Gaussian Process Modeling in Cosmology:

The Coyote Universe

Katrin Heitmann, ISR-1, LANL

In collaboration with:
Jim Ahrens, Salman Habib, David Higdon, Chung Hsu, Earl Lawrence, Charlie Nakhleh, Christian Wagner, Martin White, Brian Williams

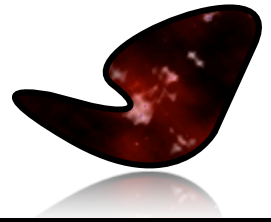
LA-UR-08-05630, LA-UR-08-07921, LA-UR-09-06131

SDSS, First Light in 1998

Visualization: Pat McCormick, CCS-1, LANL

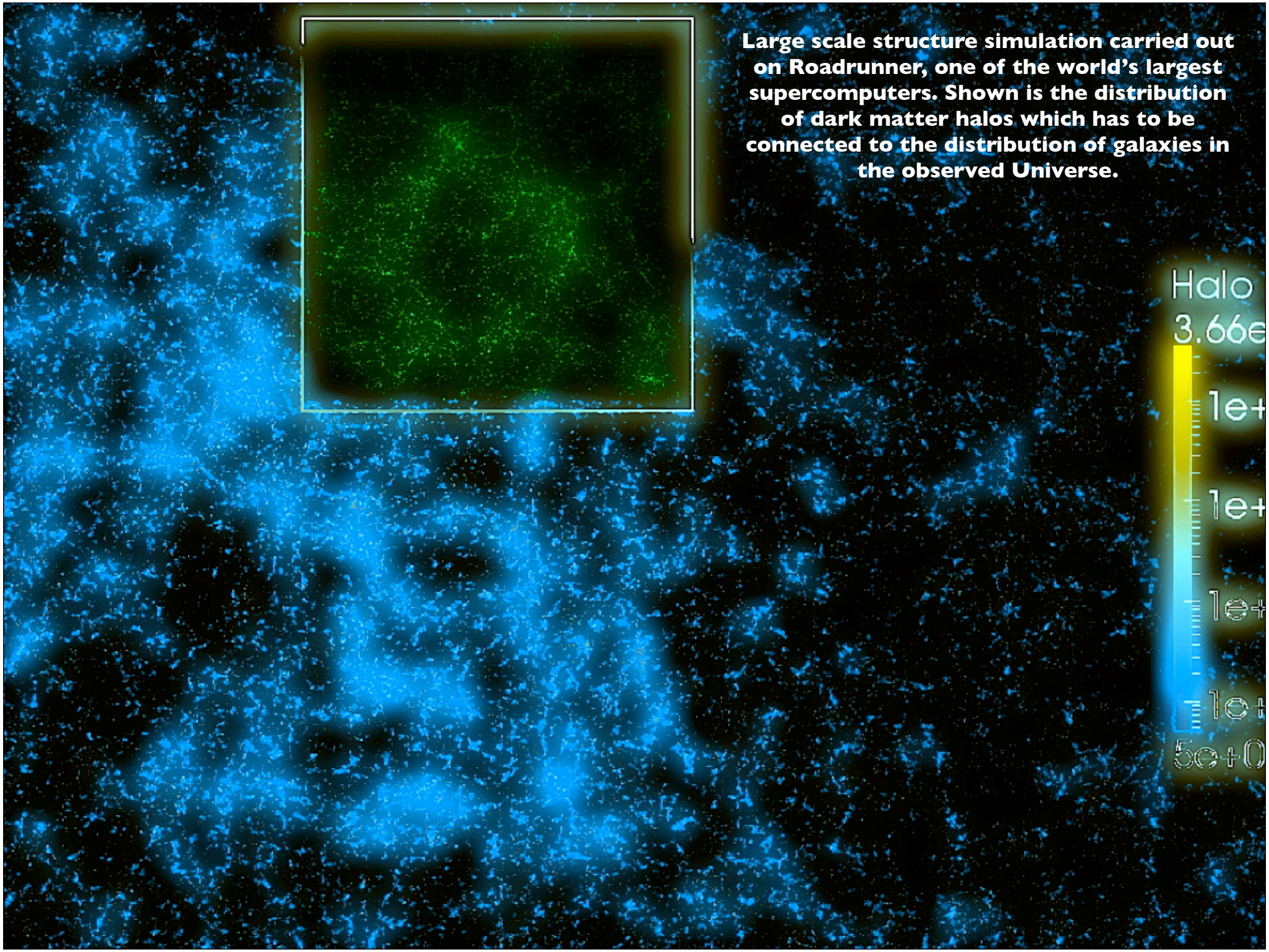
Deep Lens Survey/LSST

The Challenge



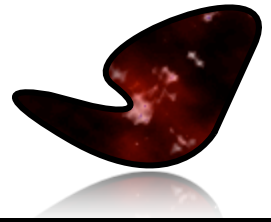
- **Cosmology has entered the era of precision science, from order of magnitude estimates to measurements at the 10% level**
- **Next step: measurements at the 1% level accuracy, theory and predictions have to keep up!**
- **Nonlinear regime of structure formation as measured by galaxy surveys holds wealth of new information on cosmology**
  - Matter power spectrum P(k), 3-point function, mass function...
- **For cosmological constraints from e.g. SDSS:**
  - Run your favorite MCMC code, e.g. CosmoMC
  - Need to calculate different statistics, e.g. P(k), ~10,000 - 100,000 times for different models
  - Current fitting functions for these statistics (tuned to simulations) accurate at the 10% level for different cosmologies, not good enough!
  - Brute force simulations: ~30 years on 2000 processor Beowulf Cluster...

Movie



Large scale structure simulation carried out on Roadrunner, one of the world's largest supercomputers. Shown is the distribution of dark matter halos which has to be connected to the distribution of galaxies in the observed Universe.

A First Step to a Solution



- **Start with a "simple" example, dark matter power spectrum P(k)**
- **Aim: predict P(k) out to scales of k~1 h/Mpc at 1% accuracy between z=0 and z=1**
  - Regime of interest for current weak lensing surveys
  - Baryon physics at these scales sub-dominant
  - Dynamic range for simulations manageable
- **Step 1: Show that simulations can be run at the required accuracy (arXiv:0812.1052)**
  - Initial conditions, force and mass resolution, ...
  - Minimal requirement: 1 billion particles, 1.3 Gpc volume, 50 kpc force resolution, ~ 20,000 CPU hours, few days on 250 processors + wait time in queue ~ 1 week per simulation on "Coyote", LANL cluster
- **Step 2: Cosmic Calibration Framework (arXiv:0902.0429)**
  - Build -- with a small number of high-precision simulations -- a prediction scheme ("emulator") which provides the power spectrum for any cosmology within the parameter space under consideration
  - Showed that ~ 40 cosmological models sufficient
- **Step 3: Cosmic Emulator (arXiv:0912.4490)**
  - Carry out large number of simulations (~1,000) at varying resolution for 38 cosmologies, one high-resolution run per cosmology, emulator is an instantaneous "look-up" table
  - Emulator available at: www.lanl.gov/projects/cosmology/CosmicEmu

The Coyote Universe

Priors:

$0.020 \leq \omega_b \leq 0.025$

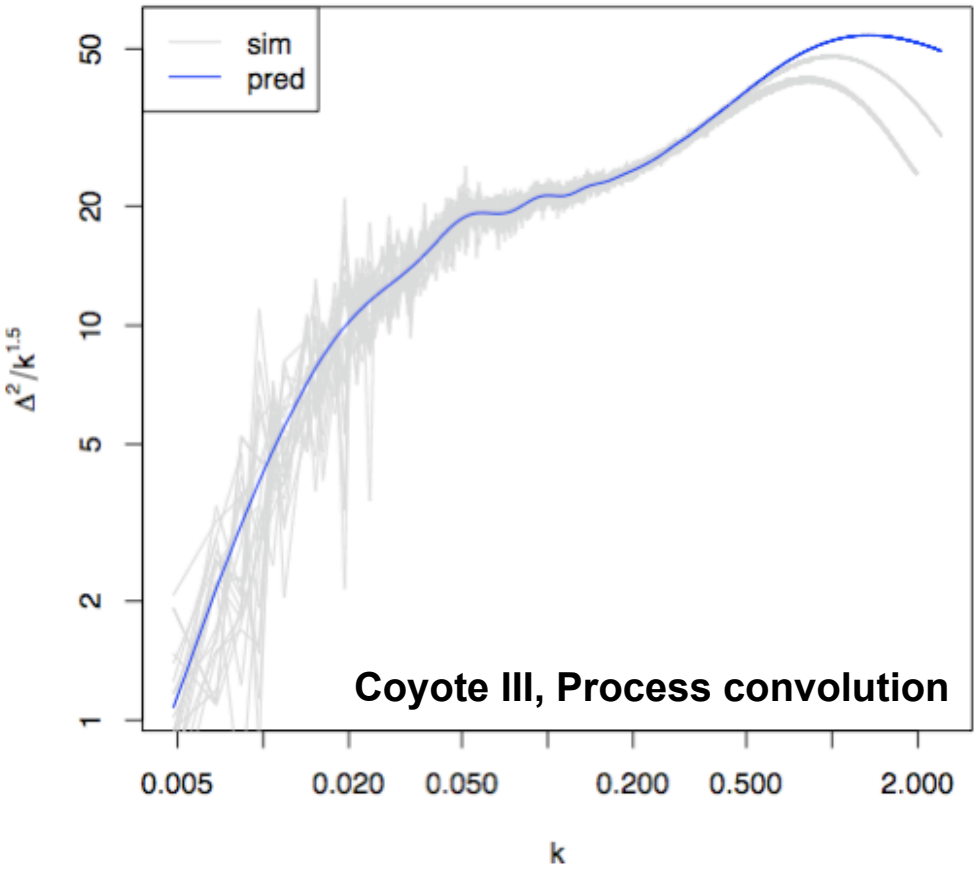
$0.11 \leq \omega_m \leq 0.15$

$0.85 \leq n_s \leq 1.05$

$-1.3 \leq w \leq -0.7$

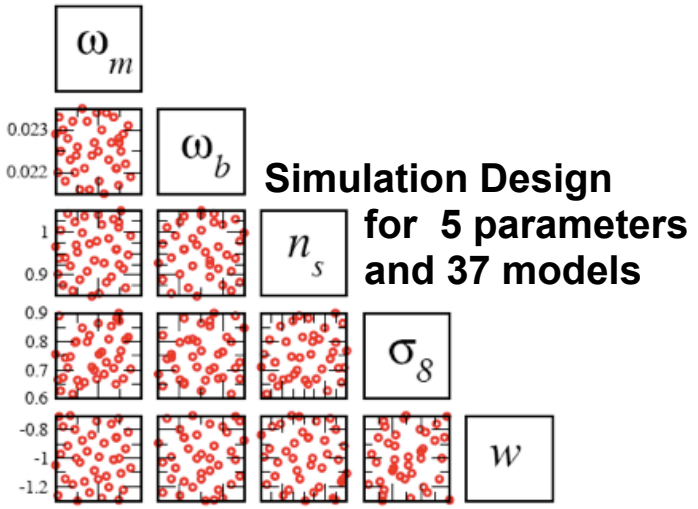
$0.6 \leq \sigma_8 \leq 0.9$

- **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**



The Cosmic Calibration Framework

- **Simulation design: for a given set of parameters to be varied and a fixed number of runs that can be done, at what settings should the simulations be performed?**
  - Orthogonal-array Latin hypercubes (OA-LH) design
- **Interpolation scheme that allows for predictions of the power spectrum for any cosmology within the priors**
  - Model simulation outputs using a p_η- dimensional basis representation
  - Find suitable set of orthogonal basis vectors $\phi_i(k,z)$, here: Principal Component Analysis
  - Model the weights, here: Gaussian Process modeling
- **Uncertainty and sensitivity analysis**
- **Calibration: combining simulations with observations**



Simulation Design for 5 parameters and 37 models

$$\ln\left\{\frac{\Delta^2(k,z)}{2\pi k^{3/2}}\right\} = \sum_{i=1}^{p_\eta} \phi_i(k,z) w_i(\theta) + \varepsilon \qquad \theta \in [0,1]^{p_\theta}$$

Dimensionless power spectrum

Number of basis function, here: 5

Cosmological parameters

Number parameters, 5

Basis functions, here: PC basis

Weights, here: GP model

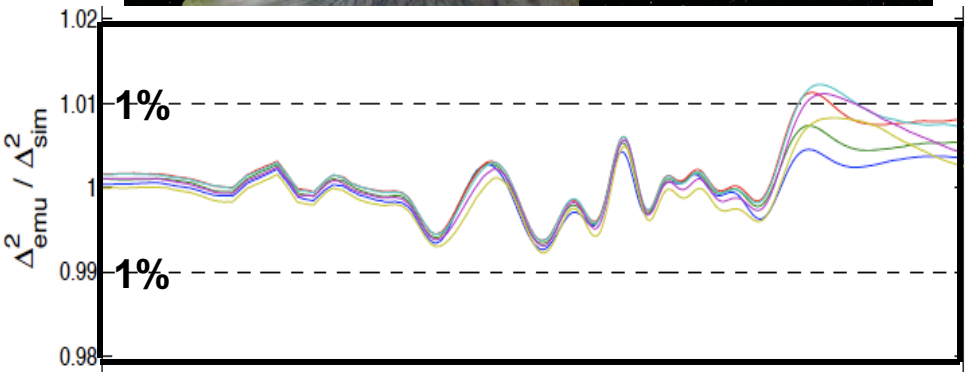
Gaussian Process Modeling:

- **Nonparametric regression scheme, particularly well suited for interpolation of smooth functions**
- **Local interpolator**
- **Extending the notion of a Gaussian distribution over scalar or vector variables into function space**

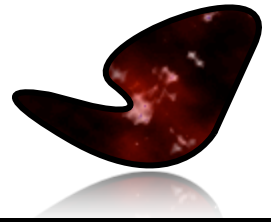
The Cosmic Emu(lator)

- **Prediction tool for matter power spectrum has been built**
- **Accuracy within specified priors between z=0 and z=1 out to k=1 h/Mpc at the 1% level reached**
- **Emulator has been publicly released as a C code**
- **Next steps**
  - Extend k-range
  - Include more physics, e.g. neutrinos
  - Other statistics, e.g. shear spectrum

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

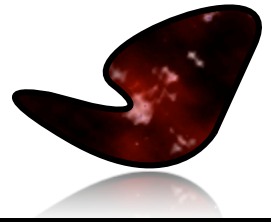


Summary and Outlook



- **Nonlinear regime of structure formation requires simulations**
  - No error controlled theory
  - Simulated skies/mock catalogs essential for survey analysis
- **Simulation requirements are demanding, but can be met**
  - Only a finite number of simulations can be performed
- **Cosmic Calibration Framework**
  - Accurate emulation of several statistics, accuracy comparable to simulation errors
  - Allows fast calibration of models vs. data
- **Challenges for the future**
  - More physics needs to be taken into account
  - Computational and storage capacities will be demanding
  - Simulation infrastructure, running very large number of simulation requires automation
  - Serving the data, simulation results should be available to broader community
  - Communication with other communities: statisticians, computer scientists, applied mathematicians, it takes some time to learn each other's language!

References



- **Gaussian Process Modeling Introduction:**
  - ***Gaussian Processes for Machine Learning,*** C.E. Rasmussen, K.I. Williams, MIT Press, www.GaussianProcess.org/gpml
- **Cosmic Calibration Framework (www.lanl.gov/projects/cosmology/CosmicEmu)**
  - ***The Coyote Universe III: Simulation Suite and Precision Emulator for the Nonlinear Matter Power Spectrum,*** E. Lawrence et al., ApJ submitted, arXiv:0912.4490
  - ***The Coyote Universe II: Cosmological Models and Precision Emulation of the Nonlinear Matter Power Spectrum,*** K. Heitmann et al., ApJ 705, 156 (2009), arXiv:0902.0429
  - ***Constraining Cosmological Parameters by Combining Simulations and Physical Observations,*** D. Higdon et al., in The Handbook of Applied Bayesian Analysis, Oxford University Press (2010)
  - ***Simulations and Cosmological Inference: A Statistical Model for Power Spectra Means and Covariances,*** M.D. Schneider et al., Phys. Rev. D78, 063529 (2008)
  - ***Cosmic Calibration: Constraints from the Matter Power Spectrum and the Cosmic Microwave Background,*** S. Habib et al., Phys. Rev. D76, 083503 (2007)
  - ***Cosmic Calibration,*** K. Heitmann et al., ApJ 646, L1 (2006)
- **Gaussian Process Modeling for other Astrophysics Problems**
  - ***Nonparametric Reconstruction of the Dark Energy Equation of State,*** T. Holsclaw et al., Phys. Rev. D.
  - ***Gaussian Process Modeling of Asteroseismic Data,*** B.J. Brewer, D. Stello, MNRAS 395, 22206 (2009)
  - ***Photometric Redshift Prediction Via Gaussian Process Regression,*** M. Way et al. ApJ 706, 623 (2009)