

Digital Twins of the Universe: Machine Learning Against Cosmological Bias

Jenny Sorce
and collaborators

Researcher at CRIStAL, Lille
Associate Researcher at IAS, Orsay

ML4Astro3, Malta, Valletta
September 4th, 2026



Digital Twins of the Universe: Machine Learning Against Cosmological Bias (Bayesian Inference+)

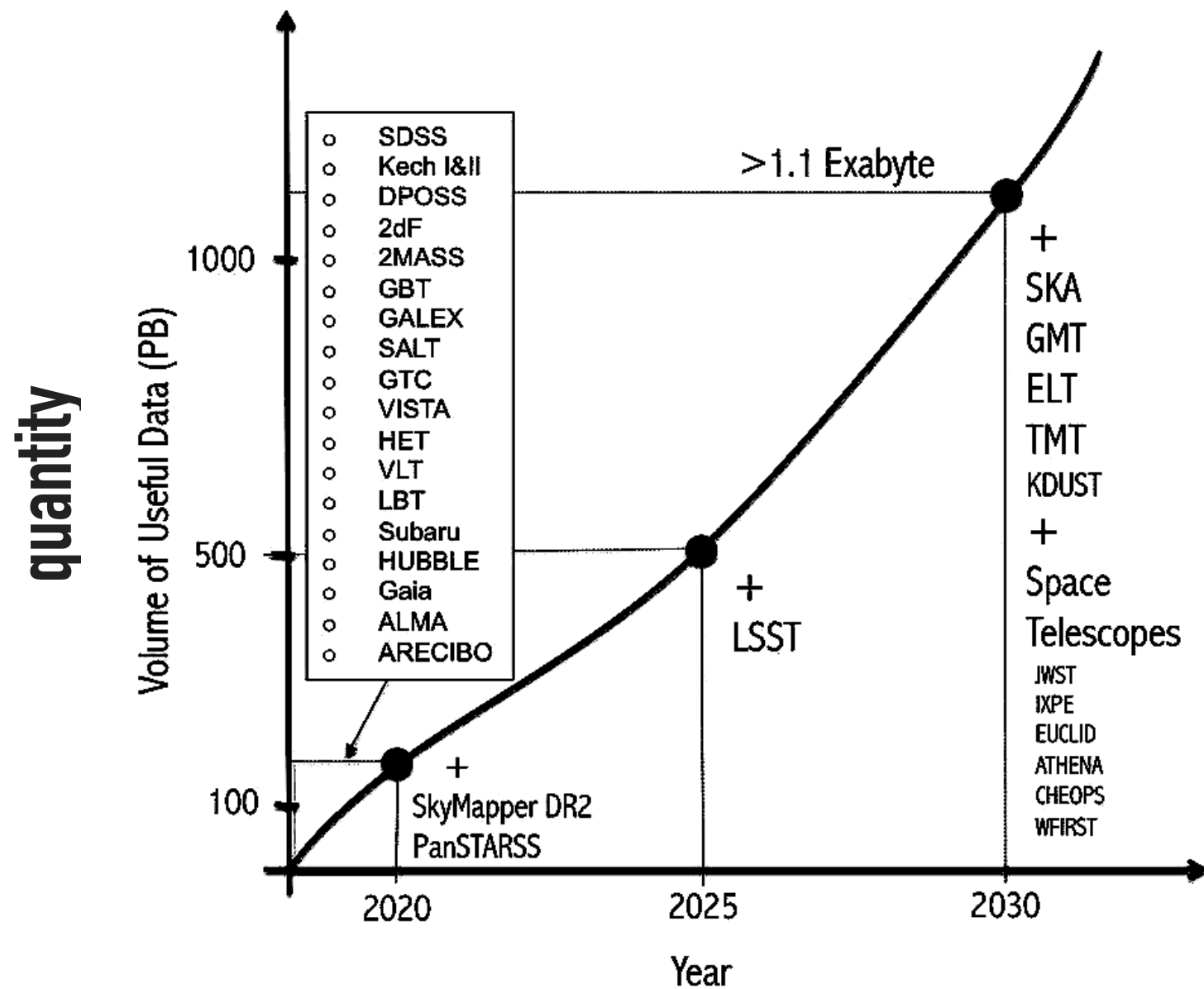
Jenny Sorce
and collaborators

Researcher at CRIStAL, Lille
Associate Researcher at IAS, Orsay

ML4Astro3, Malta, Valletta
September 4th, 2026



Observational data

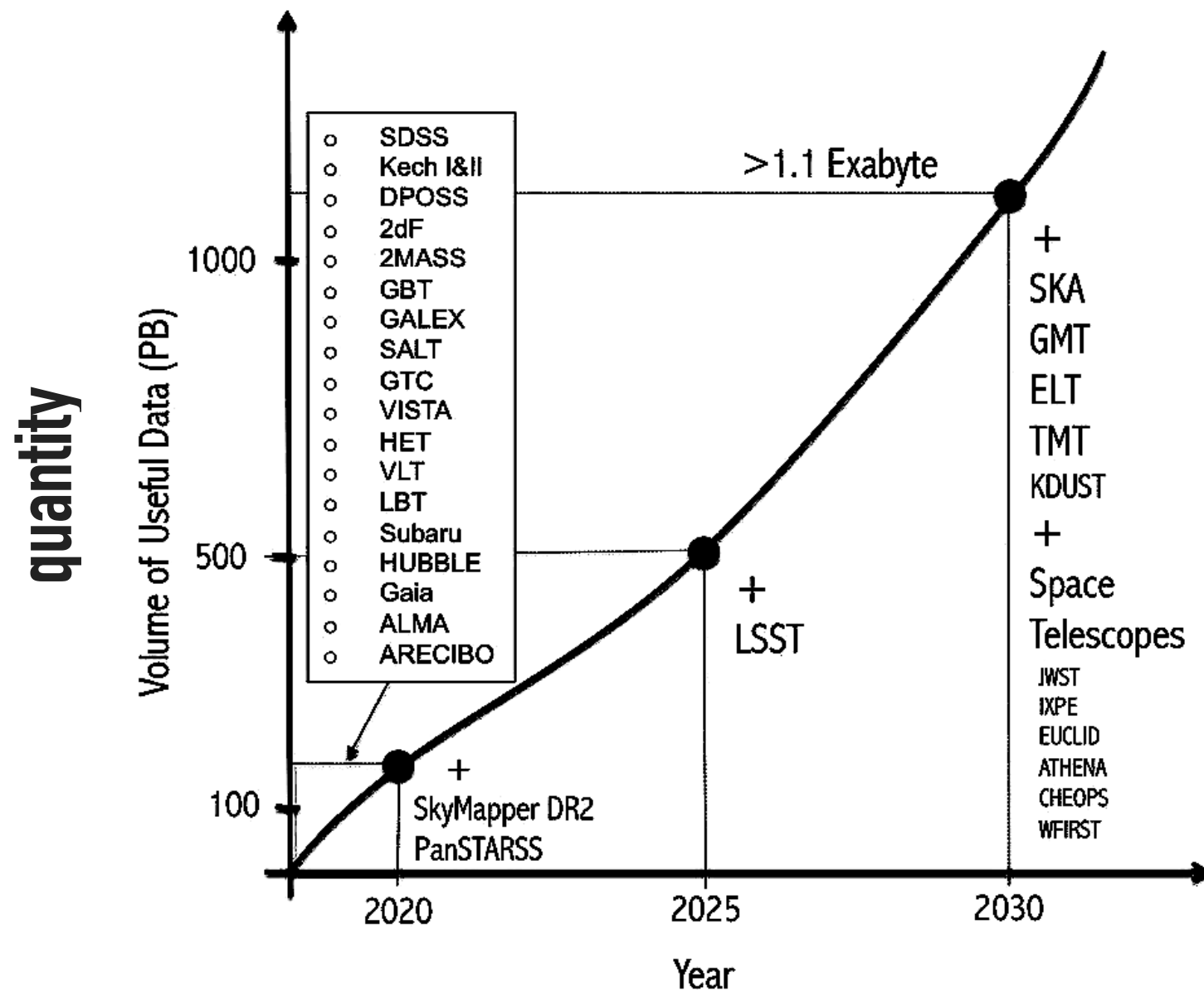


© Rosa

Interpretation?



Observational data



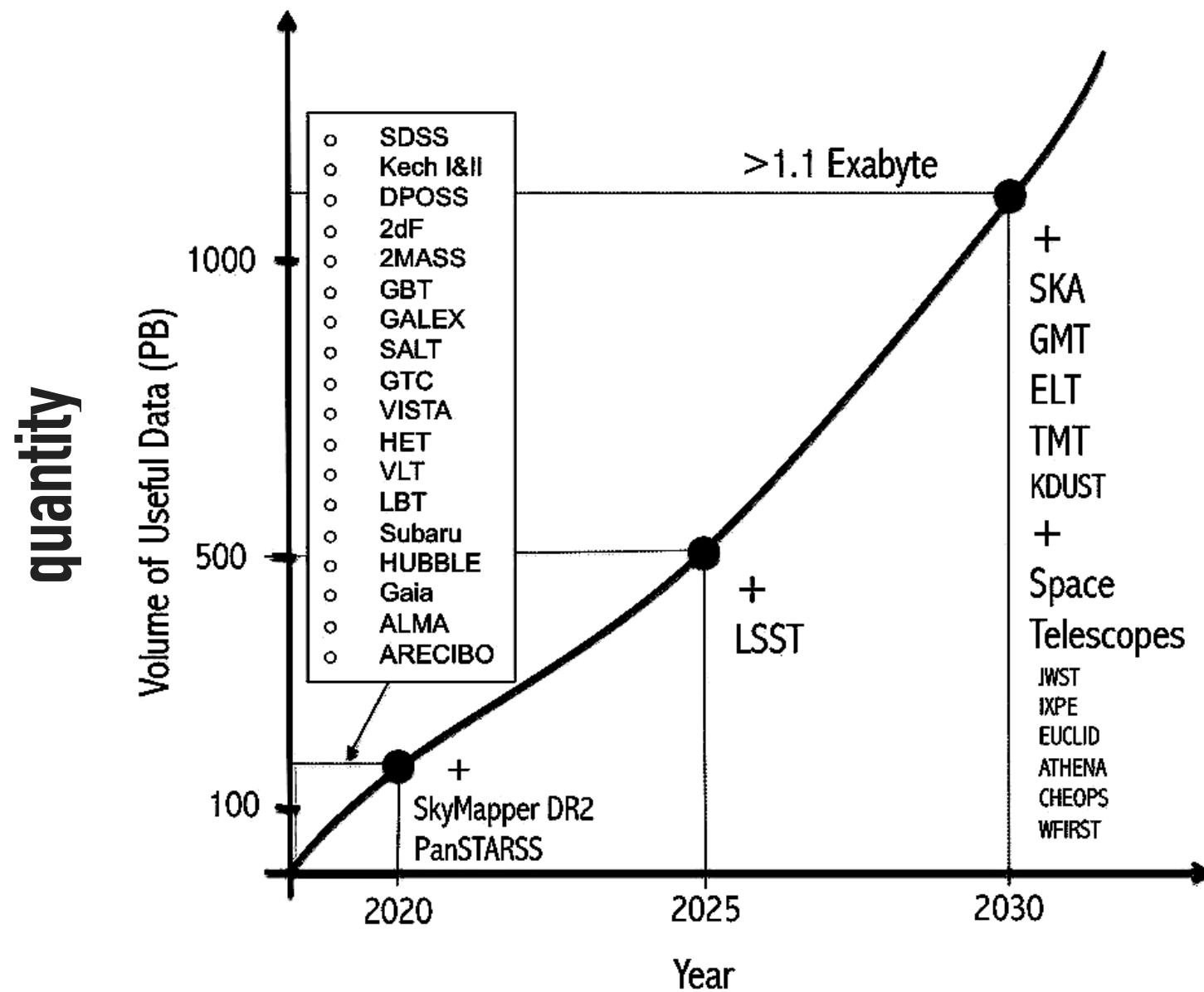
© Rosa

Data Deluge: A New Era !

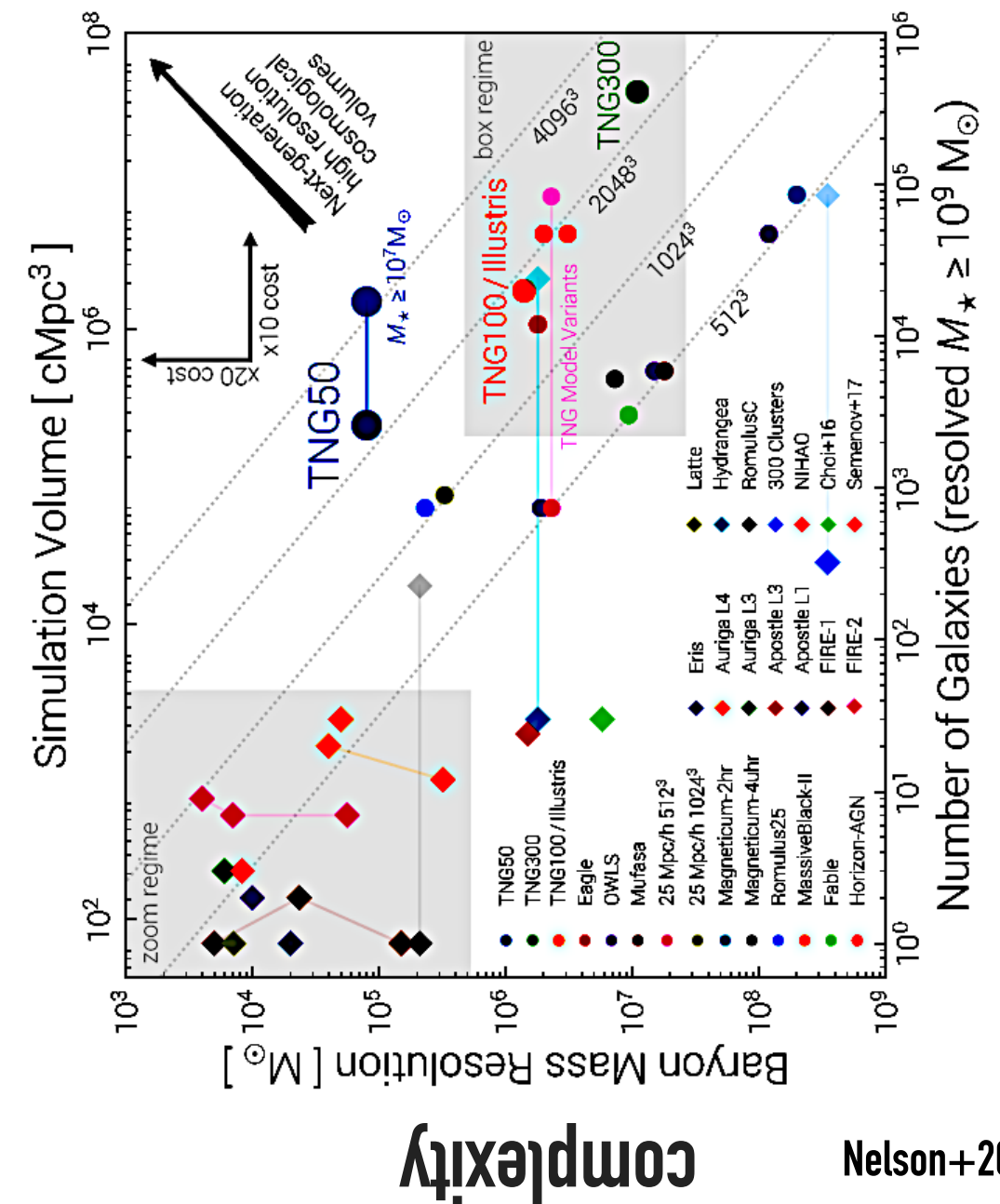
Interpretation

Observational data

Synthetic data

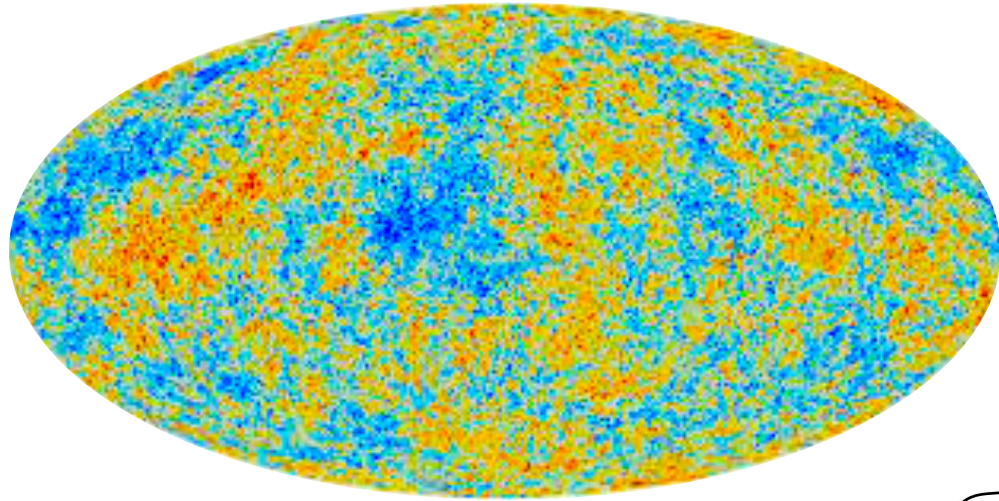


© Rosa

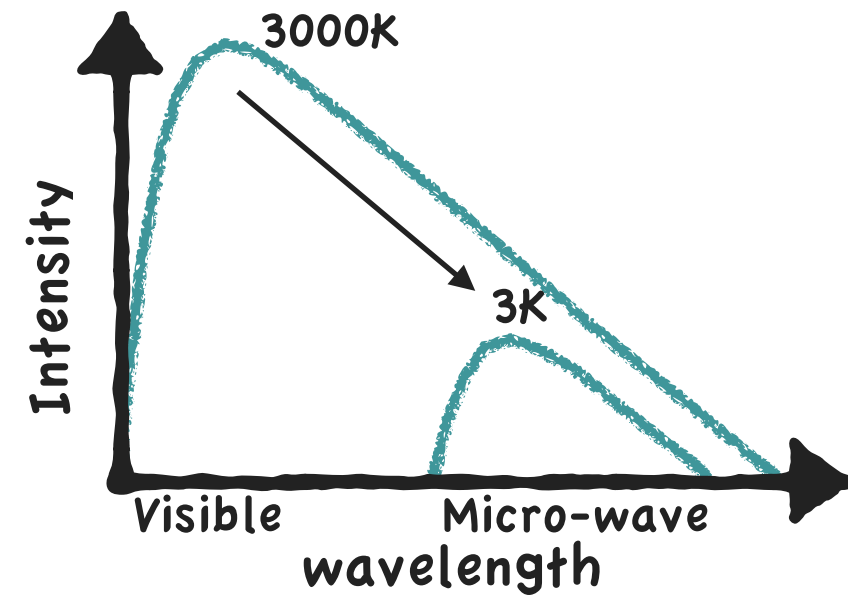


Nelson+2019

Cosmic Microwave Background

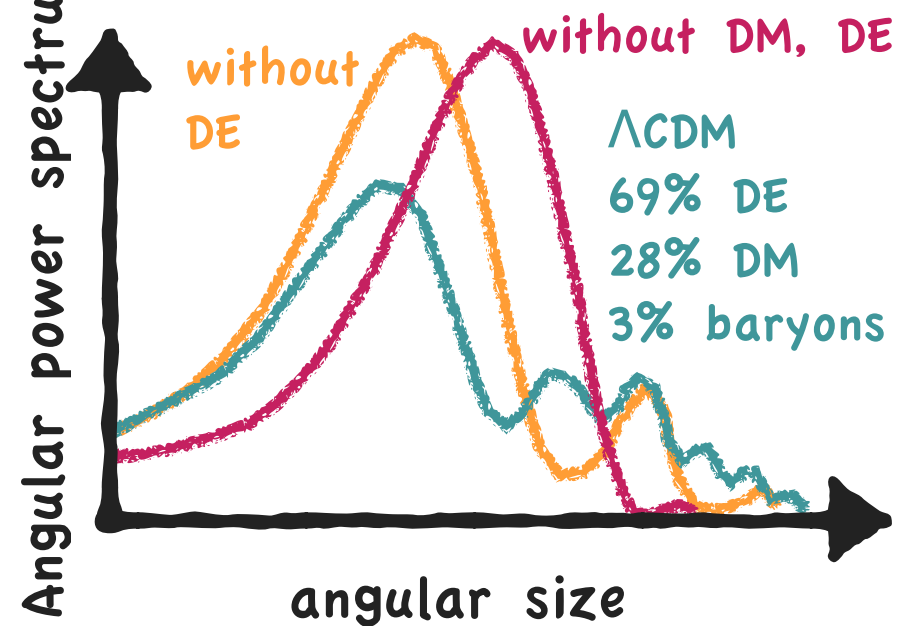


Universe expansion

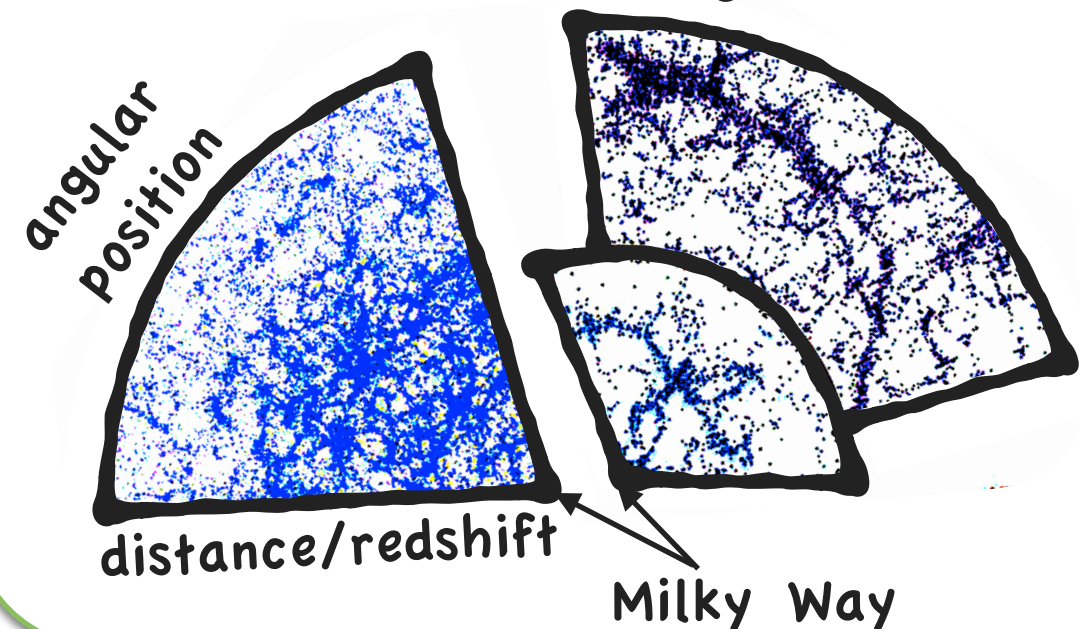


Λ CDM

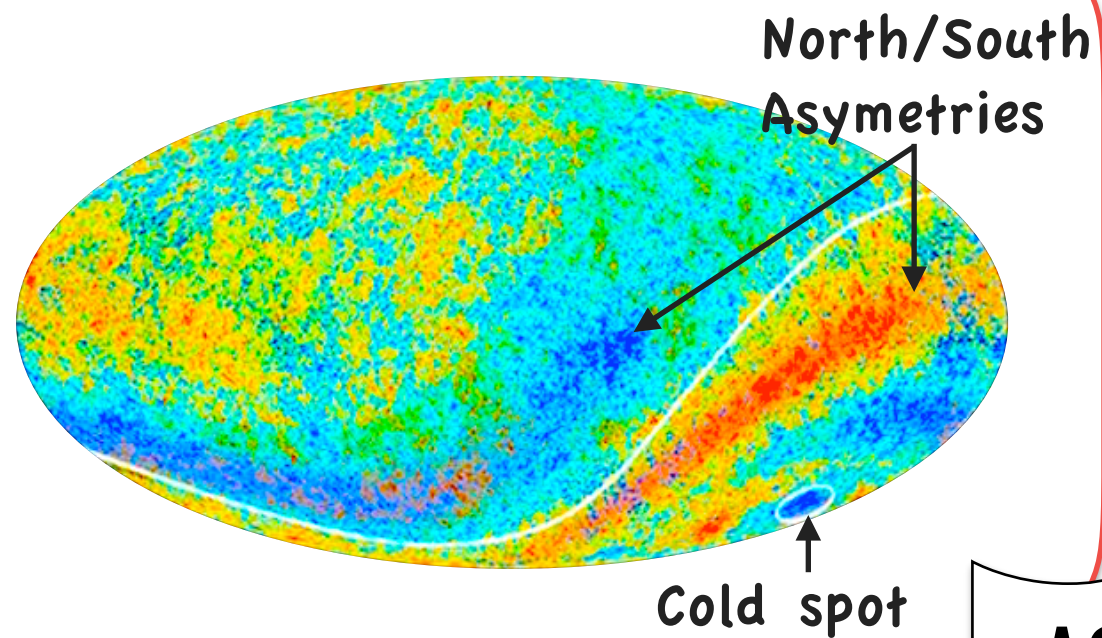
Universe content



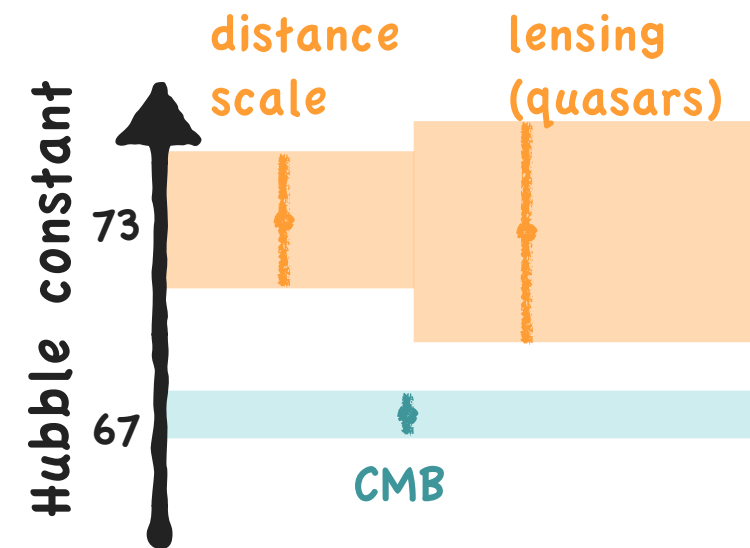
Cosmic Web and galaxies



Anomalies in the CMB

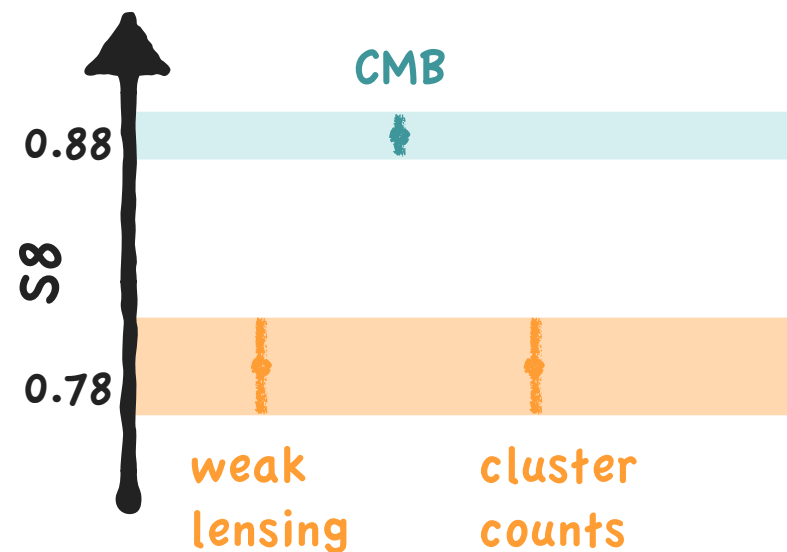


Universe expansion rate (H_0)

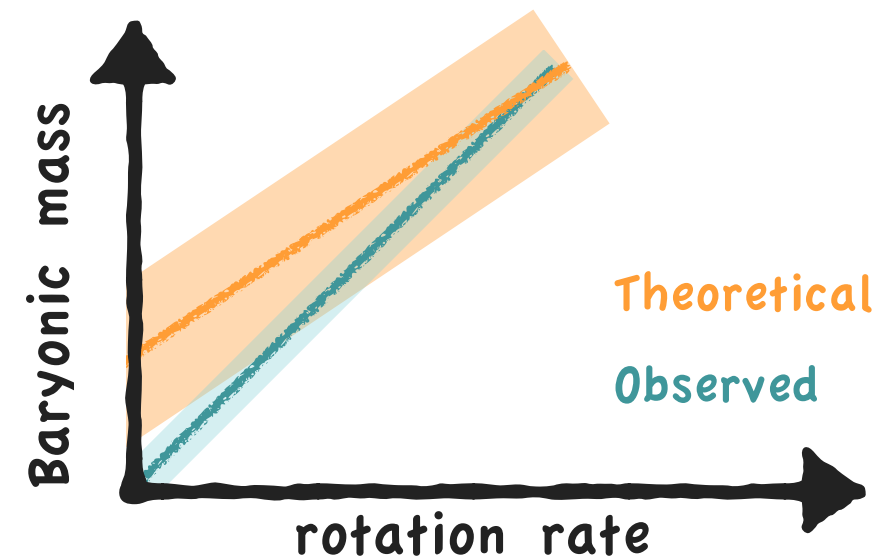


Λ CDM

S8 (σ_8, Ω_m)

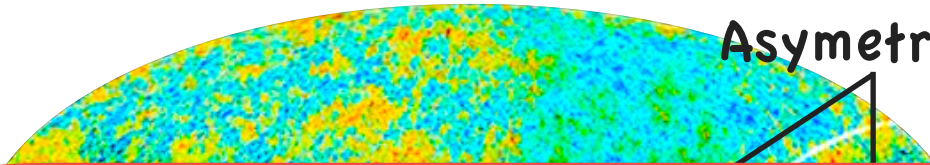


Galaxy properties



Anomalies in the CMB

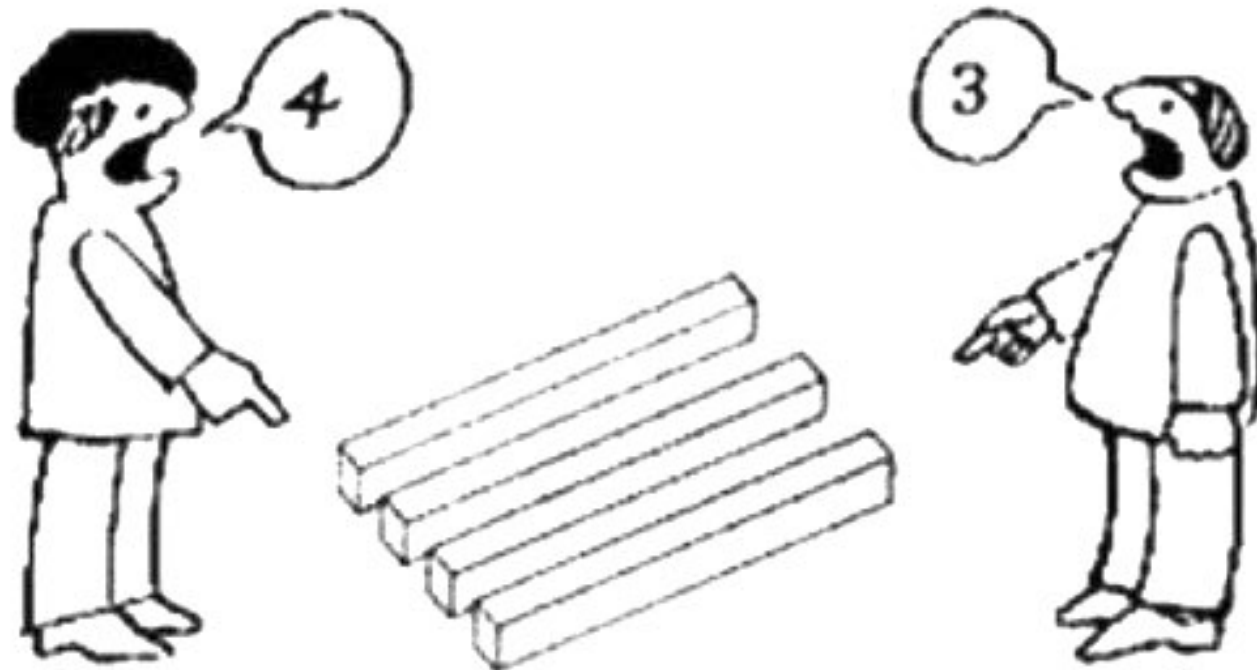
North/South
Asymmetries



Universe expansion rate (H_0)



New physics or biases/systematics?



weak
lensing

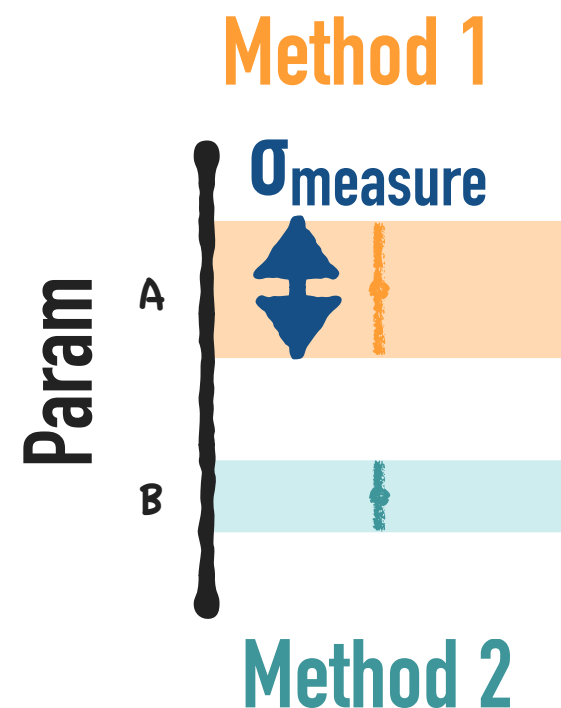
cluster
counts



Biases/Systematics ?

$$\text{Param} = X \pm \sigma_{\text{measure}}$$

- nb measurements
- instruments/tools sensitivity
= precision

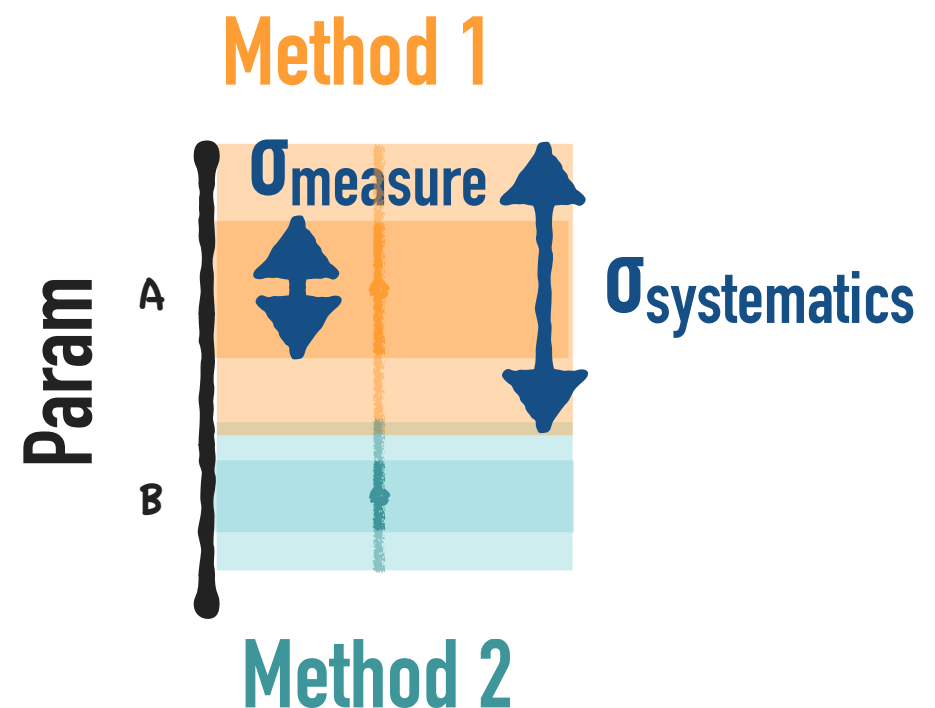


Biases/Systematics ?

$$\text{Param} = X \pm \sigma_{\text{measure}} \pm \sigma_{\text{systematics}}$$

- nb measurements
- instruments/tools sensitivity
= precision

Standard simulations give total uncertainty (cosmic variance)
but tensions vs. systematics?

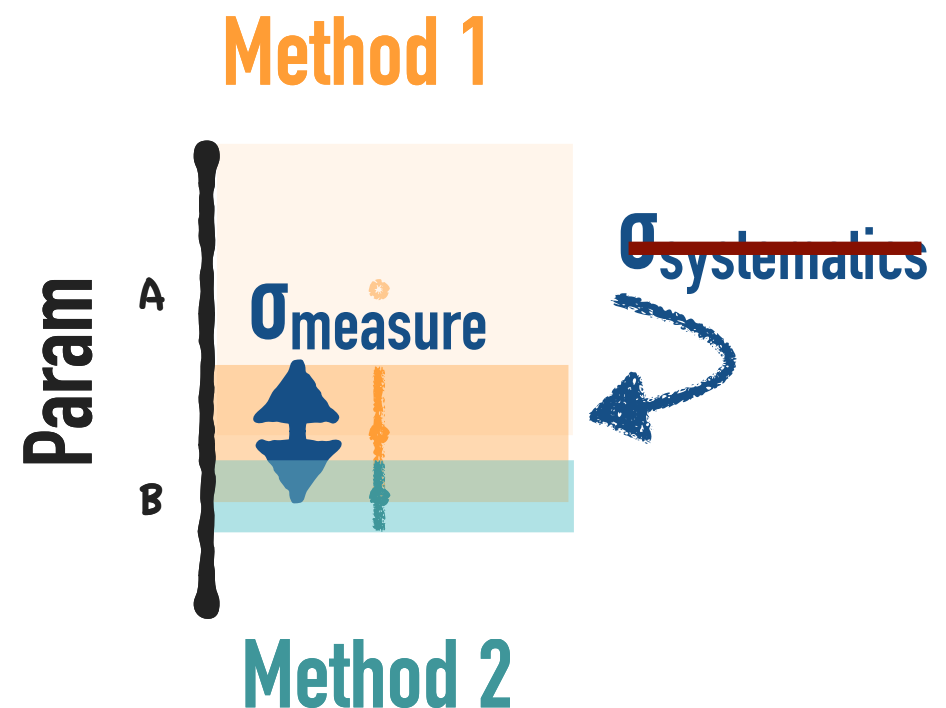


Biases/Systematics ?

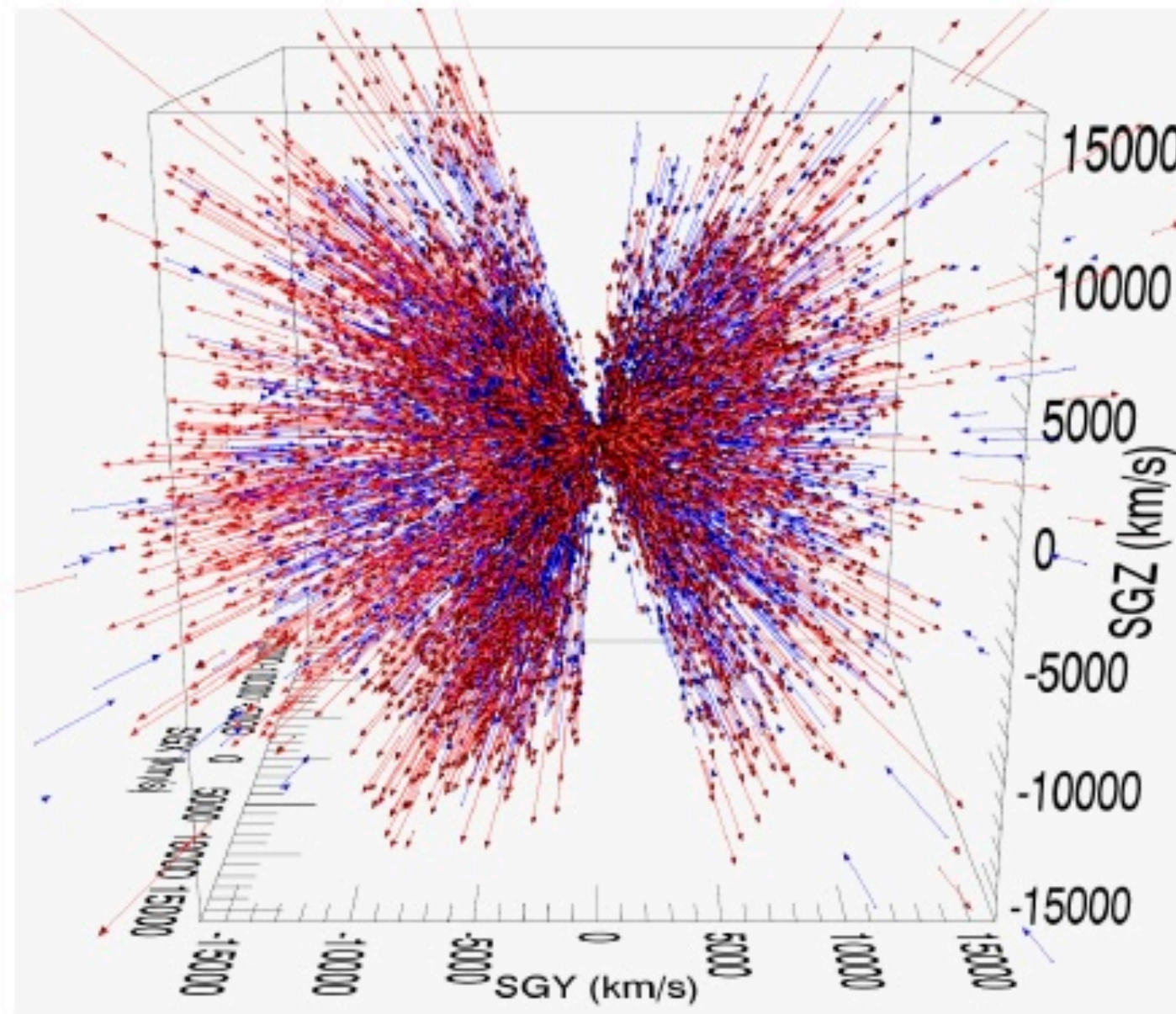
$$\text{Param} = X \pm \sigma_{\text{measure}} \pm \sigma_{\text{systematics}}$$

- nb measurements
- instruments/tools sensitivity
= precision

Digital twins or Constrained Cosmological simulations remove systematics (apple-to-apple)



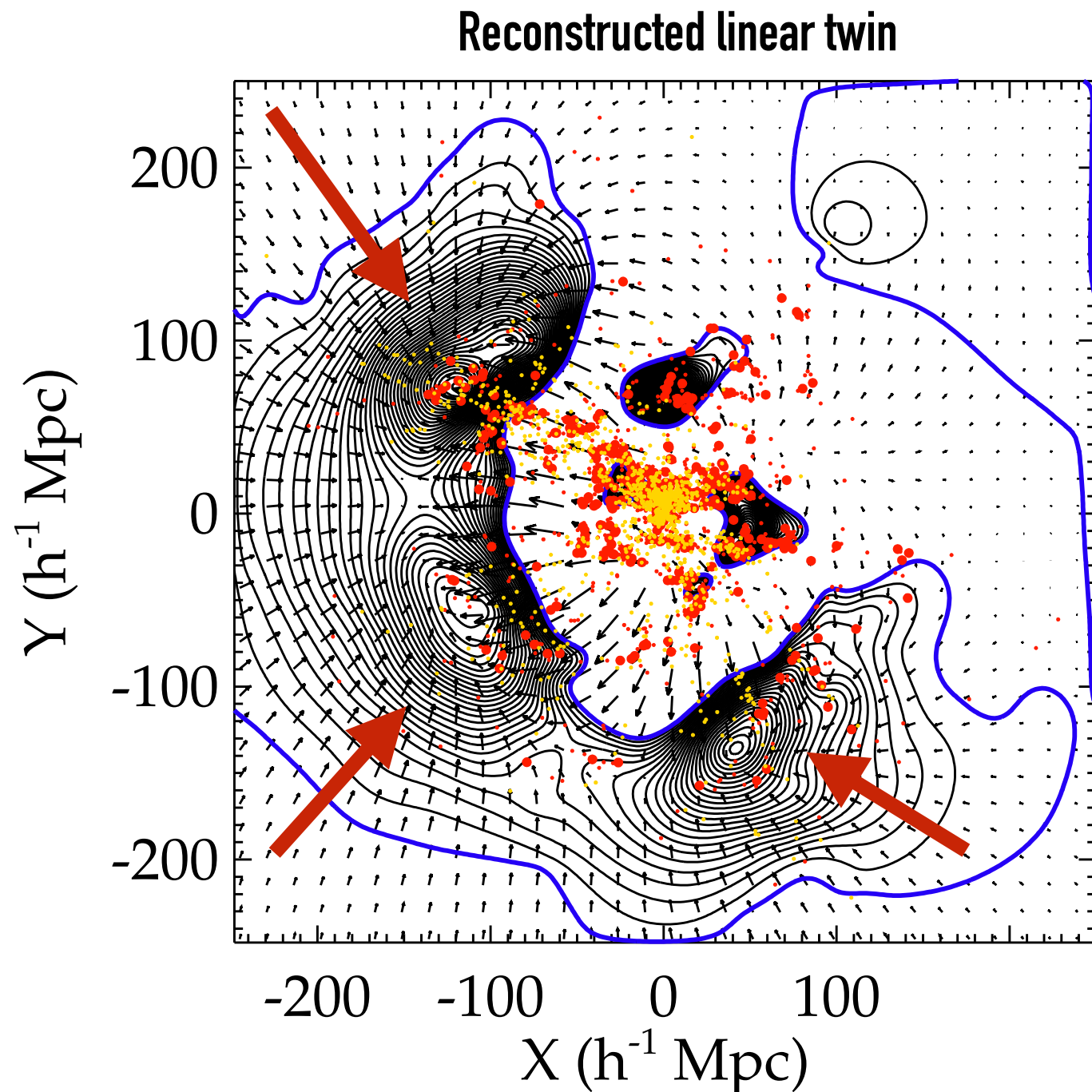
Example of the Hubble constant



Tully+2016

Let's go back to peculiar velocity fields to get twins

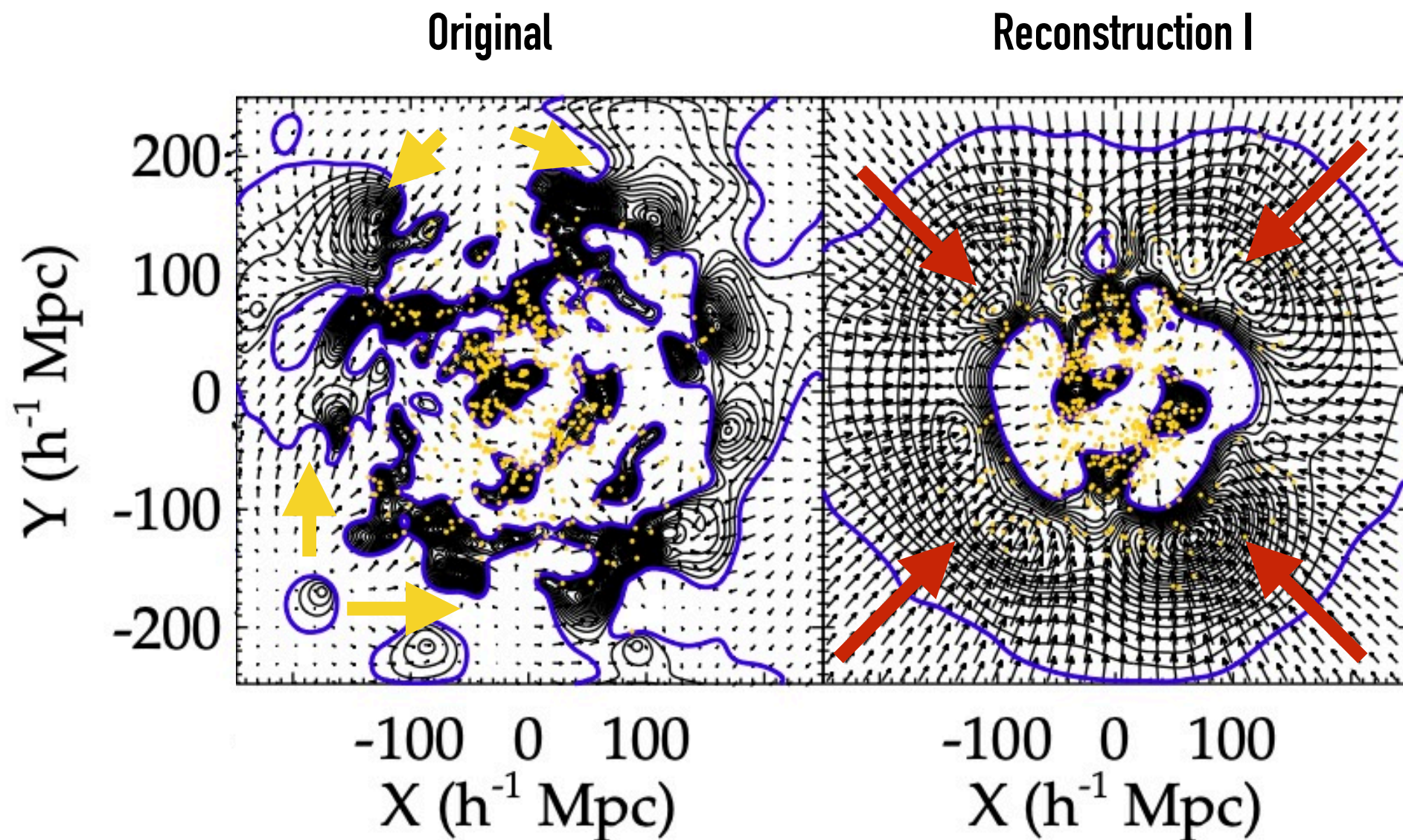
Peculiar velocity fields: Linear twin



Reconstruction $\rightarrow$ Infall $\rightarrow$ Physics?

Sorce, Stoica, Tempel 2023

Peculiar velocity fields: Linear twin



Tests with synthetic catalogs $\rightarrow$ looks like systematics!

Sorce, Stoica, Tempel 2023

Which systematics?

What do we start with? distance moduli $\rightarrow$ distances $\rightarrow$ velocities

$$1. \quad \mu = m - M = 5 \log_{10}(d_{lum} \text{ (Mpc)}) + 25$$

$$2. \quad d_{lum} = (1 + z_{cos}) \int_0^{z_{cos}} \frac{\overset{\text{redshift}}{c_l dz}}{\underset{\text{cosmological parameters}}{H_0 \sqrt{(1+z)^3 \Omega_m + \Omega_\Lambda}}}$$

$$3. \quad v_{pec} = c_l \frac{z_{obs} - z_{cos}}{1 + z_{cos}}$$

**all this observing in
spherical shells**

Which systematics?

What do we start with? distance moduli $\rightarrow$ distances $\rightarrow$ velocities

**1. LogNormal:
Asymmetric error !**

$$1. \quad \mu = m - M = 5 \log_{10}(d_{lum} \text{ (Mpc)}) + 25$$

$$2. \quad d_{lum} = (1 + z_{cos}) \int_0^{z_{cos}} \frac{\overset{\text{redshift}}{c_l dz}}{\underset{\text{cosmological parameters}}{H_0 \sqrt{(1+z)^3 \Omega_m + \Omega_\Lambda}}} \quad \overset{\text{speed of light}}{c_l}$$

$$3. \quad v_{pec} = c_l \frac{z_{obs} - z_{cos}}{1 + z_{cos}}$$

**all this observing in
spherical shells**

Which systematics?

What do we start with? distance moduli $\rightarrow$ distances $\rightarrow$ velocities

$$1. \quad \mu = m - M = 5 \log_{10}(d_{lum} \text{ (Mpc)}) + 25$$

**1. LogNormal:
Asymmetric error !**

$$2. \quad d_{lum} = (1 + z_{cos}) \int_0^{z_{cos}} \frac{\overset{\text{redshift}}{c_l dz}}{\underset{\text{cosmological parameters}}{H_0 \sqrt{(1+z)^3 \Omega_m + \Omega_\Lambda}}}$$

$$3. \quad v_{pec} = c_l \frac{z_{obs} - z_{cos}}{1 + z_{cos}}$$

all this observing in
spherical shells

**2. (In)homogeneous
Malmquist biases:
Higher probability of scatter
closer, from high to low
densities!**

Point processes & Bayesian inference

$$p(\{\mu, \sigma\} | \{d\}, c) \propto \exp[-U_D(\{\mu, \sigma\} | \{d\}, c) - U_I(\{\mu, \sigma\} | \{d\}, c)]$$

Data term

Gaussian uncertainty on the distance modulus!

Preserve realistic velocity

Interaction term

Anti-repulsion

Velocity variance minimization

$$\{\tilde{\mu}, \tilde{\sigma}\} = \arg \min \{ U_D(\{\mu, \sigma\} | \{d\}, c) + U_I(\{\mu, \sigma\} | \{d\}, c) - \log p(c) \}$$

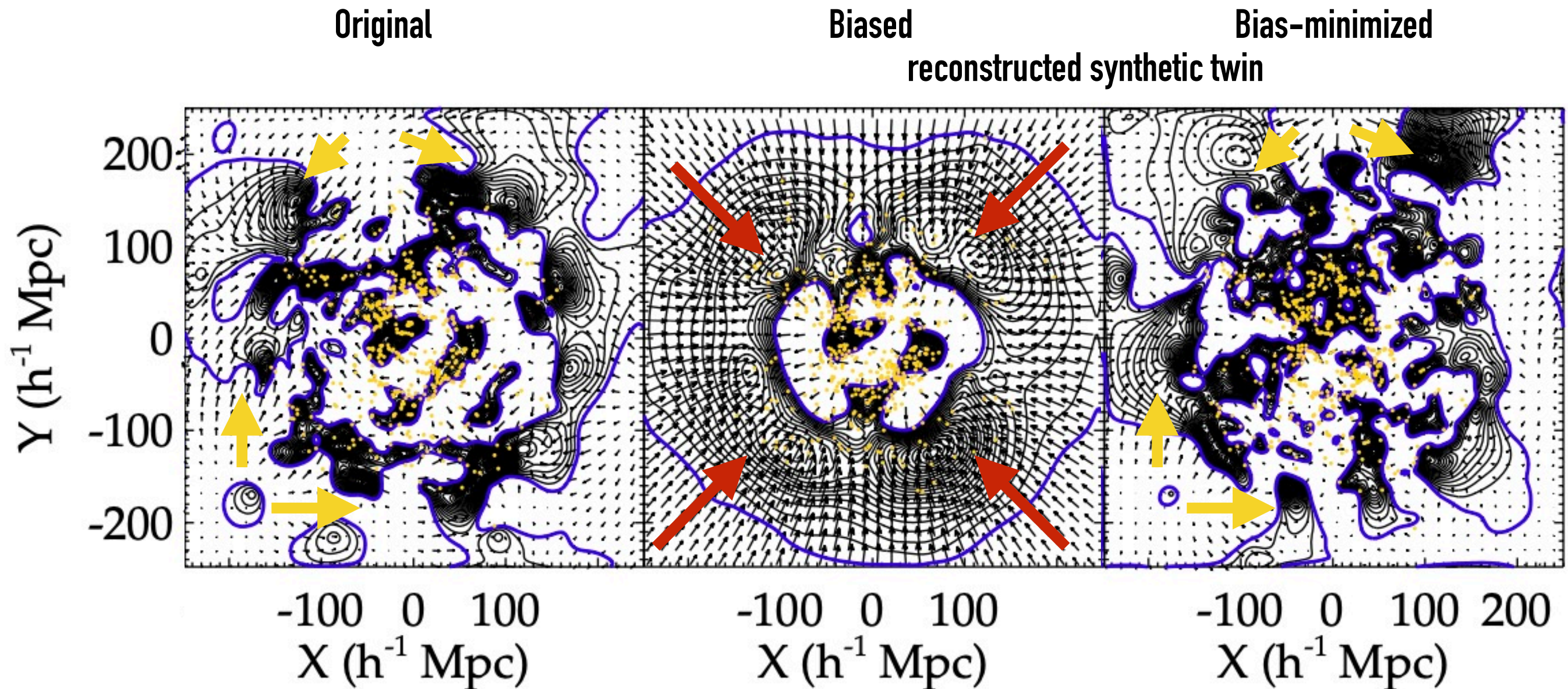
Gibbs sampling $(\tilde{\mu}_i, \tilde{\sigma}_i)$

MCMC $\sum_i (\tilde{\mu}_i, \tilde{\sigma}_i)$

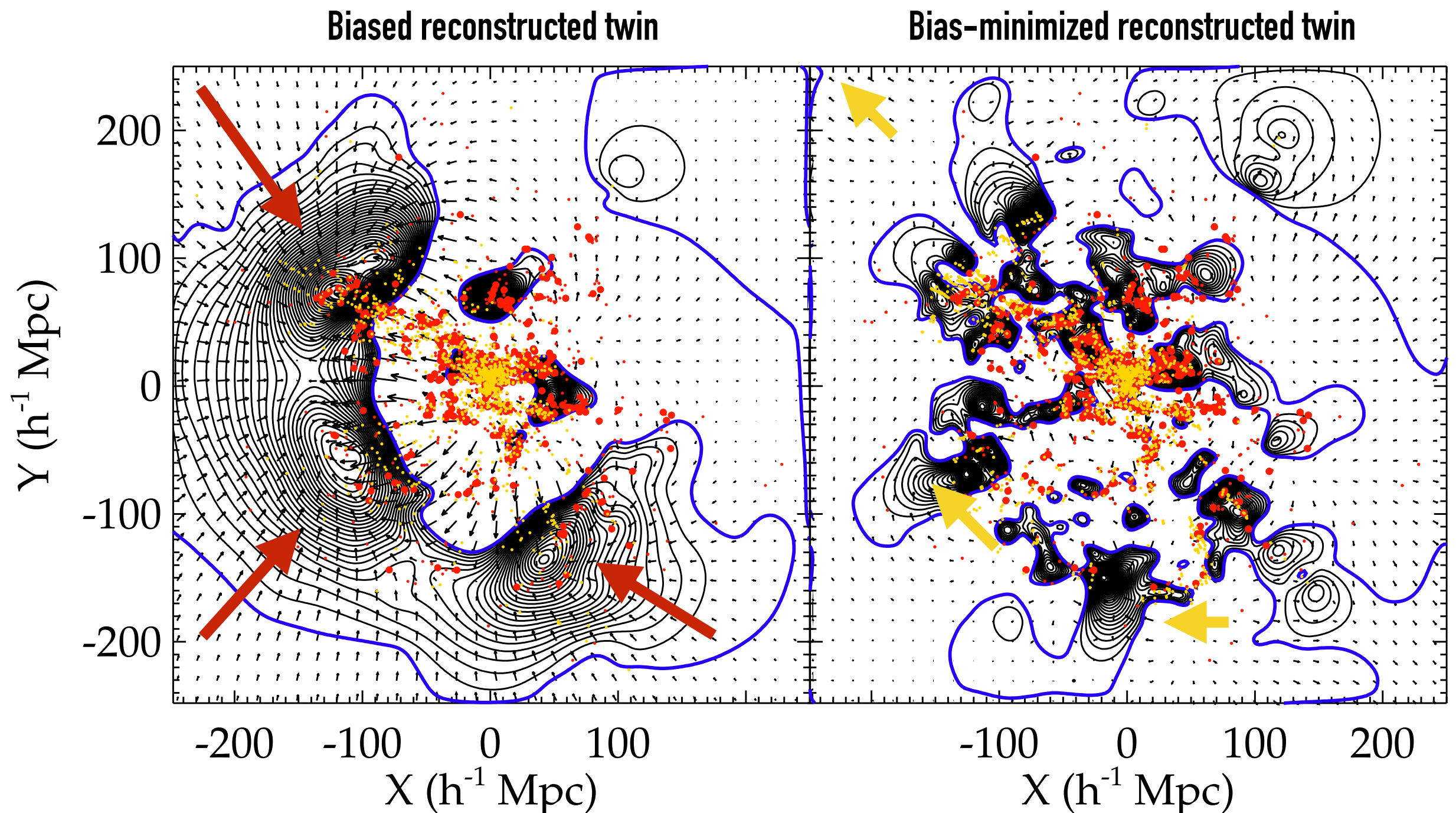
Simulated annealing (several local extrema) $\{\tilde{\mu}, \tilde{\sigma}\}$

Parameters of the model

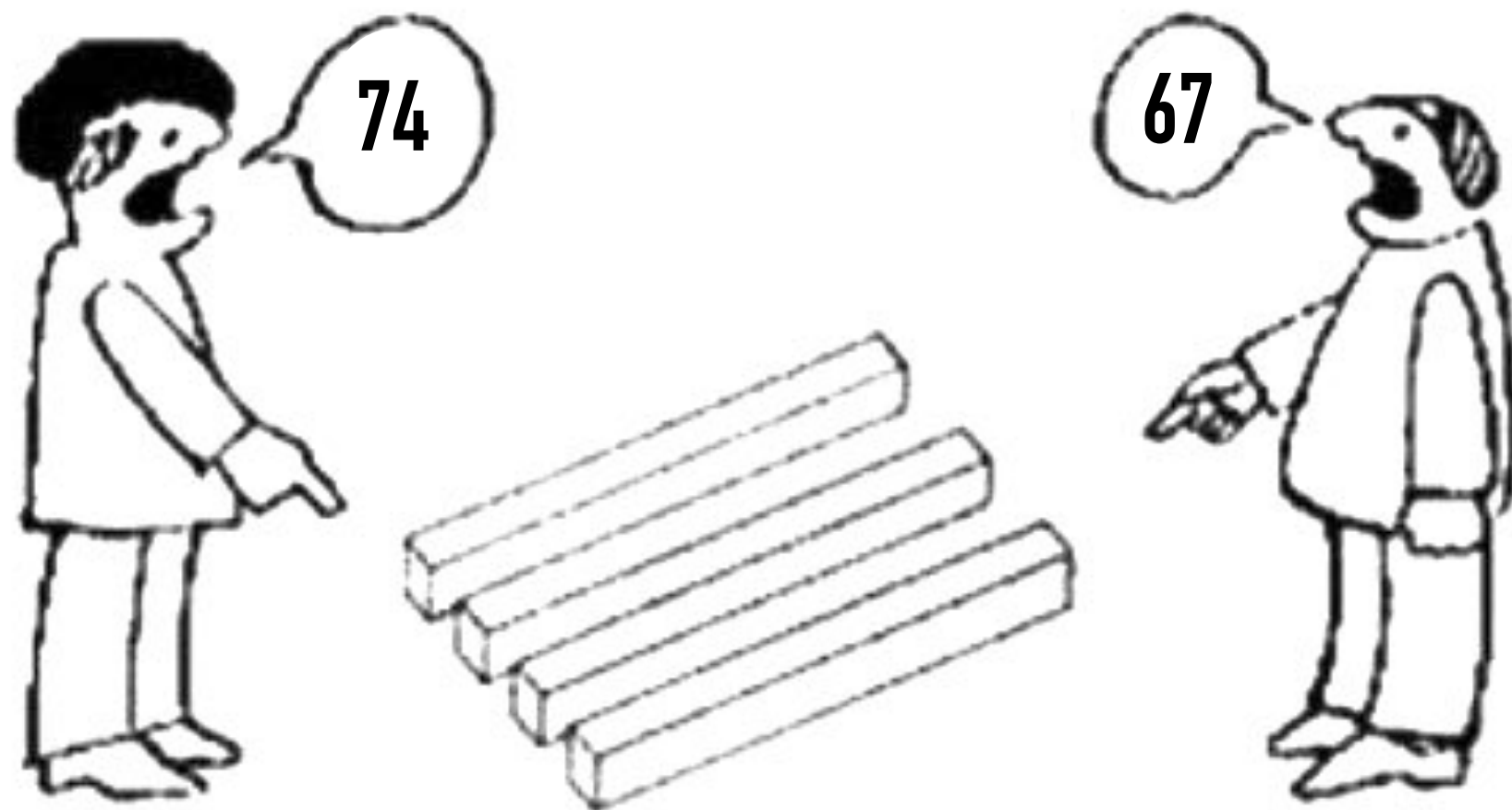
Calibrated with synthetic catalogs



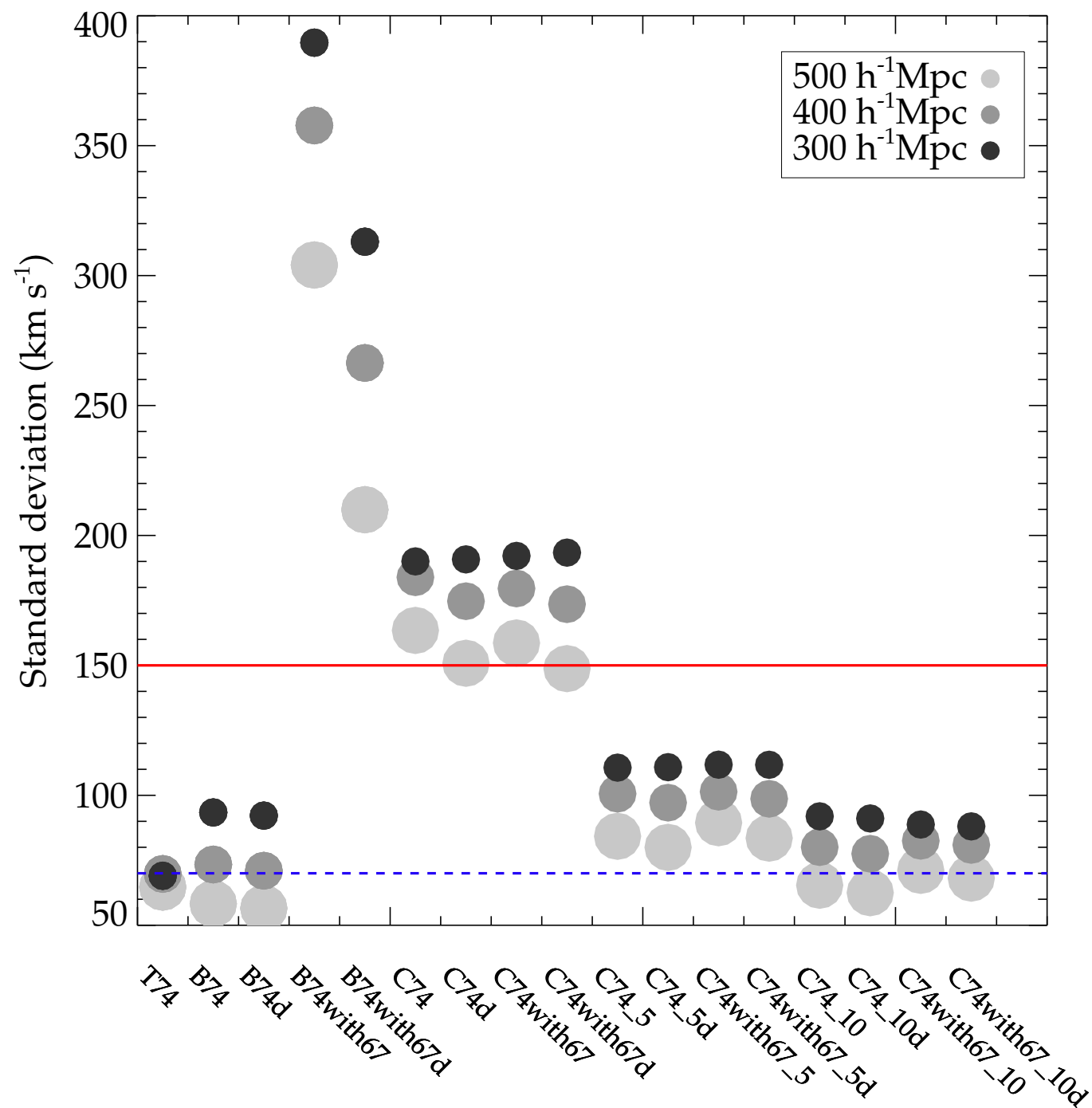
Application to synthetic data $\rightarrow$ Effect of biases minimized !



**Application to observational data: Effects of biases minimized
(sharper structures, no global infall) !**



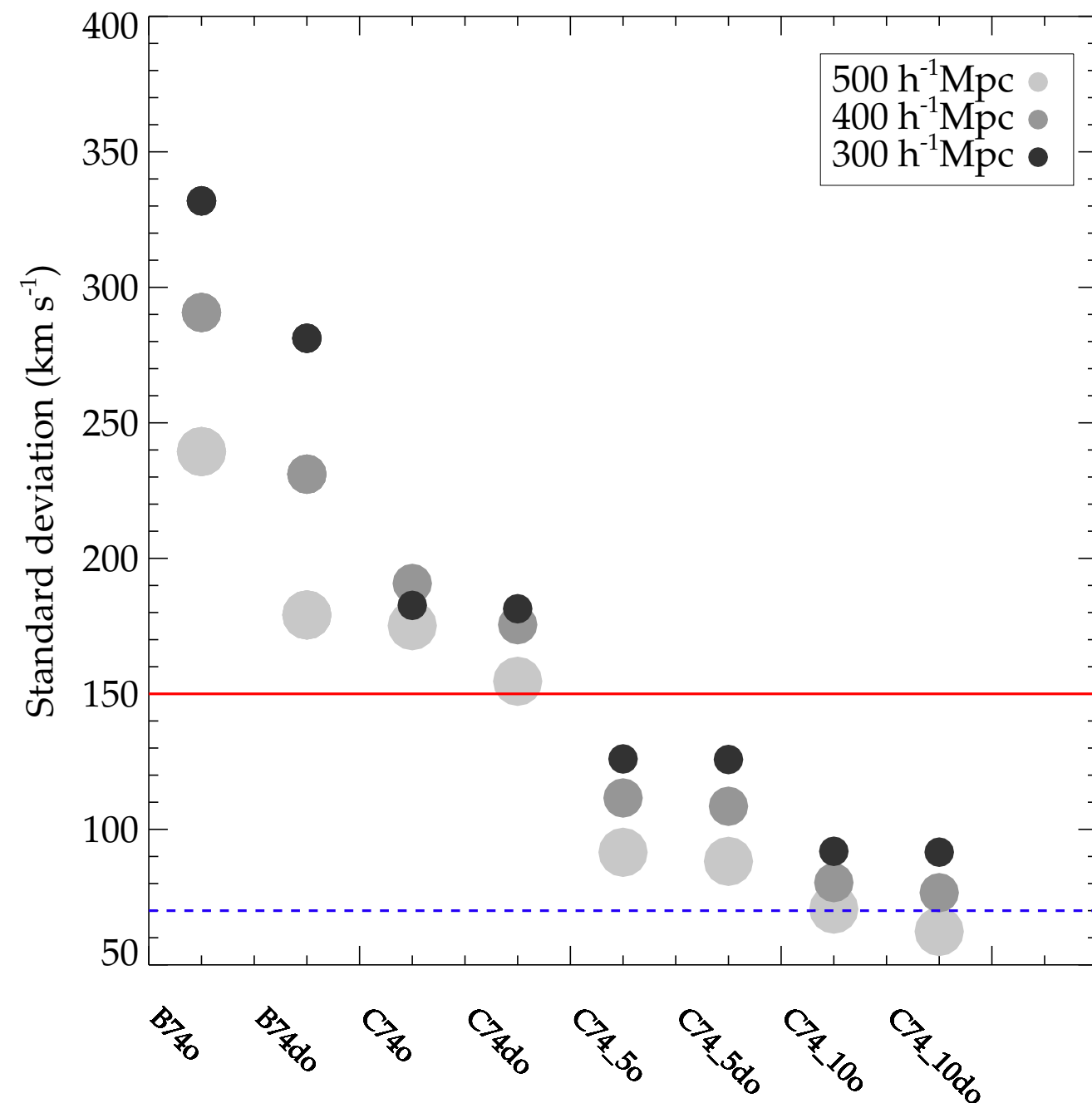
With our reconstructed twins



Synthetic data

Compare reconstructed fields
obtained with $H_0=67$ to
those obtained with 74 km/s/Mpc ,
and those assuming 74 while the
true value is 67 km/s/Mpc .

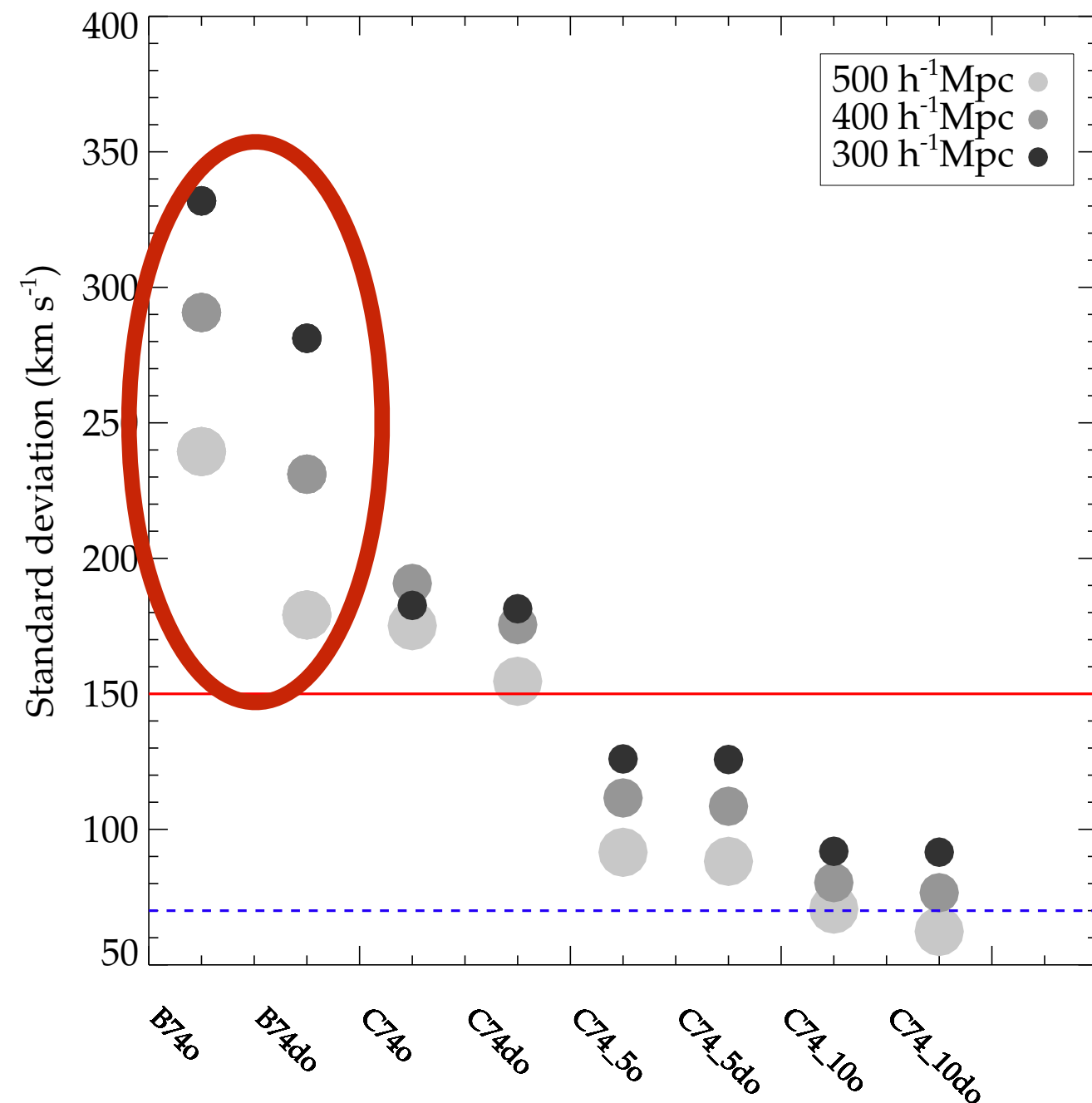
Compare reconstructed fields
obtained with $H_0=74$ and
 67km/s/Mpc , not knowing the
true value.



Observational data

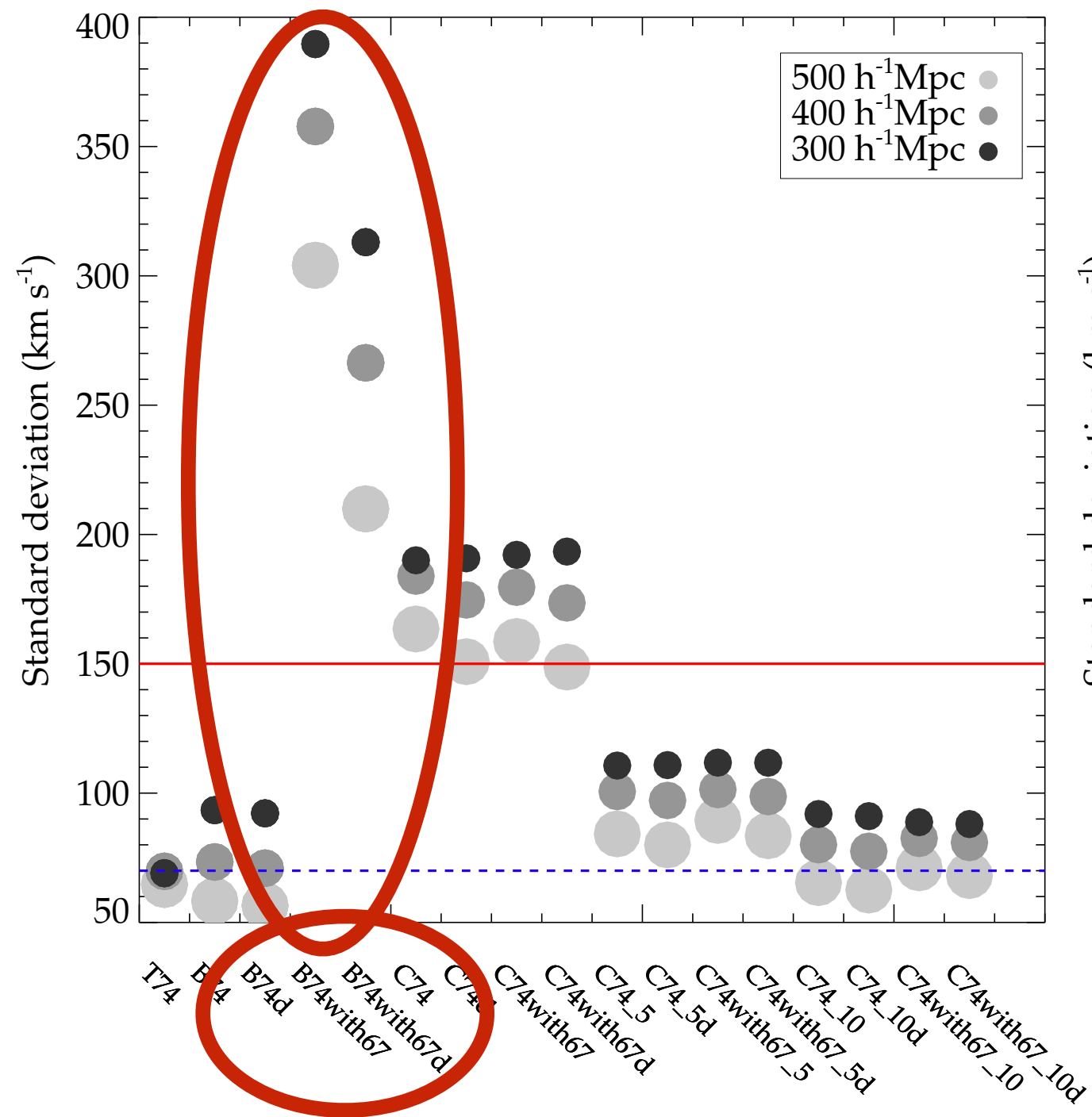
Sorce, Stoica, Tempel 2023

Compare reconstructed fields
obtained with $H_0=74$ and
 67km/s/Mpc , not knowing the
true value.

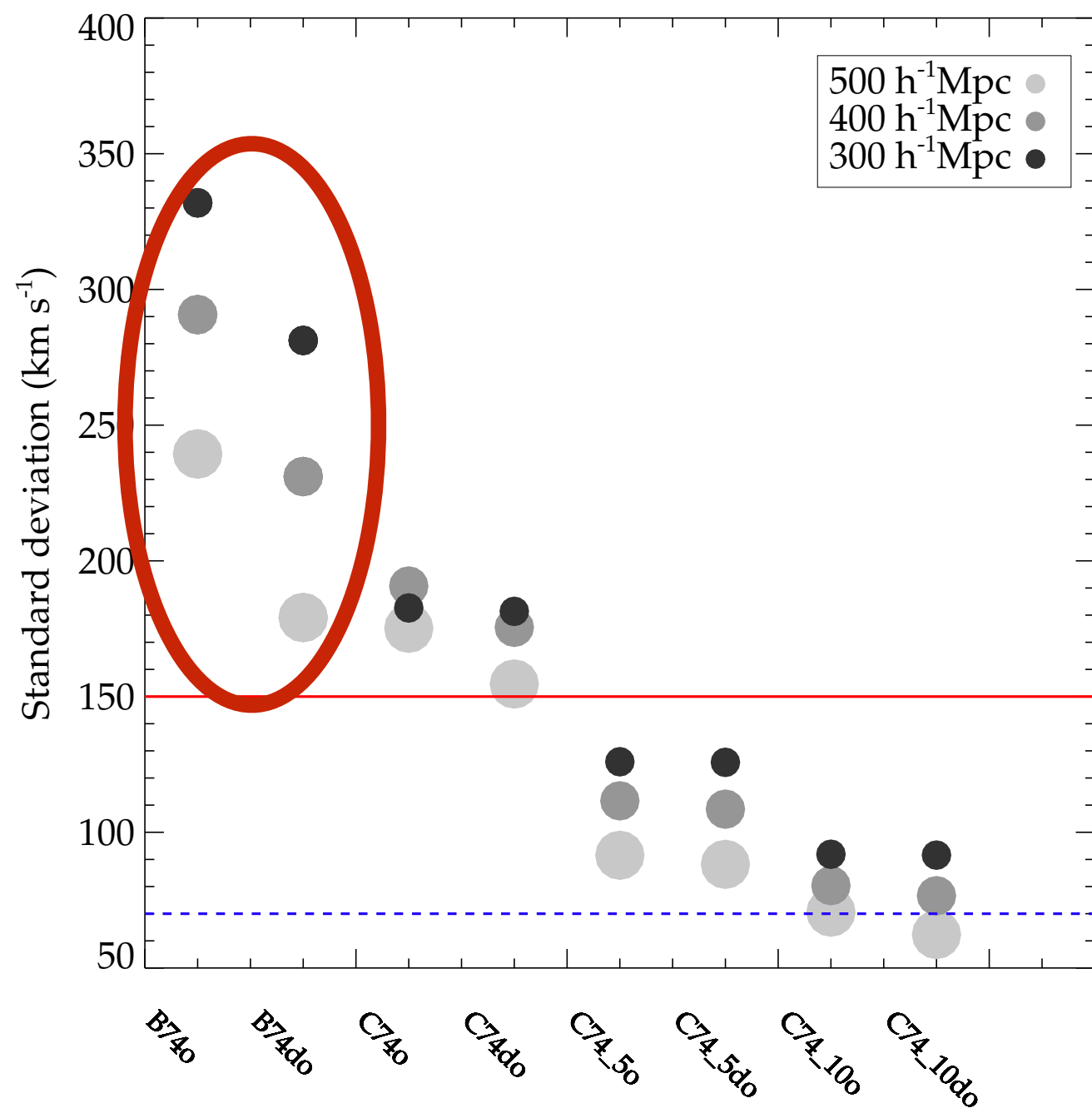


Observational data

Sorce, Stoica, Tempel 2023



Synthetic data

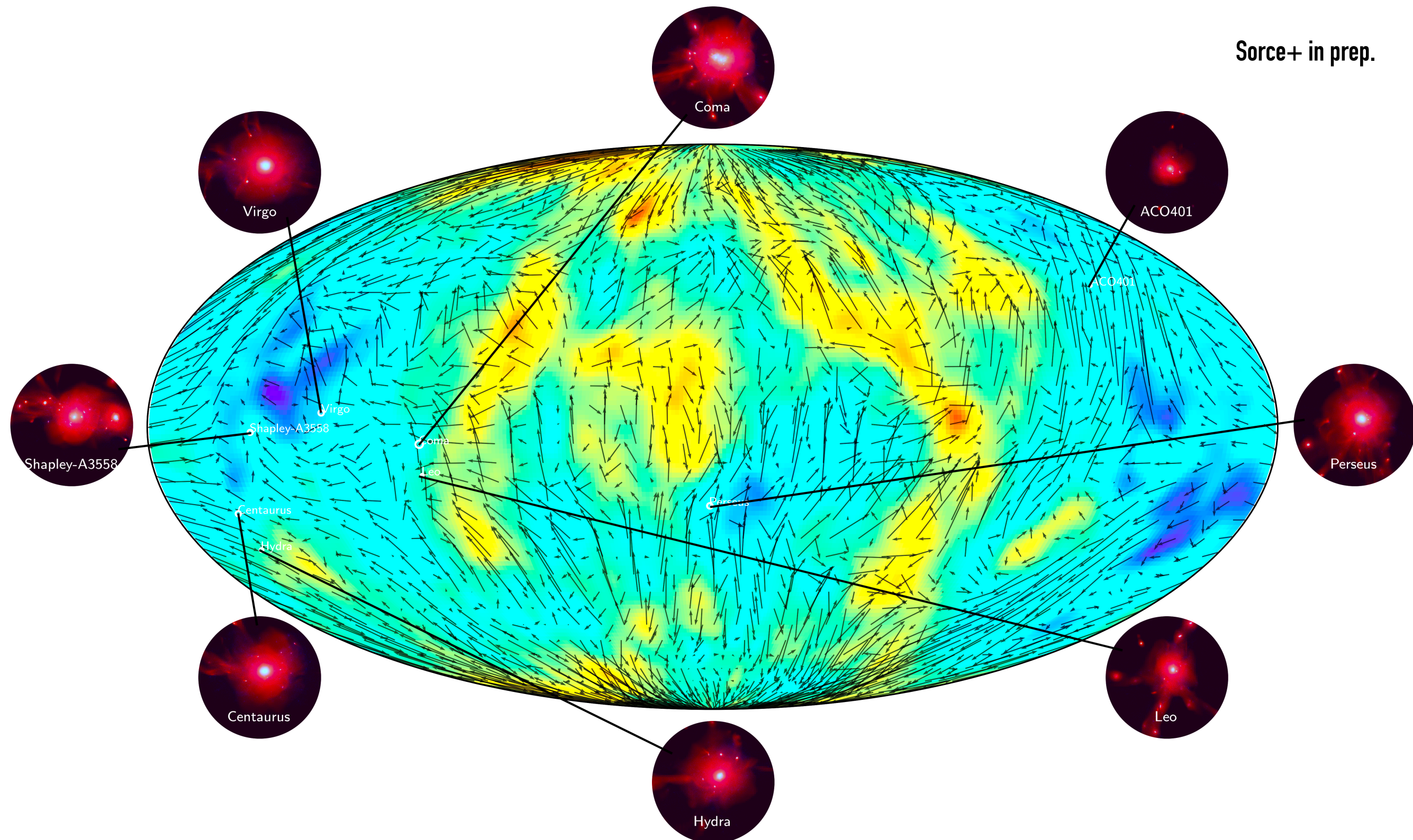


Observational data

Sorce, Stoica, Tempel 2023

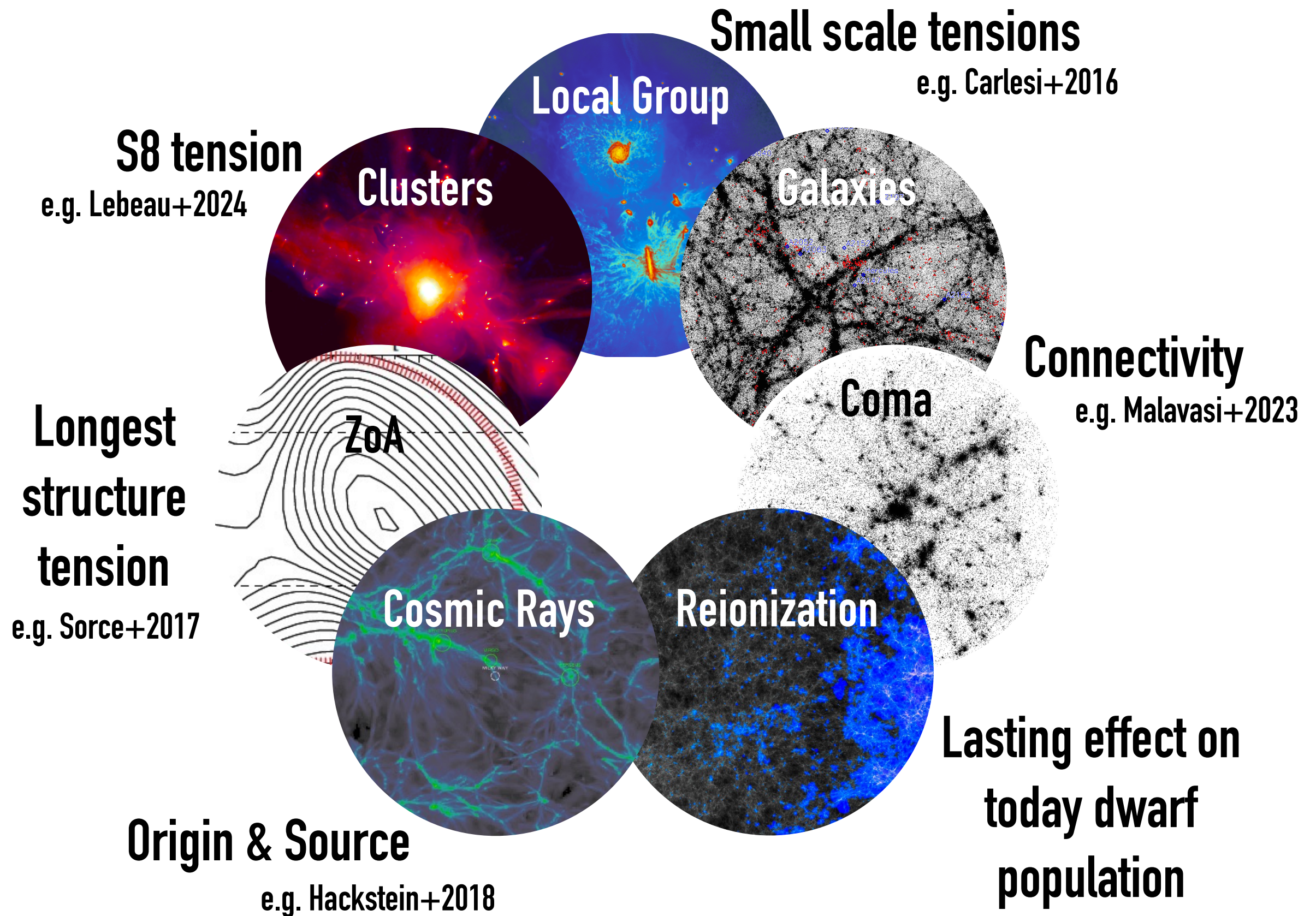
Beyond reconstructed twin: Simulated twin

Sorce+ in prep.



CLONE: cosmological hydrodynamical twin to be released stay tuned

CLONES: Multiple projects with twins



Warning: new physics or systematics? observations = lots of systematics

Bias minimization before interpretation (H_0 tension?)

CLONES = digital twins for accurate interpretation (remove cosmic variance)

Ongoing/Behind the scene:

Machine learning to speed up the inference processes at the level of

- bias minimization
- reconstruction at $z=0$ and z_{init} ,
- simulation (non-linearity, complexity)

Open Collaboration: All are welcome!

**Thank you, Merci, Grazie,
Gracias, Danke, ευχαριστώ
Mahalo, 谢谢, ありがとう, Дякую,
הודו, спасйбо, Obrigada, Dank u,
Tak, Cảm ơn, Dziękuję, 감사합니다
Kiitos, Aitäh, diolch, dankewol,
ಧನ್ಯವಾದಗಳು, شُكْرًا, ...***

*** Missing your 'thanks' spelling? It means I did not get the chance to learn how to say it so far**

Point processes & Bayesian inference

Data term

$$U_D(\{\mu\} | \{d\}, c_1) = \sum_{i=1}^n c_1 \times \frac{(\mu_i - \mu_{init,i})^2}{2\sigma_{init,i}^2} + \sum_{i=1}^n c_2 \times \frac{|v_{pec,i}|}{v_{ref}}$$

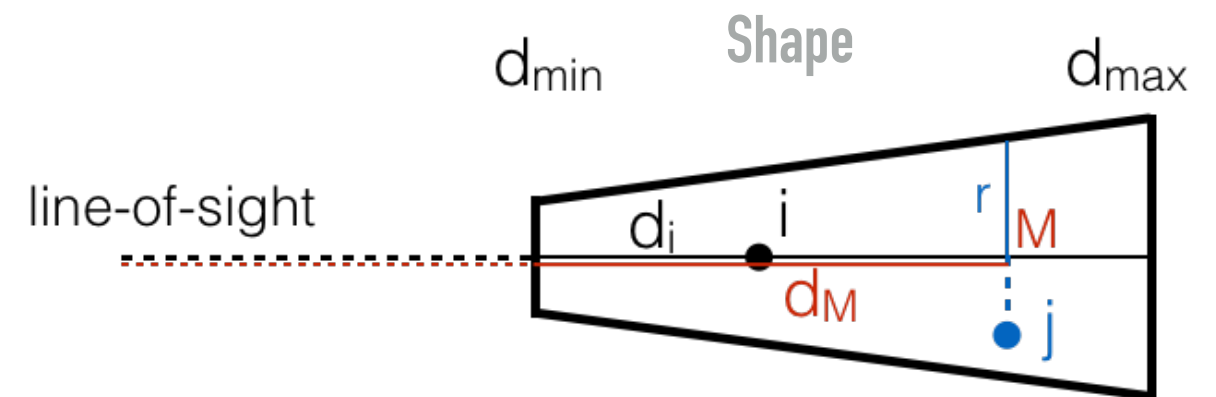
Gaussian uncertainty on
the distance modulus!

preserve realistic
velocity

Point processes & Bayesian inference

Interaction term

$$U_I(\mu \mid \mathbf{d}, c_3) = \sum_{i=1}^n c_3 (h_i/f_i + q_i)$$



all interactions

$$f_i = \sum_{j=1, j \neq i}^n \mathbf{1}\{i \sim_s j\} \times 1/\sigma_j^2$$

anti-repulsion

$$q_i = \mathbf{1}\{\forall j, j \neq i, i \not\sim_s j\}$$

velocity variance minimization

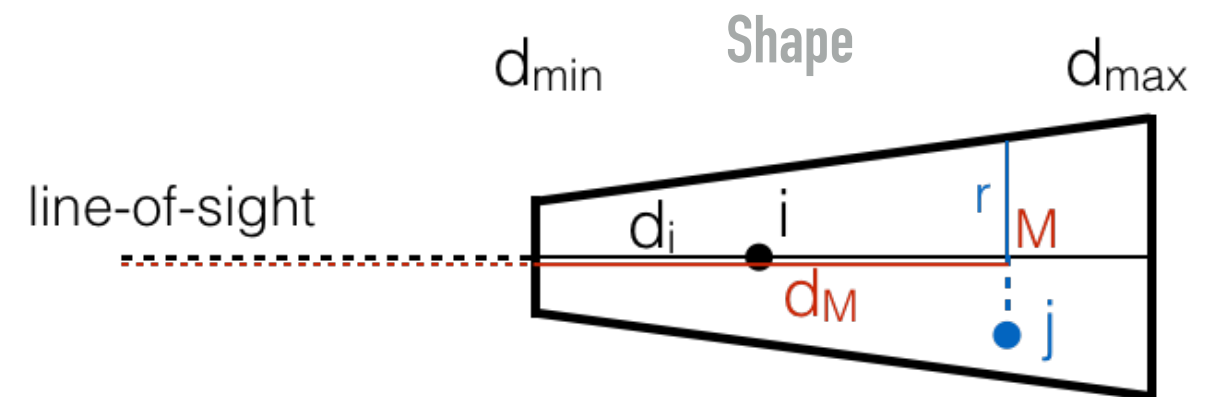
$$h_i = \sum_{j=1, j \neq i}^n \mathbf{1}\{i \sim_s j\} \times \mathbf{1}\{||v_{pec,i}| - |v_{pec,j}|| > \sigma_v\} \times 1/\sigma_j^2$$

Sorce, Stoica, Tempel 2023

Point processes & Bayesian inference

Interaction term

$$U_I(\mu \mid \mathbf{d}, c_3) = \sum_{i=1}^n c_3 (h_i/f_i + q_i)$$



all interactions

$$f_i = \sum_{j=1, j \neq i}^n \mathbf{1}\{i \sim_s j\} \times 1/\sigma_j^2$$

anti-repulsion

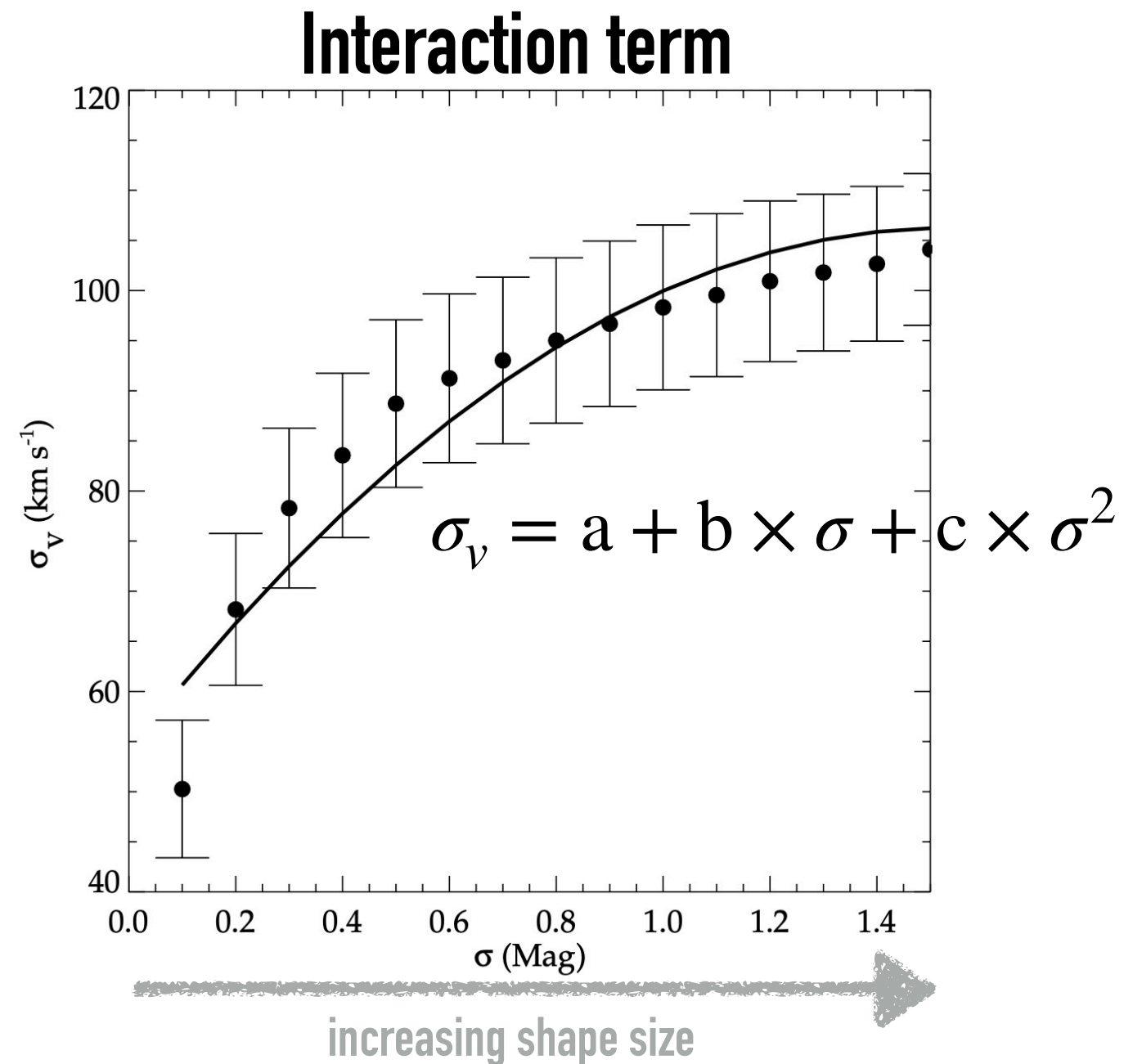
$$q_i = \mathbf{1}\{\forall j, j \neq i, i \not\sim_s j\}$$

velocity variance minimization

$$h_i = \sum_{j=1, j \neq i}^n \mathbf{1}\{i \sim_s j\} \times \mathbf{1}\{|v_{pec,i}| - |v_{pec,j}|| > \sigma_v\} \times 1/\sigma_j^2$$

Sorce, Stoica, Tempel 2023

Point processes & Bayesian inference



Sorce, Stoica, Tempel 2023

Point processes & Bayesian inference

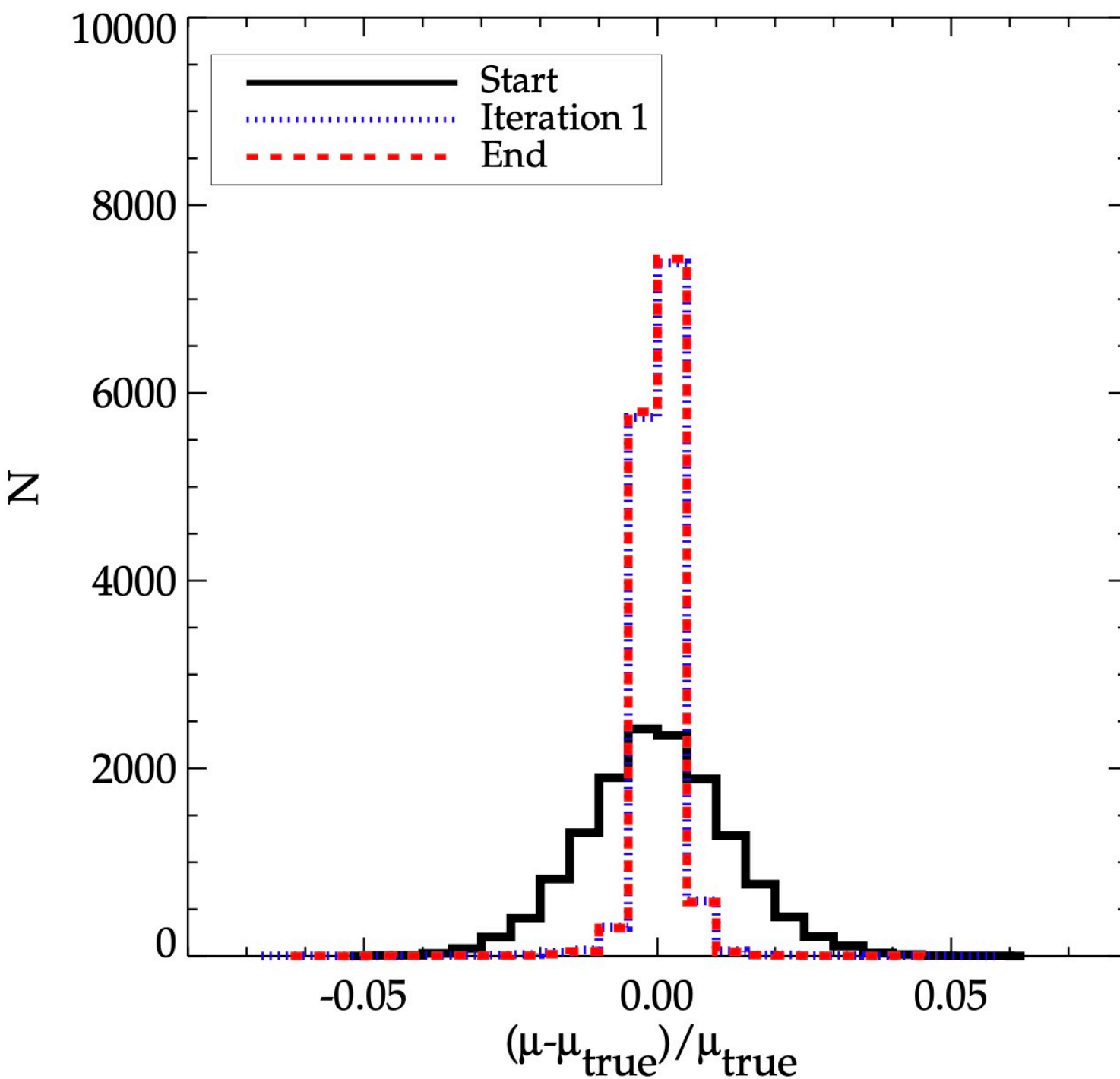
```

1 for  $t = 1 \dots n_{sa}$  do           Simulated annealing (several local extrema)
2    $T(t) = \frac{T_0}{1 + \ln(t)}$ 
3   for  $m = 1 \dots n_{mh}$  do       MCMC
4     for  $i = 1 \dots n$  do       Gibbs sampling
5        $\tilde{\mu}_i = \mu_i + U_n[-0.5, +0.5] \times \gamma \sigma_i$ 
6        $d_{lum,i}, v_{pec,i}, \tilde{d}_{lum,i}, \tilde{v}_{pec,i}$ 
7        $\tilde{\sigma}_i = \sigma_i [ (1 - p_{v,i})^2 + p_{v,i} (\frac{h_i}{f_i} + q_i) ]$  — velocity probability
8        $U_1(\dots, (\mu_i, \sigma_i), \dots), U_2(\dots, (\mu_i, \sigma_i), \dots)$ 
9        $U_1(\dots, (\tilde{\mu}_i, \tilde{\sigma}_i), \dots), U_2(\dots, (\tilde{\mu}_i, \tilde{\sigma}_i), \dots)$ 
10       $p_{\tilde{\mu}_i, \tilde{\sigma}_i}, p_{\mu_i, \sigma_i}$ 
11       $\alpha = \min\{1, (p_{\tilde{\mu}_i, \tilde{\sigma}_i} / p_{\mu_i, \sigma_i})^{1/T(t)} \sigma_i / \tilde{\sigma}_i\}$ 
12       $R = U_n[0, 1[$ 
13      if  $\alpha \geq R$ ,  $\mu_i = \tilde{\mu}_i$  &  $\sigma_i = \tilde{\sigma}_i$ 
14    end
15  end
16 end

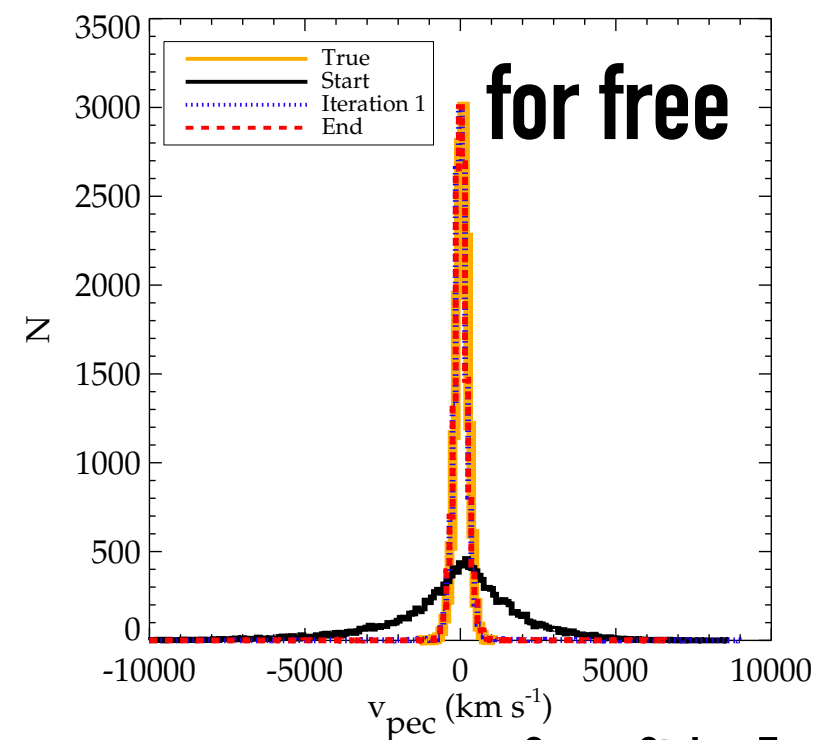
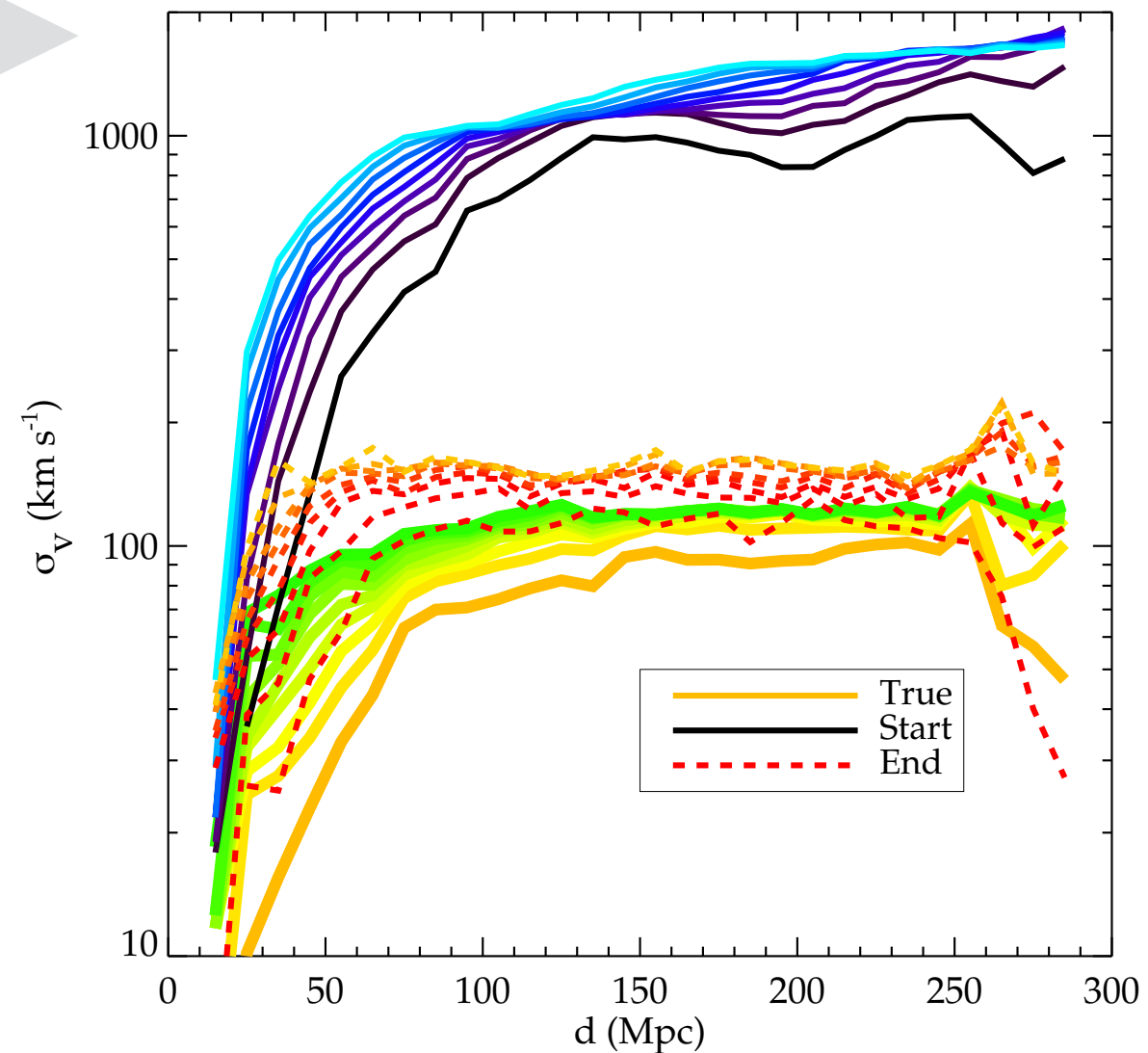
```

Calibrated with synthetic catalogs

Sorce, Stoica, Tempel 2023



Application to synthetic data



for free

Sorce, Stoica, Tempel 2023