Connecting theoretical models to observations: modeling and computing requirements

Michael D. Schneider

LLNL

Theory requirements: inputs to the Likelihood function

$$L(\text{data} \mid \text{source}, \text{noise}, \text{systematics})$$

1. **Monte Carlo simulations** for model parameter inference
2. **Definition of probes** to exploit discovery potential
3. **Validation** (accuracy & precision) of prediction codes
4. **Covariance matrices** for joint analysis of redshift slices / different probes
5. **Models for** characterization and marginalization of **systematics**

1. Optimizing or sampling from the Likelihood function

- Timeline
  - Pre-launch: Define / validate algorithms
  - Post-launch: Run at scale
- Technical / computing requirements
  - 2-point functions: Solved by CosmoLike (& similar)
  - MCMC given pixel data:

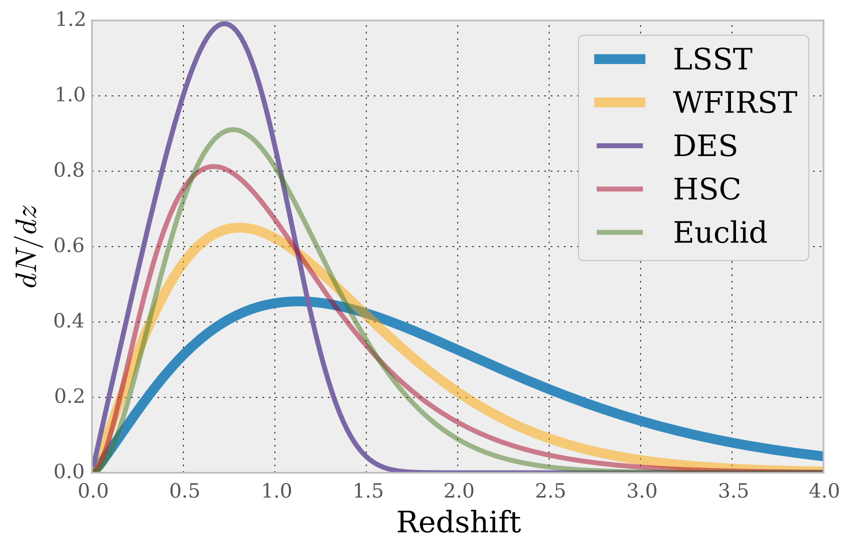
$$(10^9\,\mathrm{sources})\,(10^5\,\mathrm{samples/source})\,(10^{-2}\,\mathrm{sec/sample})\,(10^5\,\mathrm{cores})^{-1} = 10^7\,\mathrm{sec/core}$$

- Or, ~4 months to process the pixel data
  - Repeat with new epochs or combined survey information

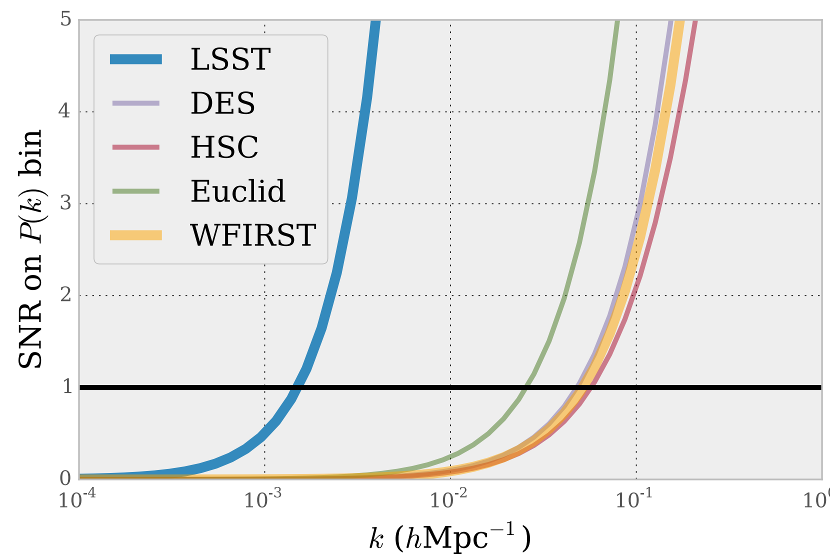
2. Defining probes to enable new discoveries

Ex: WFIRST-AFTA strengths for large-scale structure measurements

Depth



Volume



WFIRST-AFTA ideal for gravitational growth rate measurements.

2. Defining probes to enable new discoveries

- Timeline
  - Pre-launch: What can we measure?
  - Post-launch: Computing to iterate & refine estimators
- Technical / computing requirements
  - 2-point functions. Binning, weighting, source properties.
  - Mass maps, broad-band power, groups/clusters, velocity fields, combination with other data sets (e.g., CMB/LSS cross-corr).
    - Need cosmology simulations to develop methods for data interpretation.

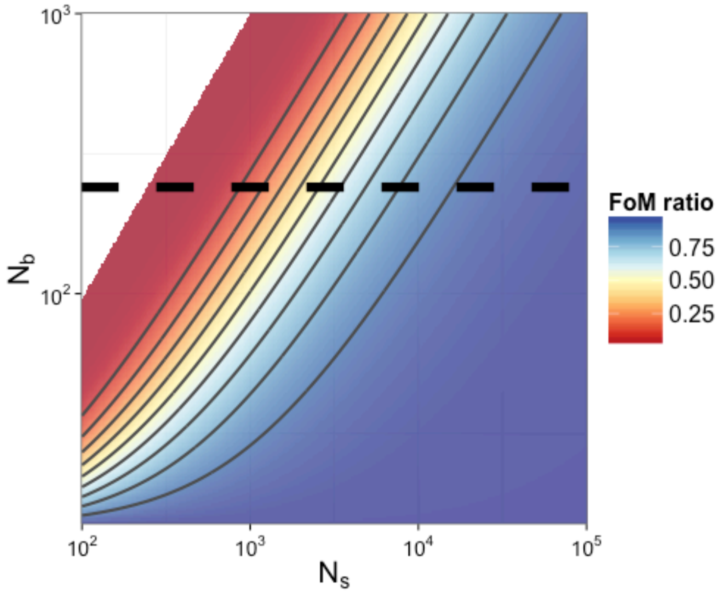
3. Validation of prediction codes

- Timeline
  - Pre-launch: Simulations
    - Internal & external consistency
  - Post-launch: Ideally no support needed!
- Technical / computing requirements
  - Compare cosmology codes with each other & with higher-accuracy simulations
    - Helps to have public code bases
    - CAMB, Coyote universe, GalSim, CosmoSIS
  - Theory / algorithm advances for alternative cosmologies
    - Ex: Modified gravity transfer functions, mass functions, halo profiles
    - Ex: nonlinear clustering with massive neutrinos
  - Compare codes with targeted data sets
    - Ex: SAM comparisons with luminosity functions
    - Ex: Atmosphere turbulence models for cosmic shear PSFs

4. Covariance matrices

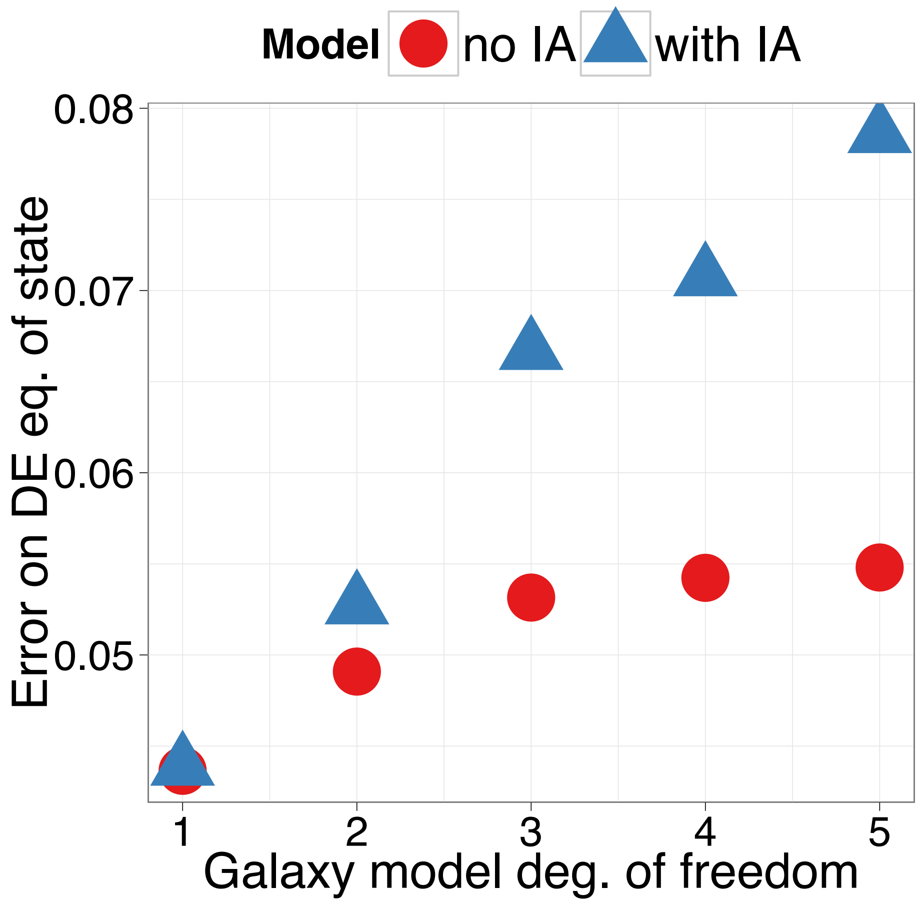
Linked with choice of probes & systematics models

- Timeline
  - Pre-launch: How precise do we need? Theory vs simulation?
  - Post-launch: Simulations at scale.
- Technical / computing requirements
  - Simulation realizations (N_s)
  - Theory models (accuracy)
    - Halo models ~20% accurate



Potentially most computationally challenging part of cosmological data interpretation.

5. Systematics characterization



Laszlo et al. (2011)

More parameters in a systematic error model (usually) leads to increased parameter uncertainties.

5. Systematics characterization

- Timeline
  - Pre-launch: Characterize physics models/simulations
  - Post-launch: Assess (data – model) residuals
- Technical / computing requirements (Eifler)
  - Parameterization
    - Impacts on marginal posteriors of cosmo parameters
      - Number of parameters
    - Scale / redshift dependencies
    - Prospects for self-calibration
  - Model validations with ancillary data
  - Algorithms to incorporate ancillary data during analysis