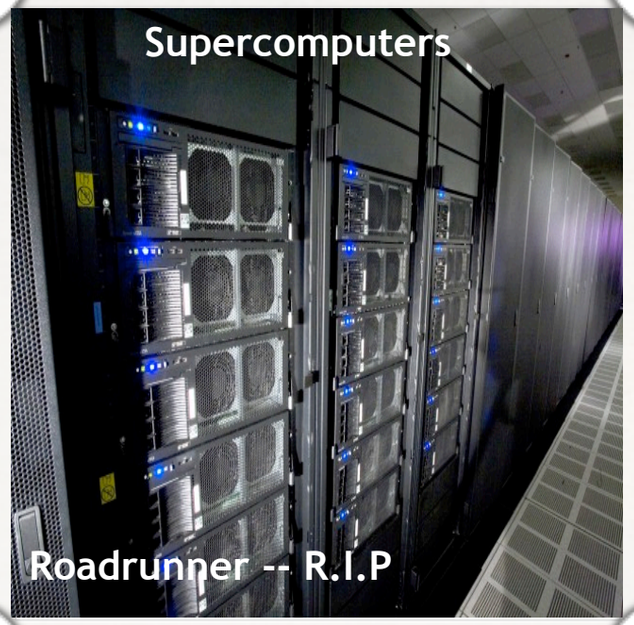
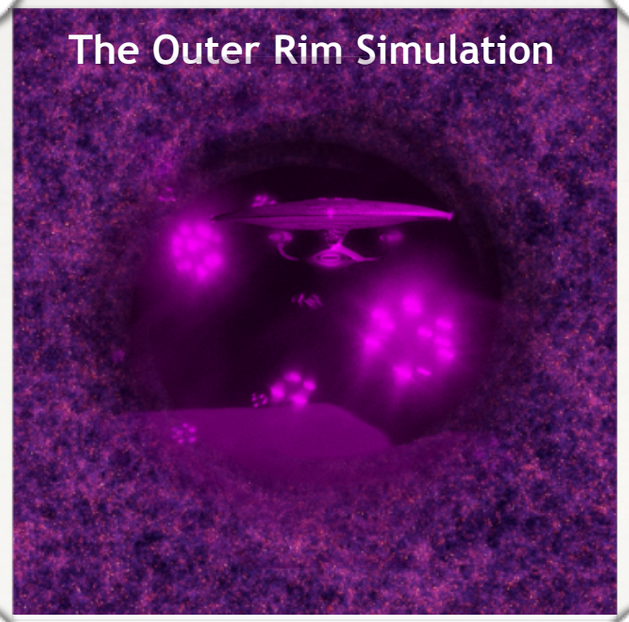
Cosmic Calibration

Katrin Heitmann

in collaboration with Derek Bingham, Salman Habib, David Higdon, Earl Lawrence, Juliana Kwan, Christian Wagner, Martin White

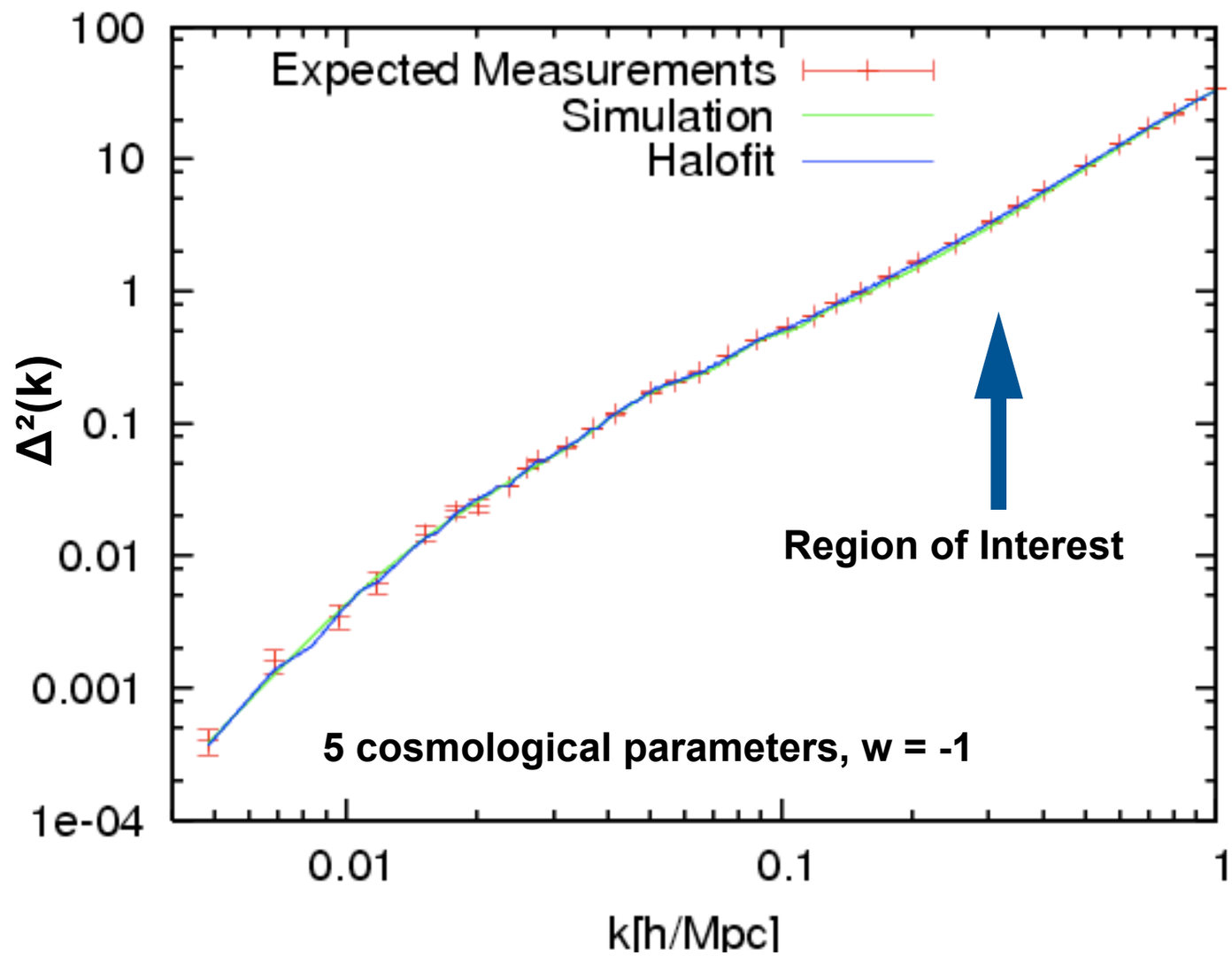
DESC Meeting, Pittsburgh Dec 5, 2013





Why do we need higher accuracy in our predictions?

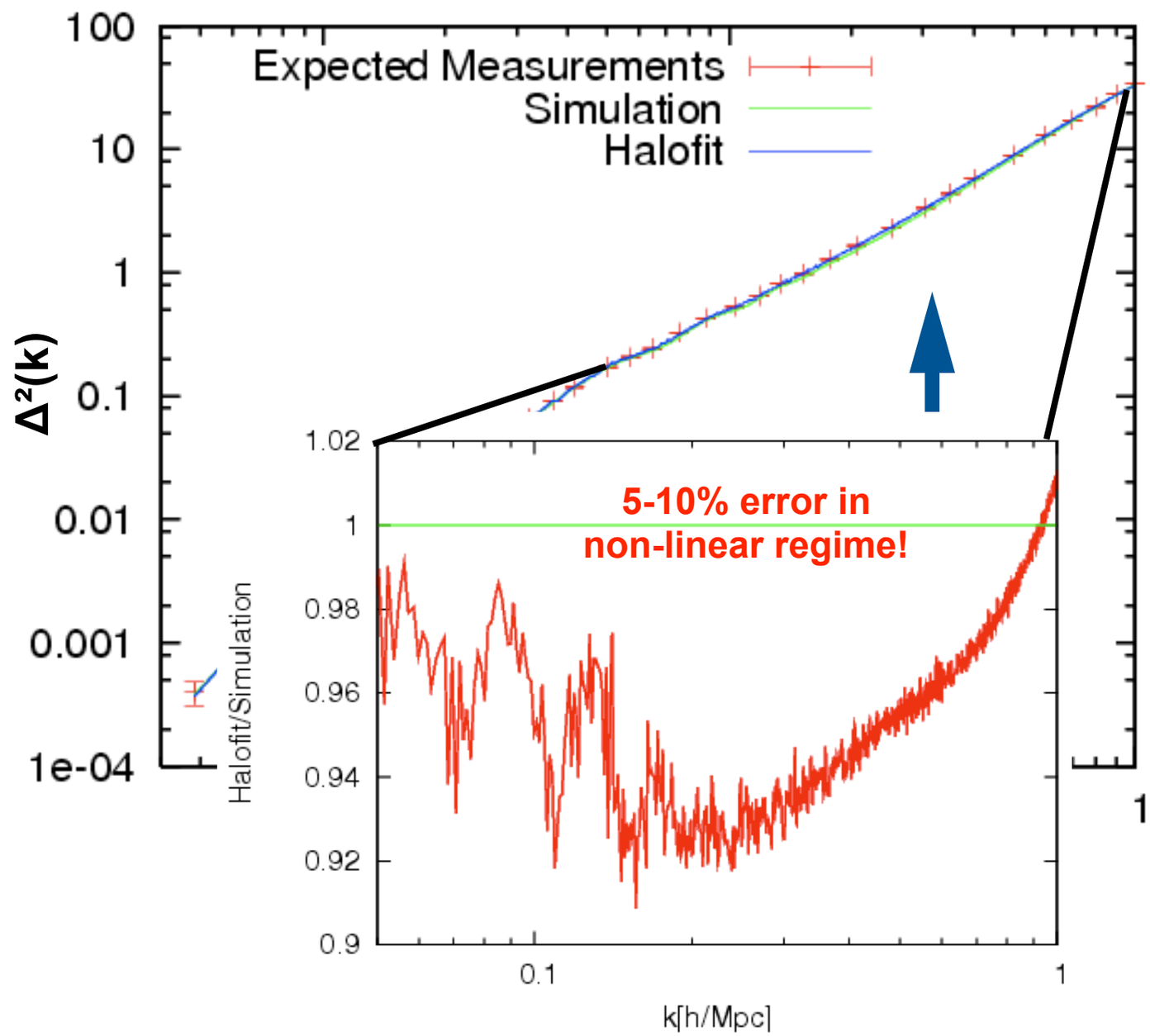
- Example in this talk: matter power spectrum
- Question: how badly will our constraints on dark energy be biased if we *do not* reach the same accuracy in our modeling as we might have in our data?
- Generate mock data set with the expected 1% error
- Analyze data with current method using HaloFit to model the matter power spectrum
  - HaloFit (Smith et al. 2003): semi-analytic fit for the power spectrum, based on modeling approach and tuned to simulations, accurate at the 5-10% level



$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}; \; P(\vec{k}) = \langle \delta^2(\vec{k}) \rangle$$

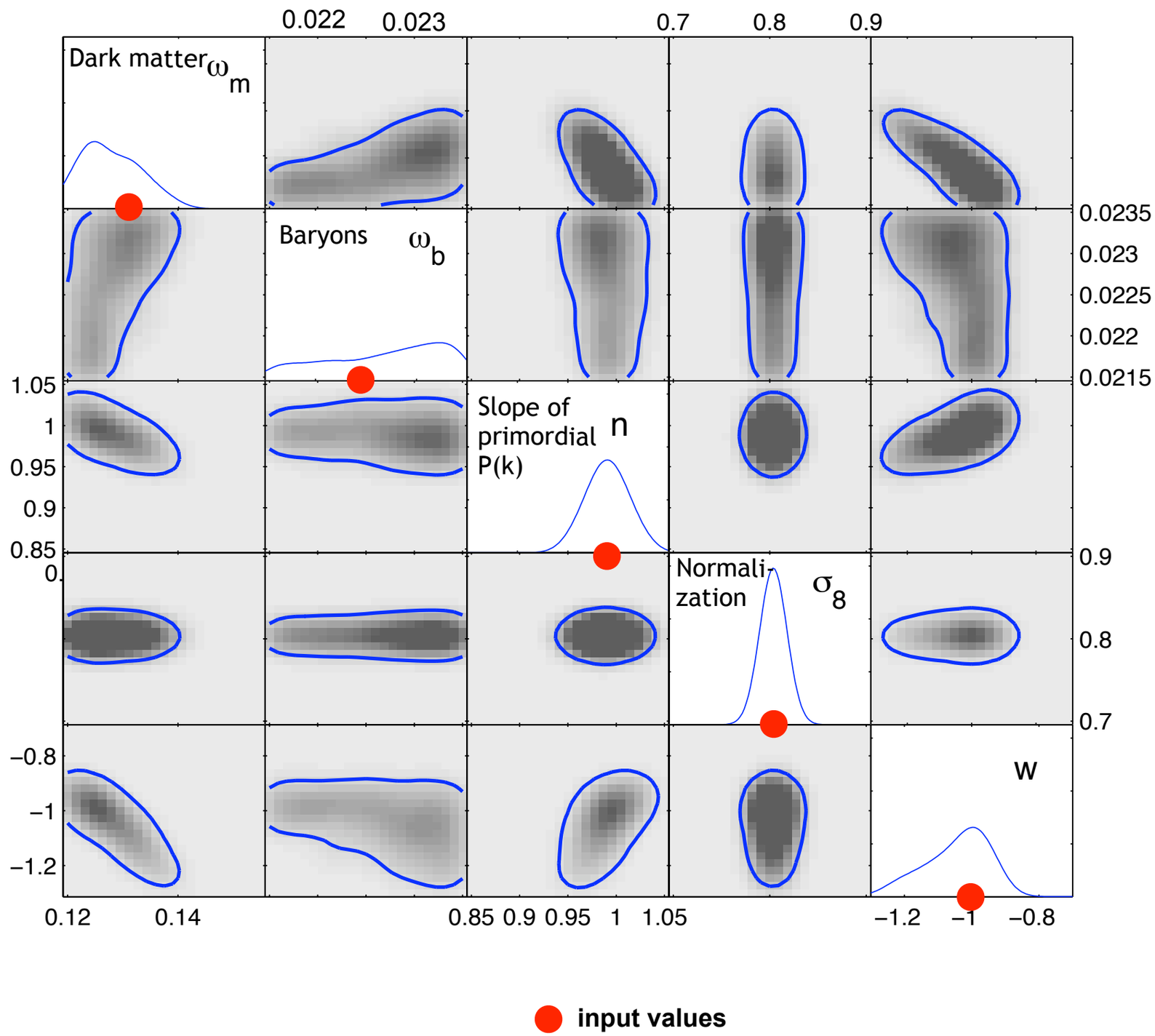
Why do we need higher accuracy in our predictions?

- Example in this talk: matter power spectrum
- Question: how badly will our constraints on dark energy be biased if we *do not* reach the same accuracy in our modeling as we might have in our data?
- Generate mock data set with the expected 1% error
- Analyze data with current method using HaloFit to model the matter power spectrum
  - HaloFit (Smith et al. 2003): semi-analytic fit for the power spectrum, based on modeling approach and tuned to simulations, accurate at the 5-10% level



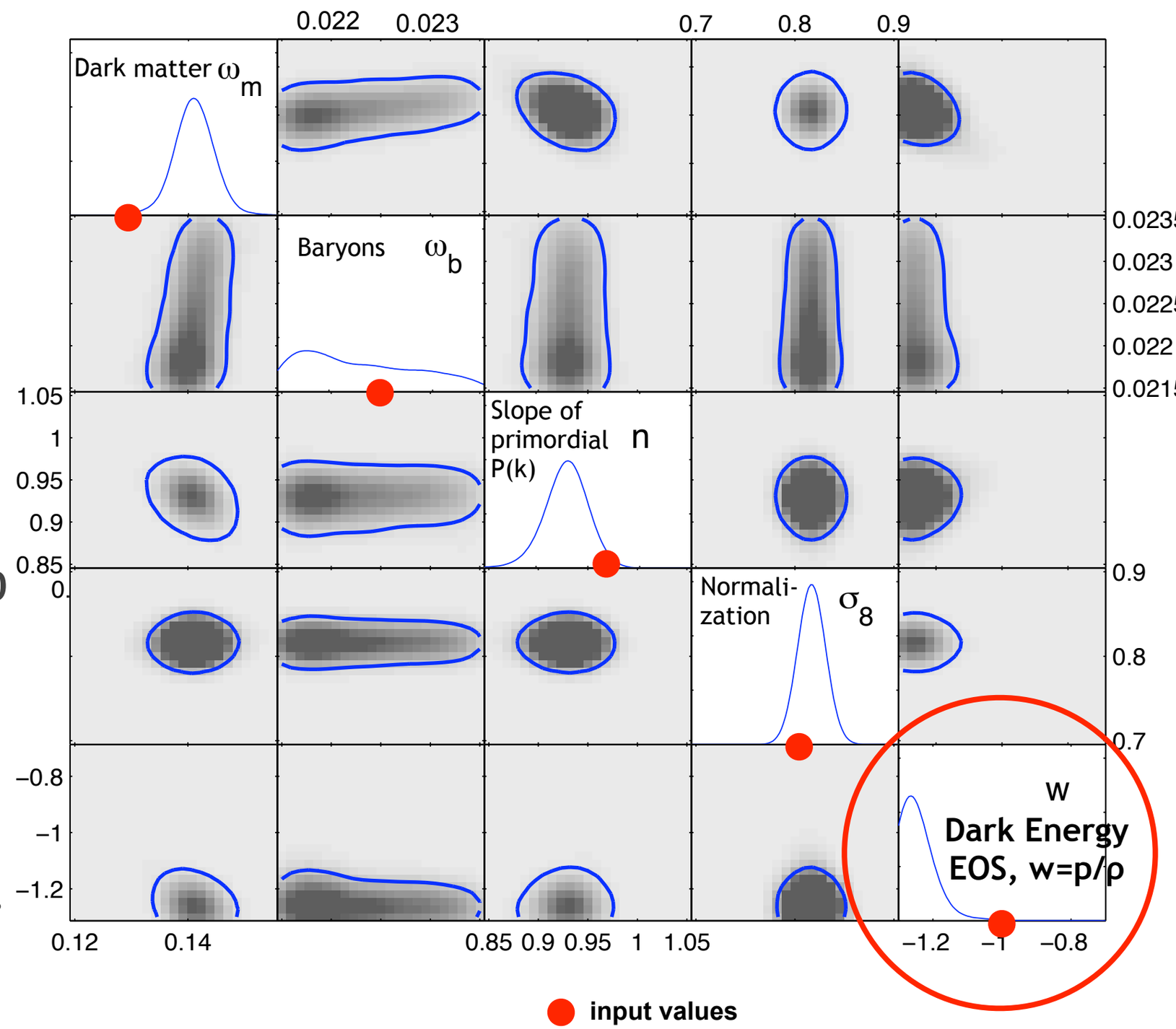
Warm up

- Generate mock data from Halofit
- Analyze data with Halofit
- Results look pretty good!



Analysis of the "True data"

- Generate mock data from high-resolution simulation
- Use Halofit for analysis; remember, halofit ~5-10% inaccurate on scales of interest
- Parameters are up to 20% wrong!
- Only solution: precision simulations
- Analysis takes at least 10,000 input power spectra for MCMC, each simulation takes ~20,000 CPU hours
- With a 2000 node cluster running 24/7, our analysis will take ~30 years, hmmm...

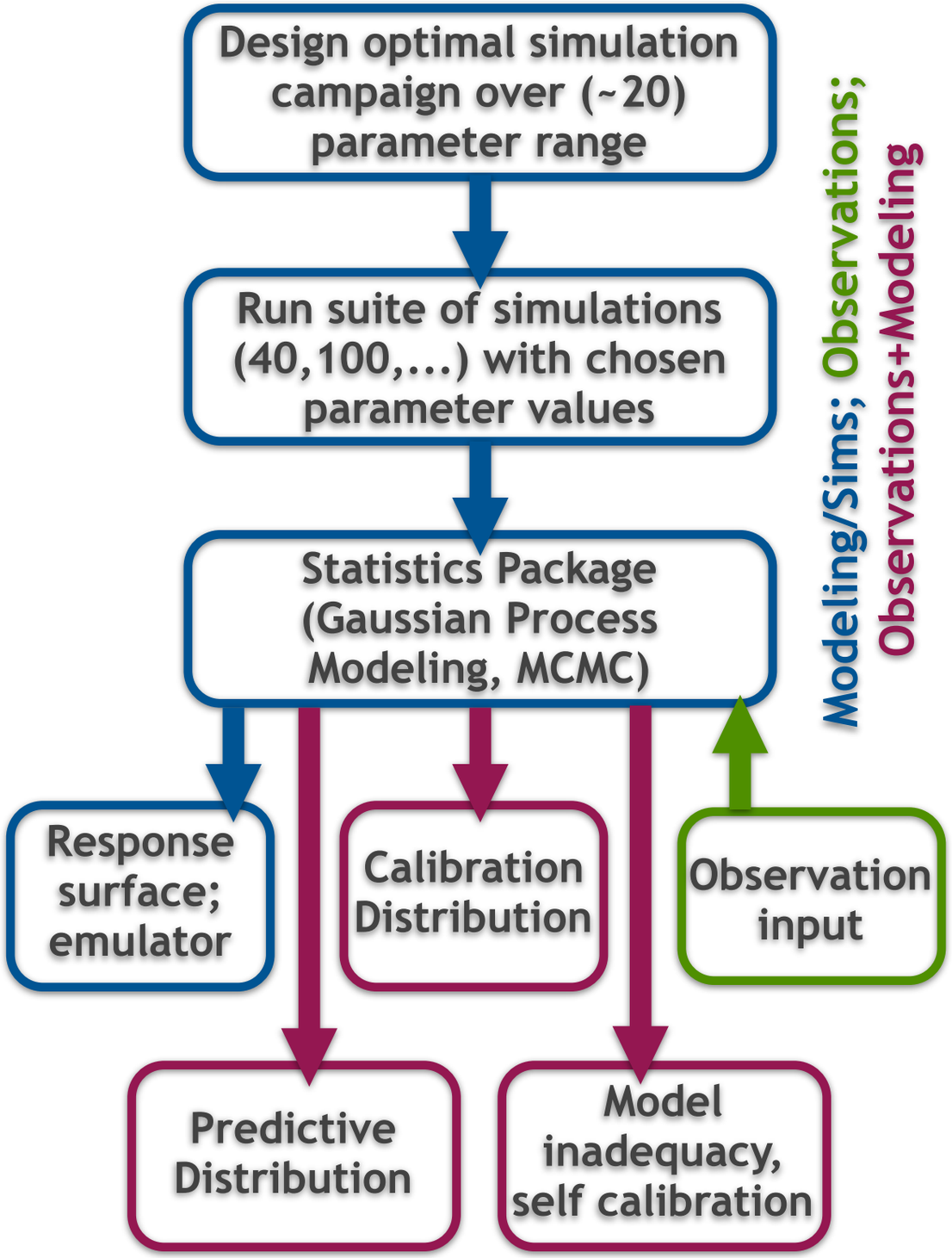


Cosmic Calibration: Solving the Inverse Problem

- **Challenge:** To extract cosmological constraints from observations in nonlinear regime, need to run Marko Chain Monte Carlo code; input: 10,000 - 100,000 different models
- **Current strategy:** Fitting functions for e.g. P(k), accurate at 10% level, as we saw this is not good enough!
- **Our alternative:** Emulators, fast prediction schemes built from a manageable set of simulations
- **Example here: Power spectrum emulator**
  - Step 1: Show simulations have required accuracy (Heitmann et al. 2005, 2008, 2010)
  - Step 2: Determine minimum number of simulations needed and develop sophisticated interpolation scheme that provides the power spectrum for any cosmology within a given parameter space prior (Heitmann et al. 2006, 2009; Habib et al. 2007)
  - Step 3: Carry out simulation and build final emulator (Lawrence et al. 2010, Heitmann et al. 2013)

Cosmic Calibration Framework

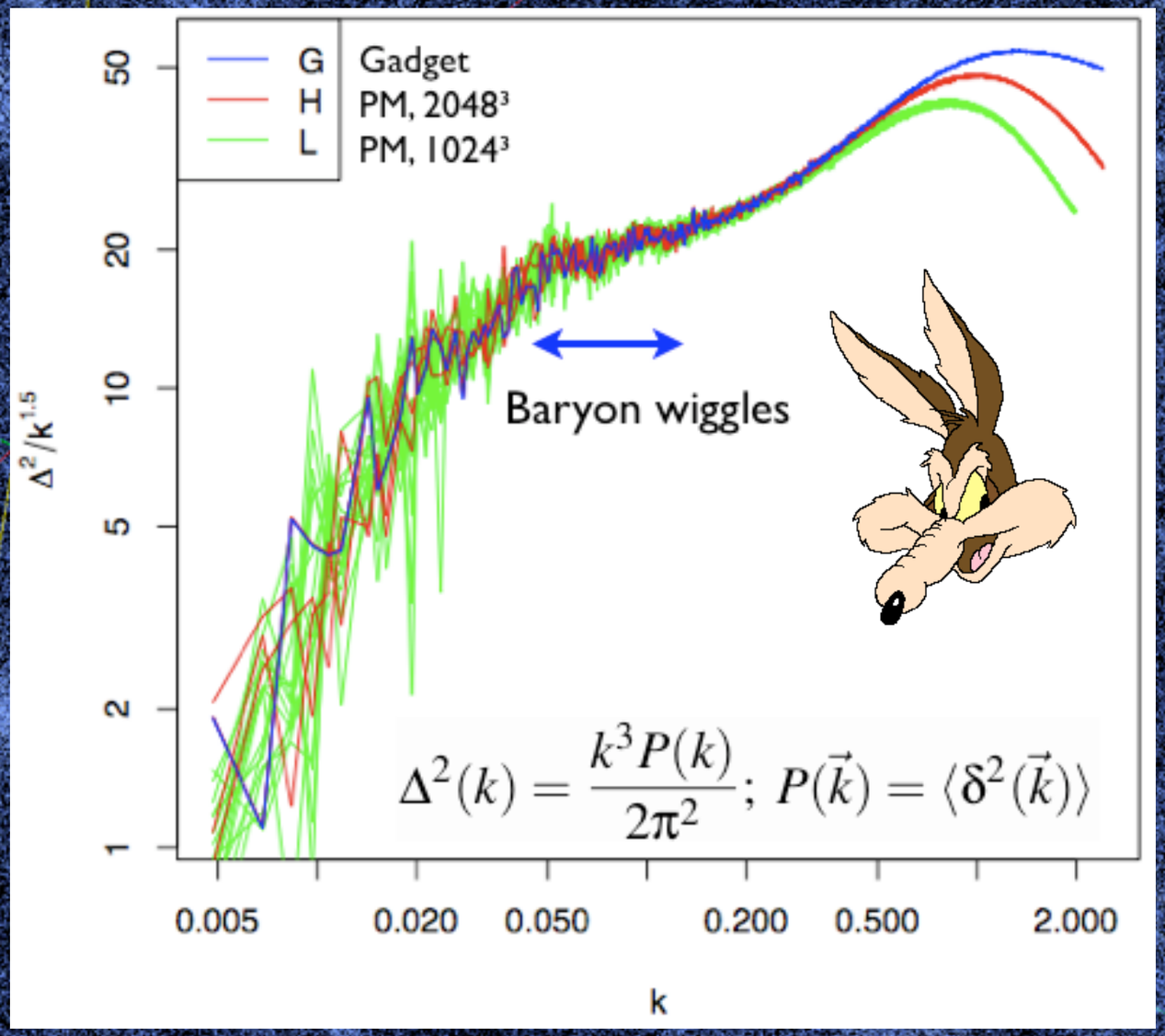
- Step 1: Design simulation campaign, rule of thumb: O(10) models for each parameter
- Step 2: Carry out simulation campaign and extract quantity of interest, in our case, power spectrum
- Step 3: Choose suitable interpolation scheme to interpolate between models, here Gaussian Processes
- Step 4: Build emulator
- Step 5: Use emulator to analyze data, determine model inadequacy, refine simulation and modeling strategy...



The Original Coyote Universe

Priors:

$0.020 \leq \omega_b \leq 0.025$
$0.11 \leq \omega_m \leq 0.15$
$0.85 \leq n_s \leq 1.05$
$-1.3 \leq w \leq -0.7$
$0.6 \leq \sigma_8 \leq 0.9$



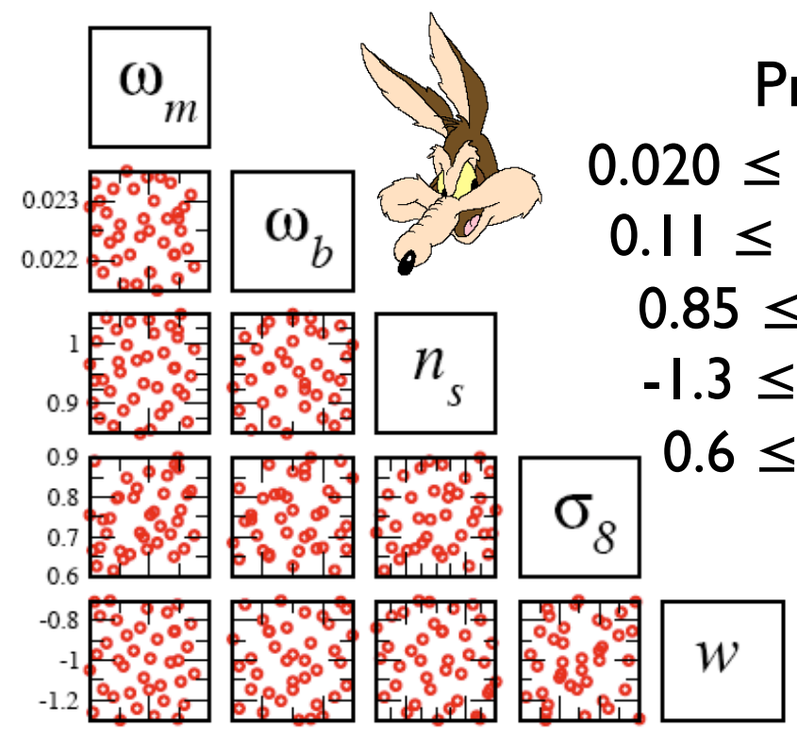
- 37 model runs + ΛCDM
  - 16 low resolution realizations (green)
  - 4 medium resolution realizations (red)
  - 1 high resolution realization (blue)
  - 11 outputs per run between z = 0 - 3
- Restricted priors to minimize necessary number of runs
- 1.3 Gpc boxes, $m_p \sim 10^{11} M_\odot$
- ~1000 simulations, 60TB

Background Visualization with ParaView by J. Woodring

The Simulation Design for wCDM Cosmologies

- "Simulation design": For a given set of parameters to be varied and a fixed number of runs, at what settings should the simulations be performed?
- Example: Five cosmological parameters, tens of high-resolution runs are affordable
- First idea: Grid
  - Space filling but poor projection properties
- Second idea: Random sampling
  - Good projection properties but poor space coverage
- Our approach: Orthogonal-array Latin hypercubes (OA-LH) design
  - Stratified random sampling approach
  - Good projection properties AND space filling

The Coyote Universe



Priors:

$0.020 \leq \omega_b \leq 0.025$

$0.11 \leq \omega_m \leq 0.15$

$0.85 \leq n_s \leq 1.05$

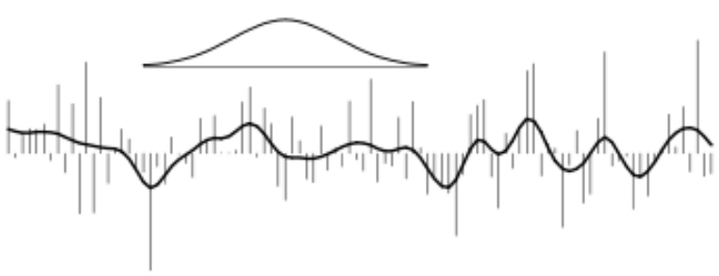
$-1.3 \leq w \leq -0.7$

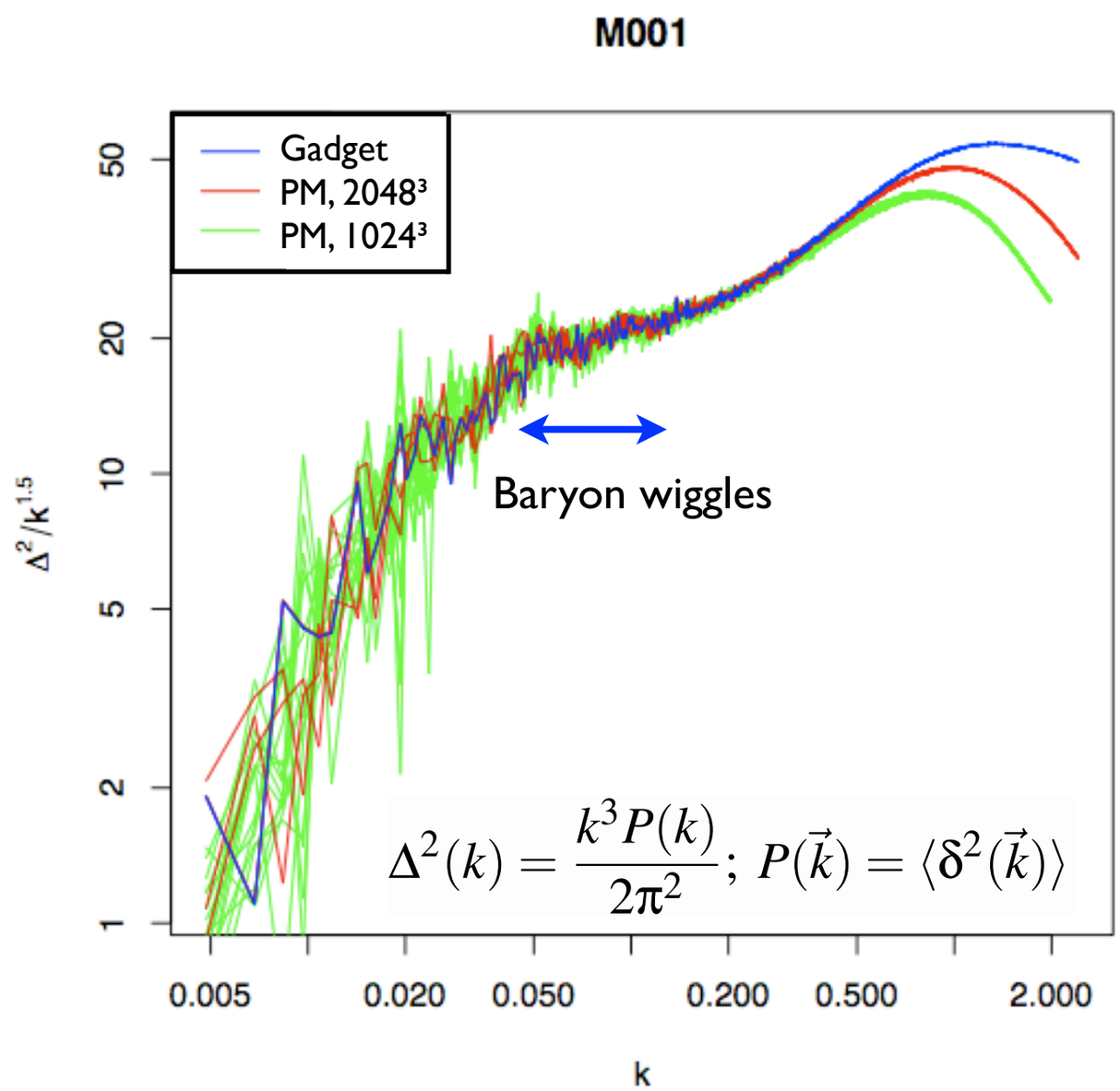
$0.6 \leq \sigma_8 \leq 0.9$

Priors are informed by current cosmological constraints, the tighter the priors, the easier to build a prediction tool. Restriction in number of parameters also helps!

Next step: Smooth Power Spectrum

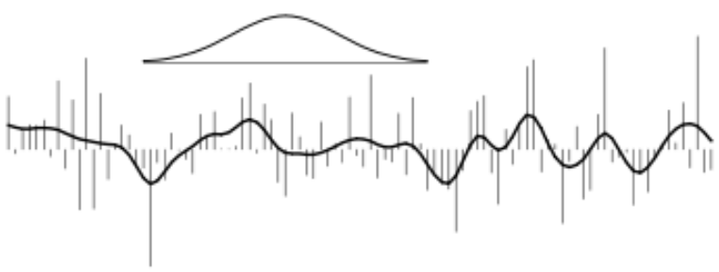
- Each simulation represents one possible realization of the Universe in a finite volume
- Need smooth prediction for building the emulator for each model
- Major challenge: Make sure that baryon features are not washed out or enhanced due to realization scatter
- Construct smooth power spectra using a process convolution model (Higdon 2002)
- Basic idea: calculate moving average using a kernel whose width is allowed to change to account for nonstationarity

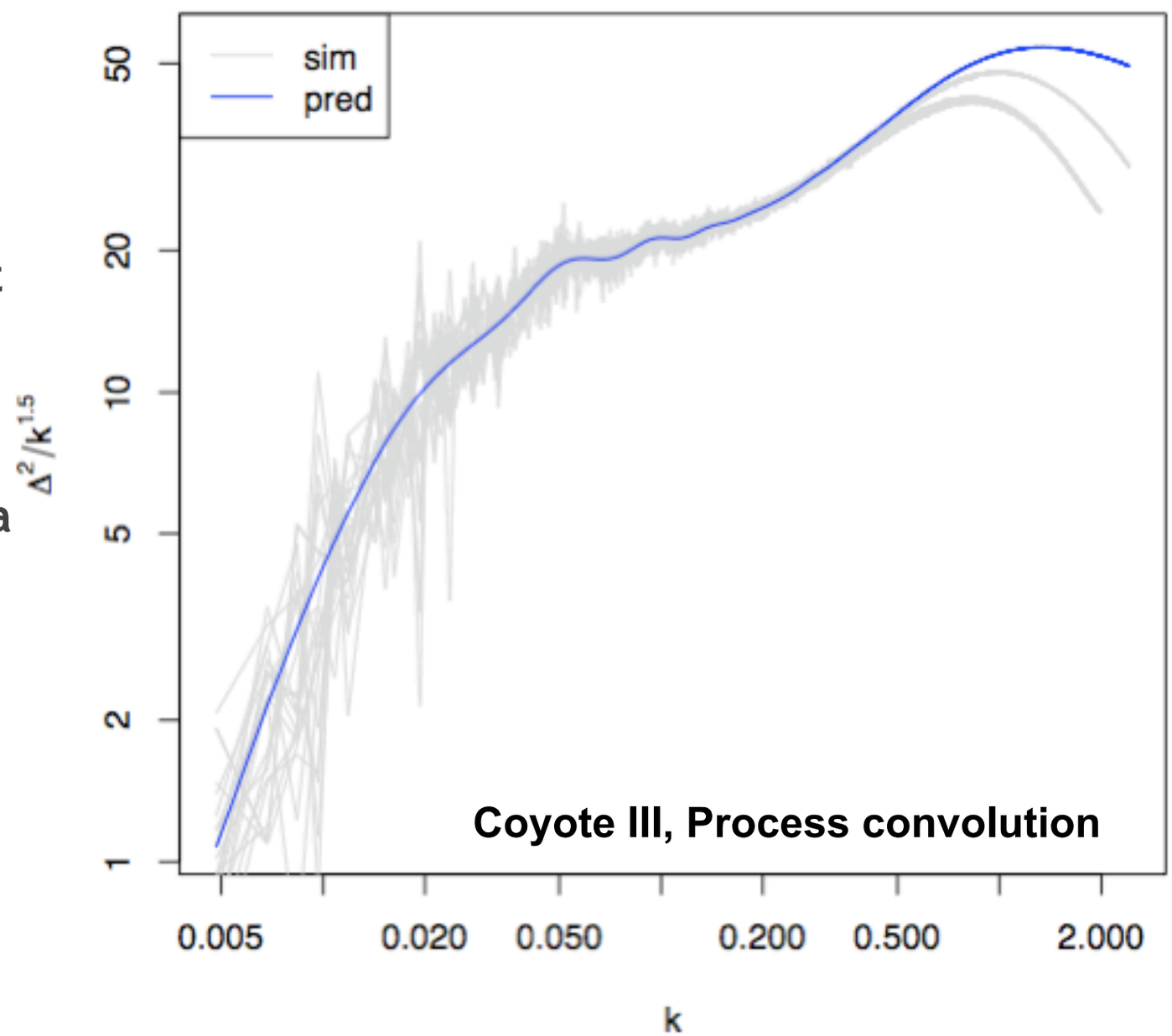




Next step: Smooth Power Spectrum

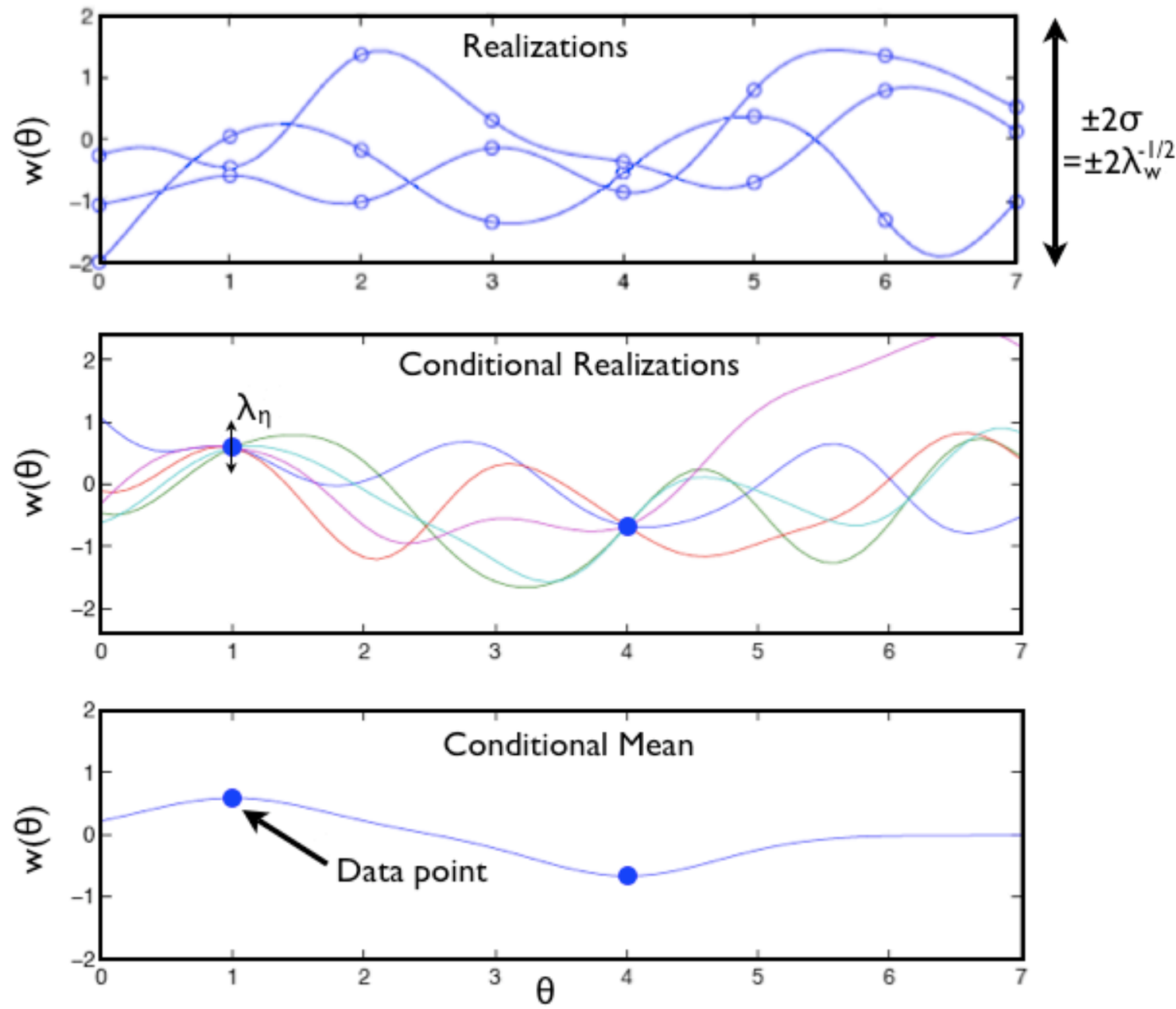
- Each simulation represents one possible realization of the Universe in a finite volume
- Need smooth prediction for building the emulator for each model
- Major challenge: Make sure that baryon features are not washed out or enhanced due to realization scatter
- Construct smooth power spectra using a process convolution model (Higdon 2002)
- Basic idea: calculate moving average using a kernel whose width is allowed to change to account for nonstationarity





The Interpolation Scheme: Gaussian Processes + PCA

- After simulation design specification: Build non-parametric interpolation scheme
- Gaussian Process (GP): fits in function space
- GP involves matrix inversion in conditioning step ("curse of dimensionality")
- Data compression: Express power spectra in terms of principal component (PC) basis (can use other basis too)
- GP over over PC coefficients



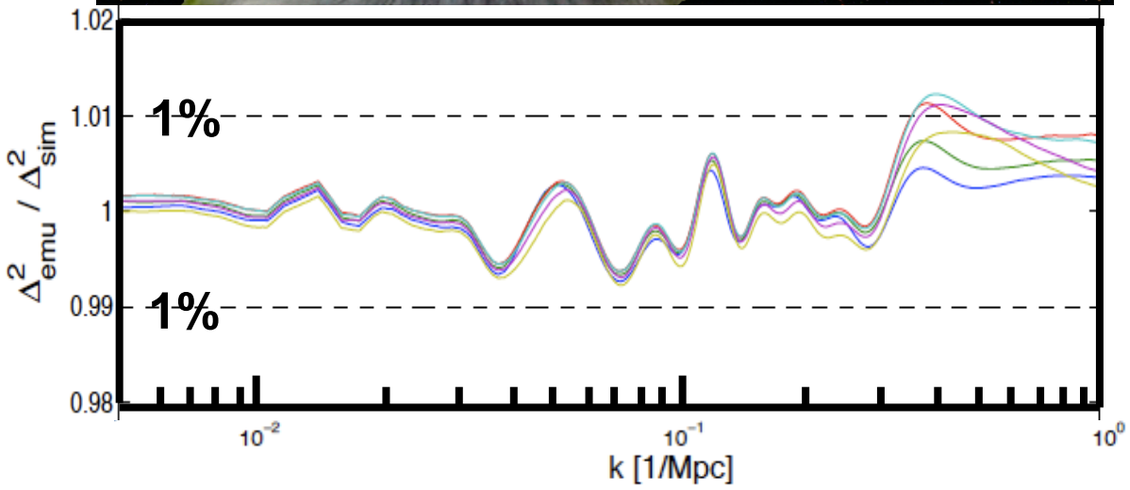
Cosmic Emulator in Action

- Instantaneous ‘oracle’ for nonlinear power spectrum, reduces compute time from weeks to negligible, accurate at 1% out to k~1/Mpc for wCDM cosmologies
- Enables direct MCMC with results from full simulations

The Cosmic Emu(lator)

- Prediction tool for matter power spectrum has been constructed
- Accuracy within specified priors between z=0 and z=1 out to k=1 *h*/Mpc at the 1% level achieved
- Emulator has been publicly released, C code
- Extension: Include *h* as sixth parameter, out to k=10 *h*/Mpc and z=4
  - Nested simulations to cover large k-range
  - Approach degrades accuracy to ~3%

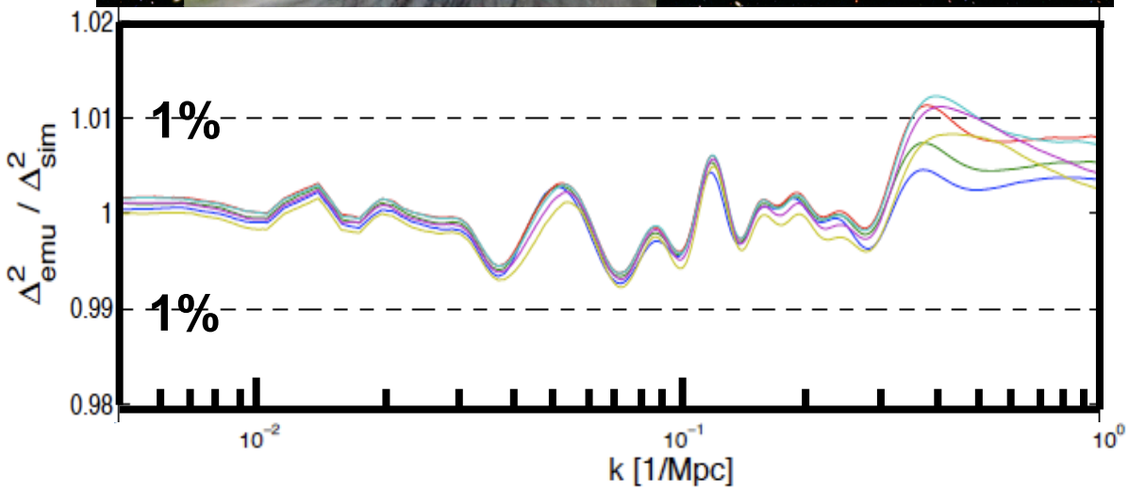
Emulator performance: Comparison of prediction and simulation output for a model not used to build emulator at 6 redshifts.



The Cosmic Emu(lator)

- Prediction tool for matter power spectrum has been constructed
- Accuracy within specified priors between z=0 and z=1 out to k=1 *h*/Mpc at the 1% level achieved
- Emulator has been publicly released, C code
- Extension: Include *h* as sixth parameter, out to k=10 *h*/Mpc and z=4
  - Nested simulations to cover large k-range
  - Approach degrades accuracy to ~3%

Emulator performance: Comparison of prediction and simulation output for a model not used to build emulator at 6 redshifts.

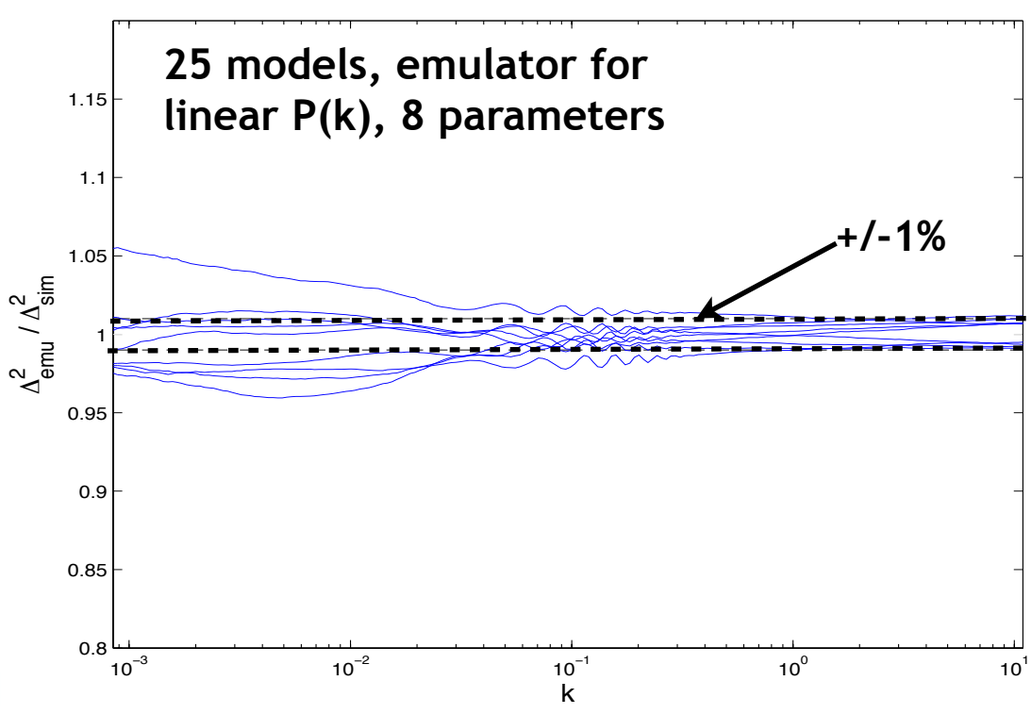


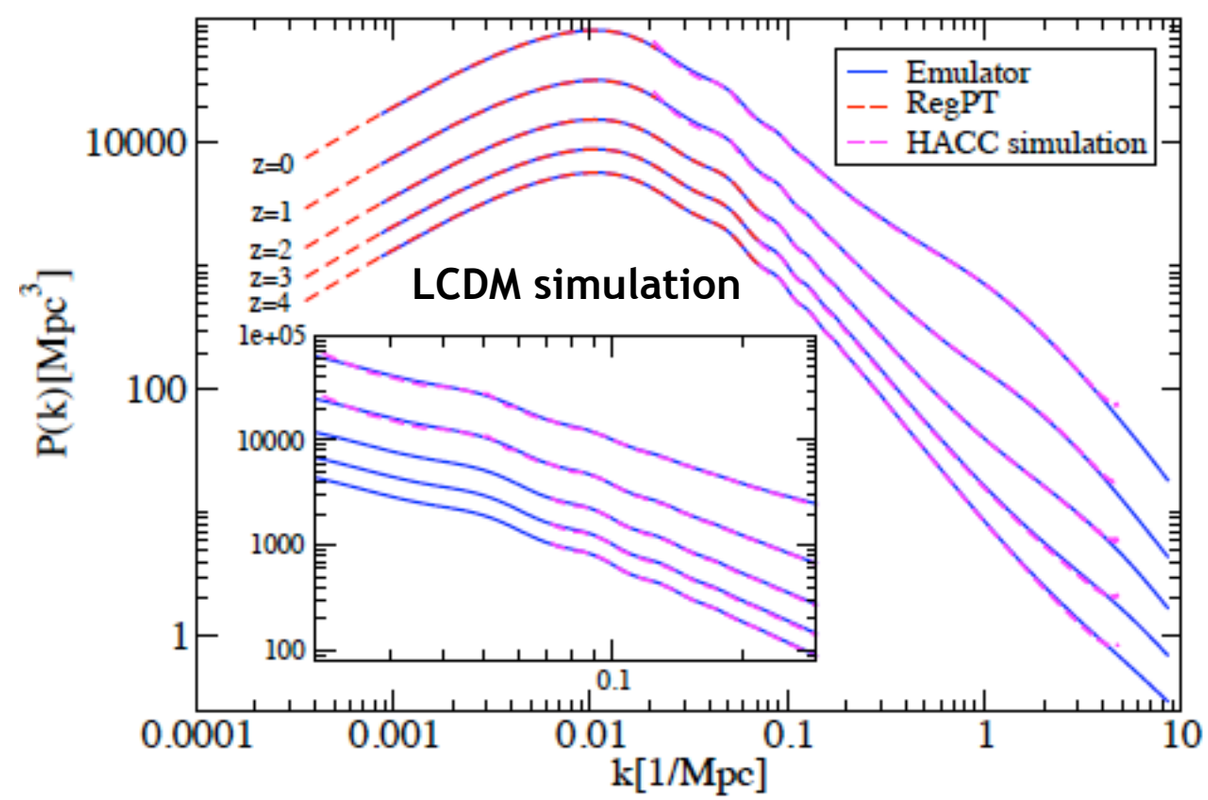
The Next Step: The Mira Universe

- Extend parameter space to include varying w(z) and massive neutrinos
- Build "nested designs": enable to build emulator from first set of 25 models, improve with additional 27 models, final precision with 99 models overall
- Various emulators for P(k), mass function, c-M relation, derived quantities...
- LCDM done, finalizing set-up based on this run

Parameters

$$0.12 \leq \omega_m \leq 0.155$$
$$0.0215 \leq \omega_b \leq 0.0235$$
$$0.7 \leq \sigma_8 \leq 0.9$$
$$0.55 \leq h \leq 0.85$$
$$0.85 \leq n_s \leq 1.05$$
$$-1.3 \leq w_0 \leq -0.7$$
$$-1.5 \leq w_a \leq 1.15$$
$$0.0 \leq \omega_\nu \leq 0.01.$$



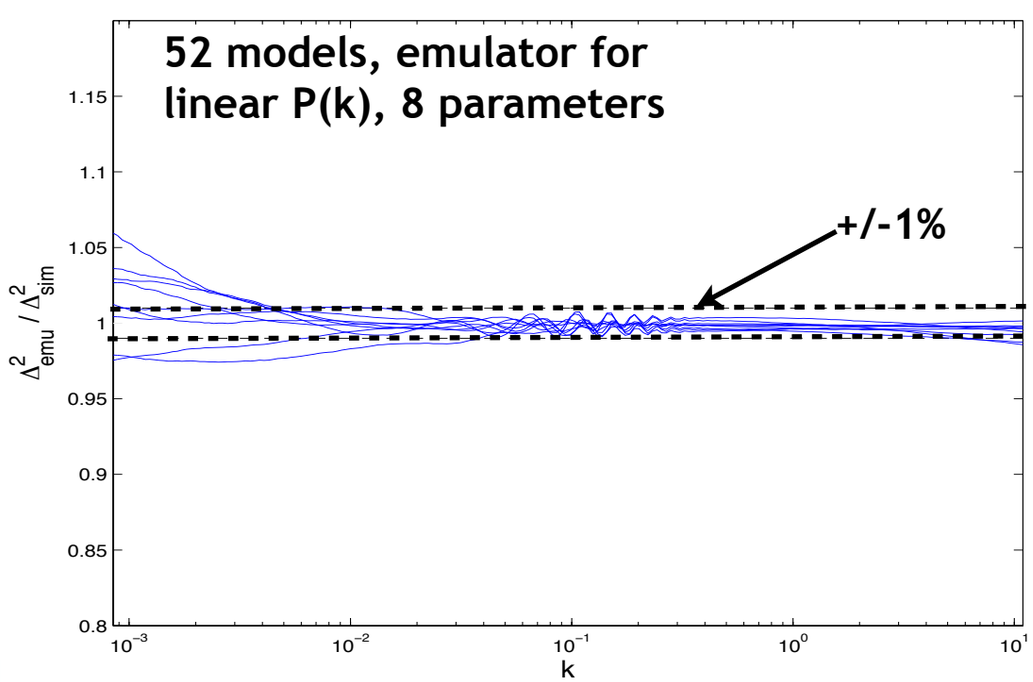


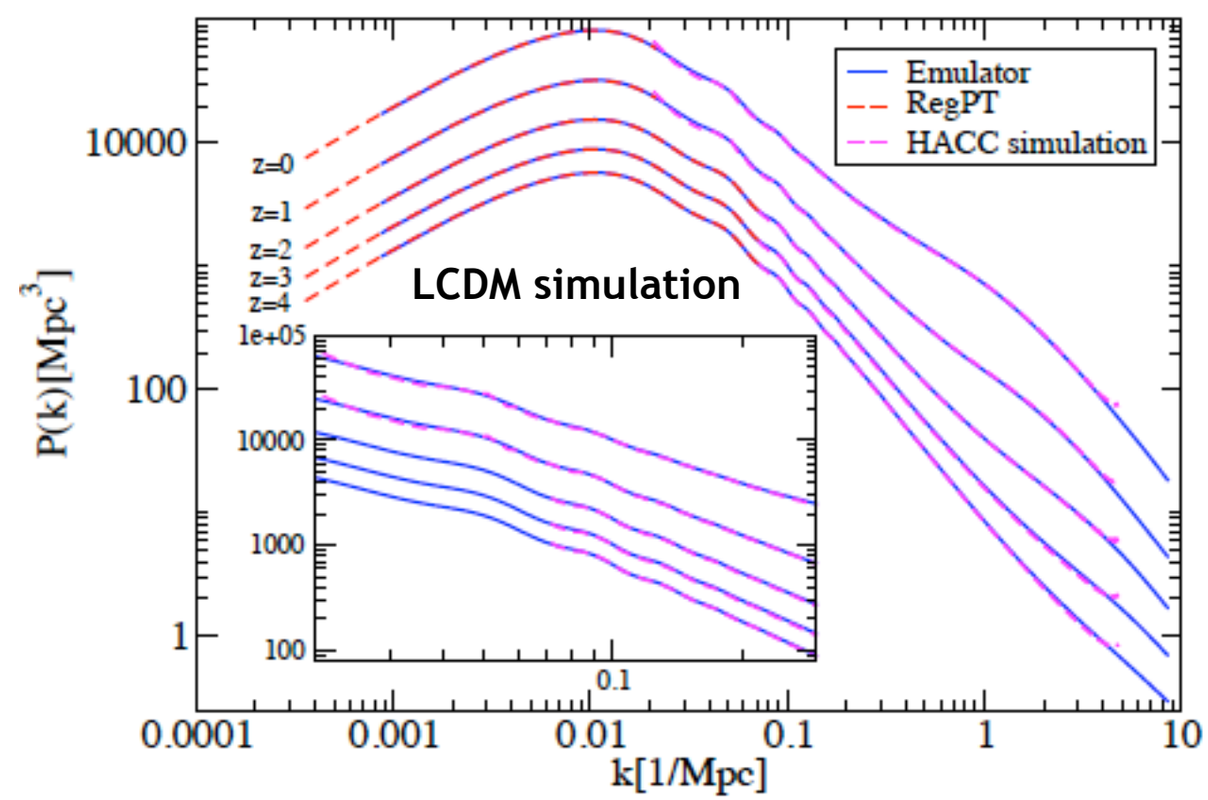
The Next Step: The Mira Universe

- Extend parameter space to include varying w(z) and massive neutrinos
- Build "nested designs": enable to build emulator from first set of 25 models, improve with additional 27 models, final precision with 99 models overall
- Various emulators for P(k), mass function, c-M relation, derived quantities...
- LCDM done, finalizing set-up based on this run

Parameters

$$0.12 \leq \omega_m \leq 0.155$$
$$0.0215 \leq \omega_b \leq 0.0235$$
$$0.7 \leq \sigma_8 \leq 0.9$$
$$0.55 \leq h \leq 0.85$$
$$0.85 \leq n_s \leq 1.05$$
$$-1.3 \leq w_0 \leq -0.7$$
$$-1.5 \leq w_a \leq 1.15$$
$$0.0 \leq \omega_\nu \leq 0.01.$$



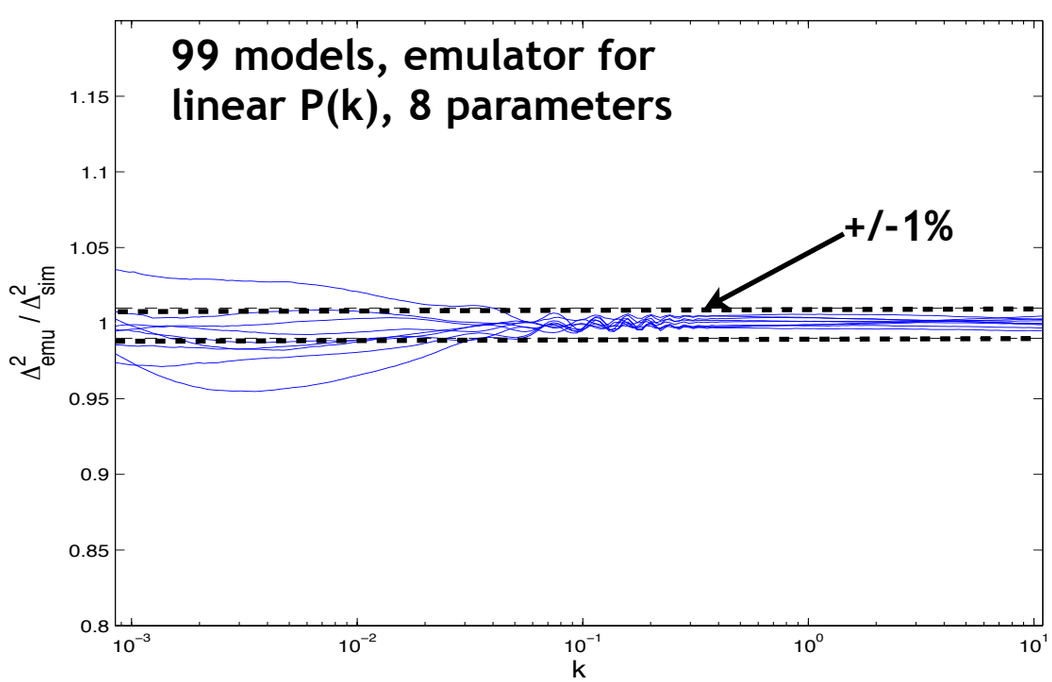


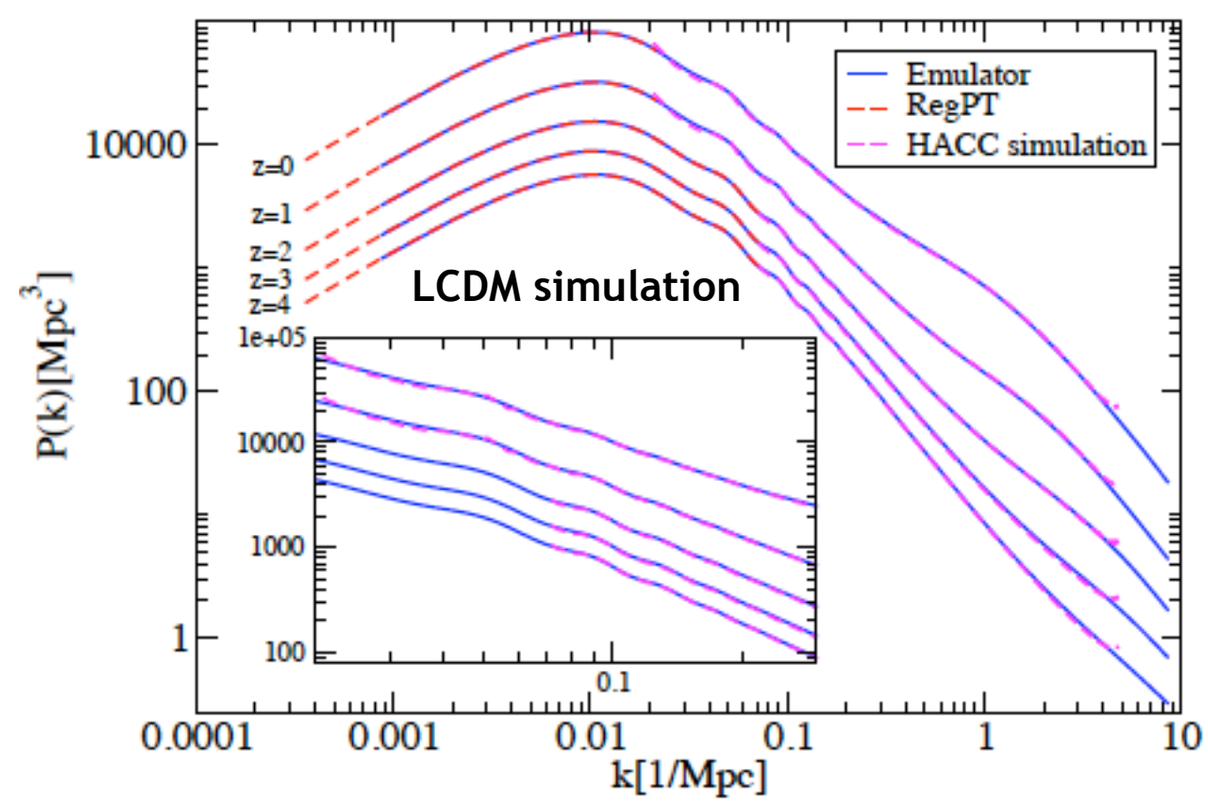
The Next Step: The Mira Universe

- Extend parameter space to include varying w(z) and massive neutrinos
- Build "nested designs": enable to build emulator from first set of 25 models, improve with additional 27 models, final precision with 99 models overall
- Various emulators for P(k), mass function, c-M relation, derived quantities...
- LCDM done, finalizing set-up based on this run

Parameters

$$0.12 \leq \omega_m \leq 0.155$$
$$0.0215 \leq \omega_b \leq 0.0235$$
$$0.7 \leq \sigma_8 \leq 0.9$$
$$0.55 \leq h \leq 0.85$$
$$0.85 \leq n_s \leq 1.05$$
$$-1.3 \leq w_0 \leq -0.7$$
$$-1.5 \leq w_a \leq 1.15$$
$$0.0 \leq \omega_\nu \leq 0.01.$$





Summary and Outlook

- Precision cosmology needs high accuracy predictions! Can we avoid being theory limited?
- Cosmic Calibration Framework allows us to build fast prediction tools for ongoing and future surveys
- *The Mira Universe* will lead to an unprecedented set of simulations, spanning 8 cosmological parameters, including different dark energy models and neutrinos
- In this talk: Focus on power spectrum science but other emulators can be easily built, we have built in addition to matter power spectrum -- galaxy power spectrum, c-M relation (Kwan et al. 2012, Kwan et al. 2013)