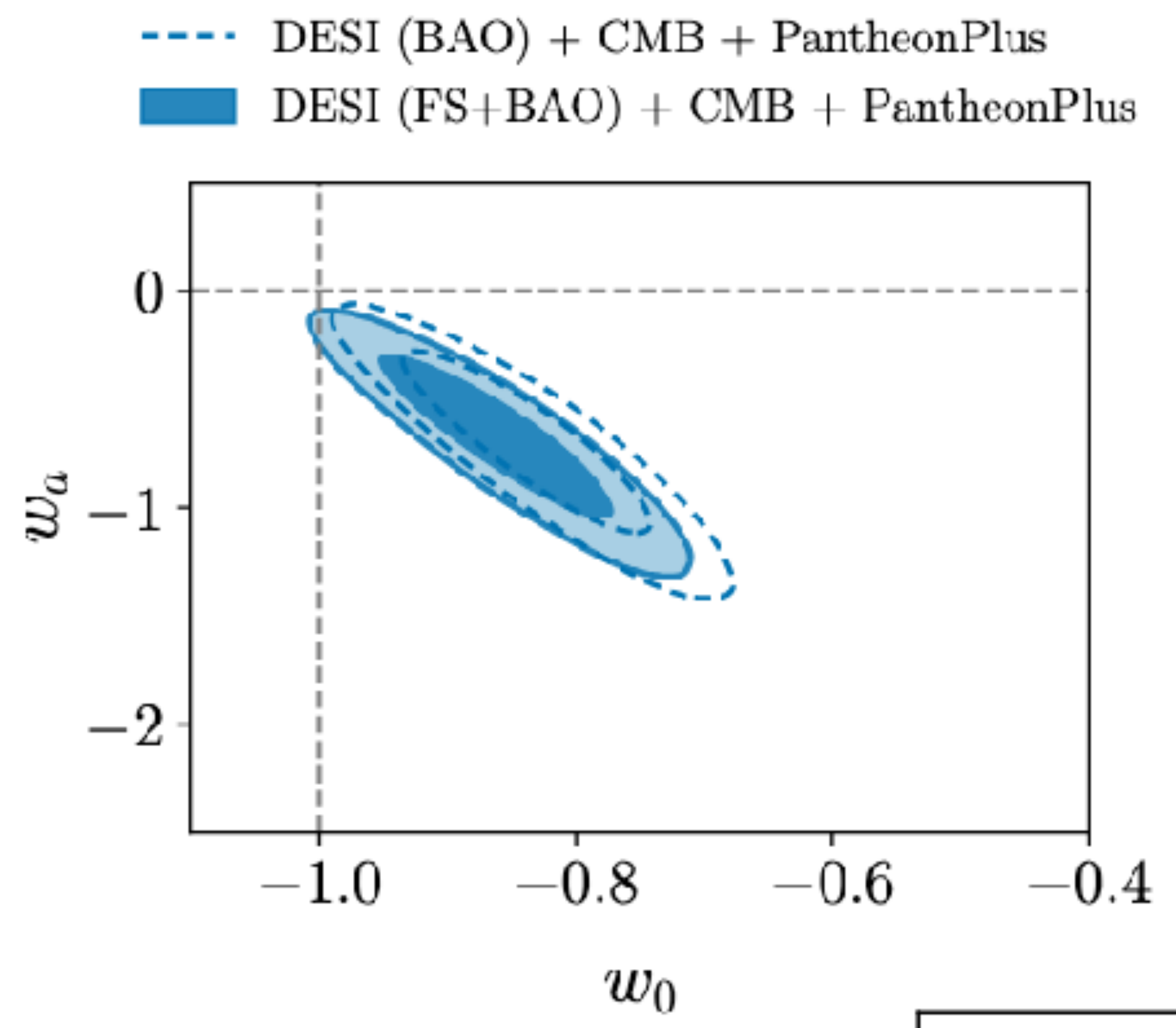
- Results are **stable** ($< 0.5\sigma$ shifts) under changes to the bias model
    - *i.e.* fixing $a_0, b_{\Gamma_3}, \tilde{c}$
    - *Or* removing the hexadecapole
- Results are **stable** under switching from **unwindowed** to (conventional) **windowed** estimators
    - (Windowed are much slower however!)

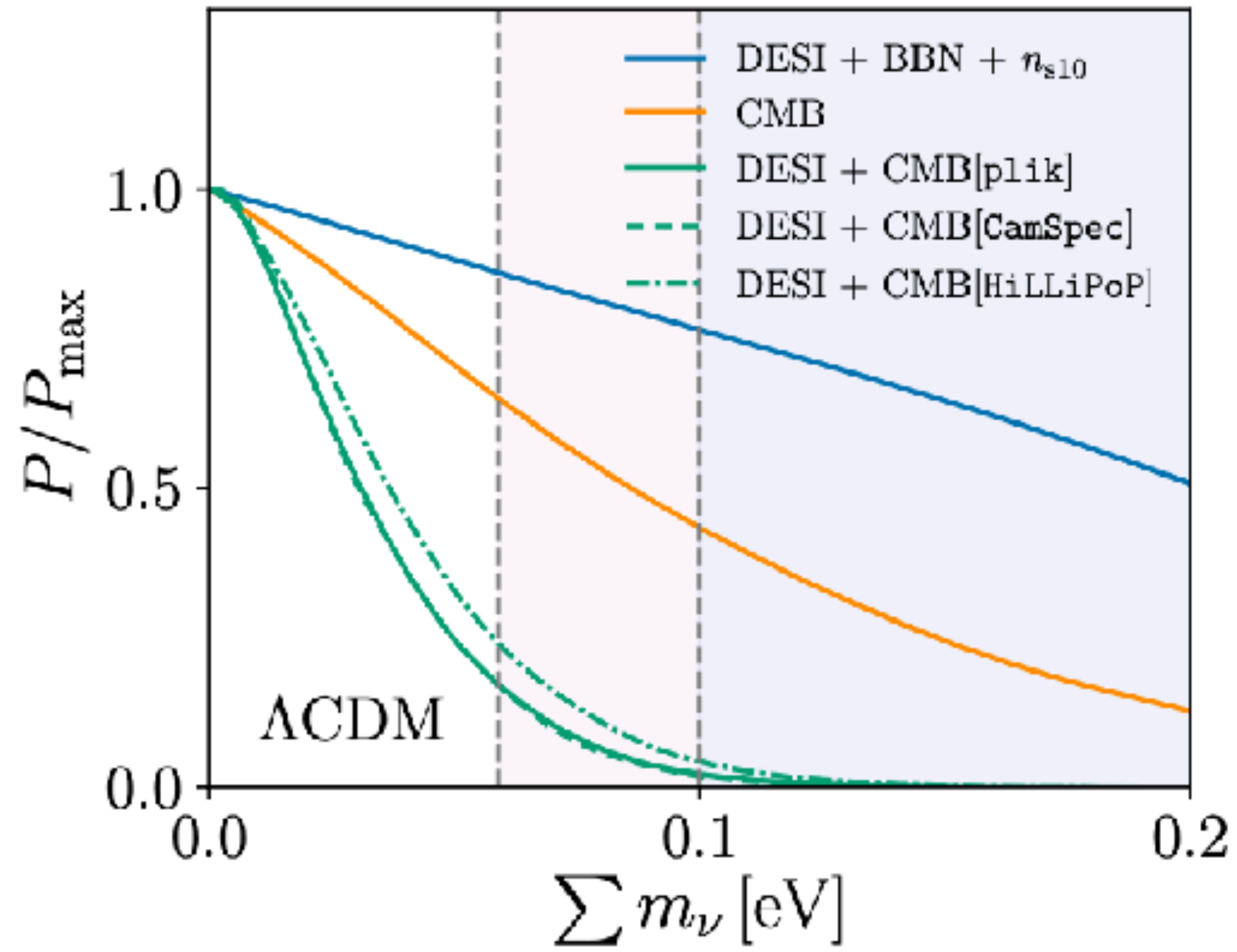


Next works

- There are **many** more models to explore, e.g.,
  - Curvature Ω_k
  - Dark Energy $w(a)$
  - Neutrinos $\sum m_\nu$
  - Primordial non-Gaussianity, $f_{\rm NL}^{\rm loc,\, eq,\, orth}$
- There are **many** more datasets to explore, e.g.,
  - Combined BAO and full-shape data
  - Bispectrum multipoles, B_ℓ
  - Smaller scale bispectra (one-loop)

Can we independently reproduce these?





SIMONS FOUNDATION

COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Stanford University

DESI (edited)

Summary

- We perform a **full renalysis** of the **public** DESI DR1 (full-shape), using **independent** *estimators*, *theory codes*, and *covariances*!

- We find **consistency** within $\approx 0.5\sigma$, and we add **new data** $(P_4 + B_0)$

- There's **a lot** more to explore with the data!

Contact: ohep2@cantab.ac.uk