Christopher J. Miller
Asst. Professor,
Astronomy

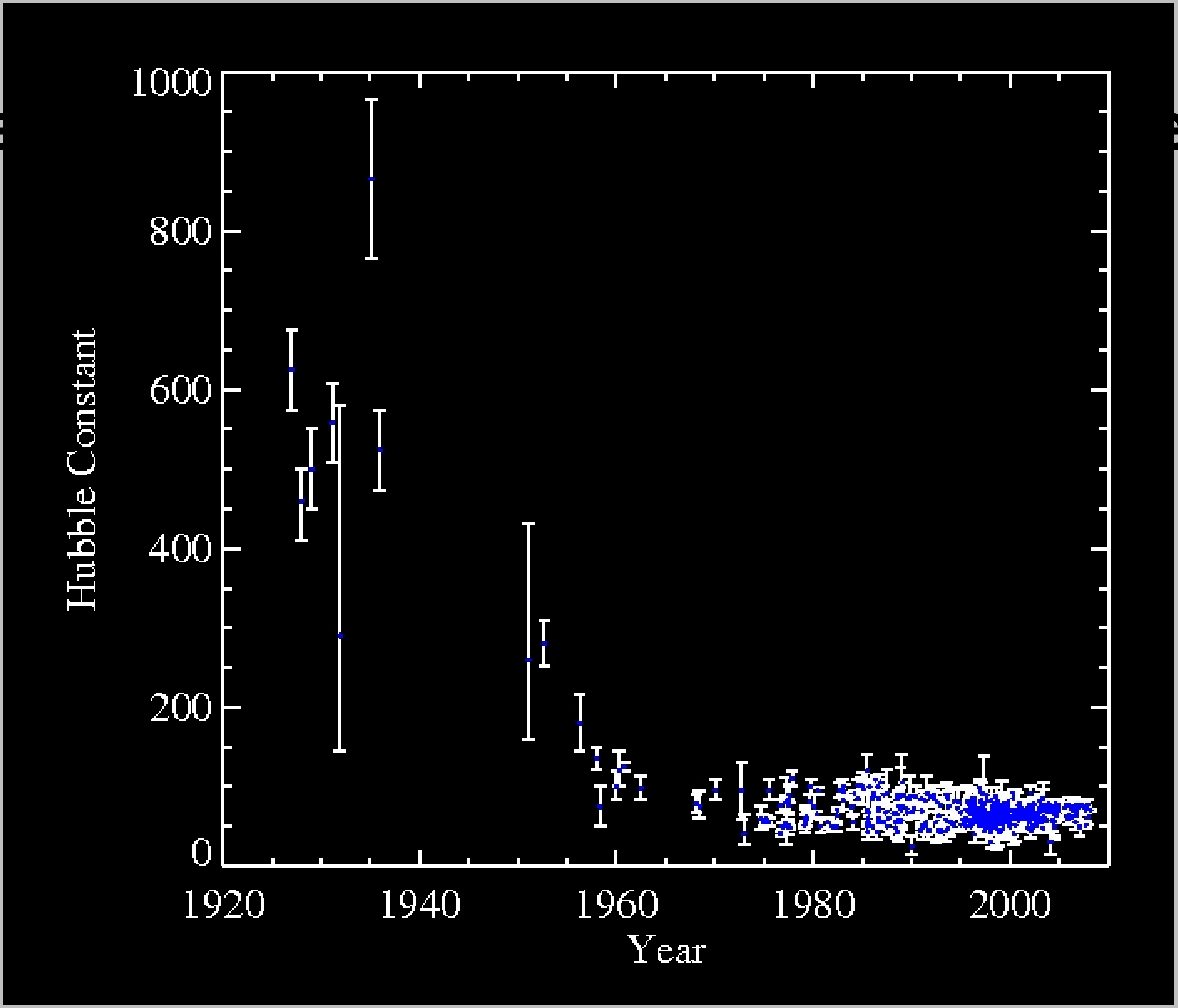
Computational, Statistical, and Mathematical Challenges in Astronomy

The Challenges

- **The Demand for Higher Precision Science**
  - The Hubble Constant

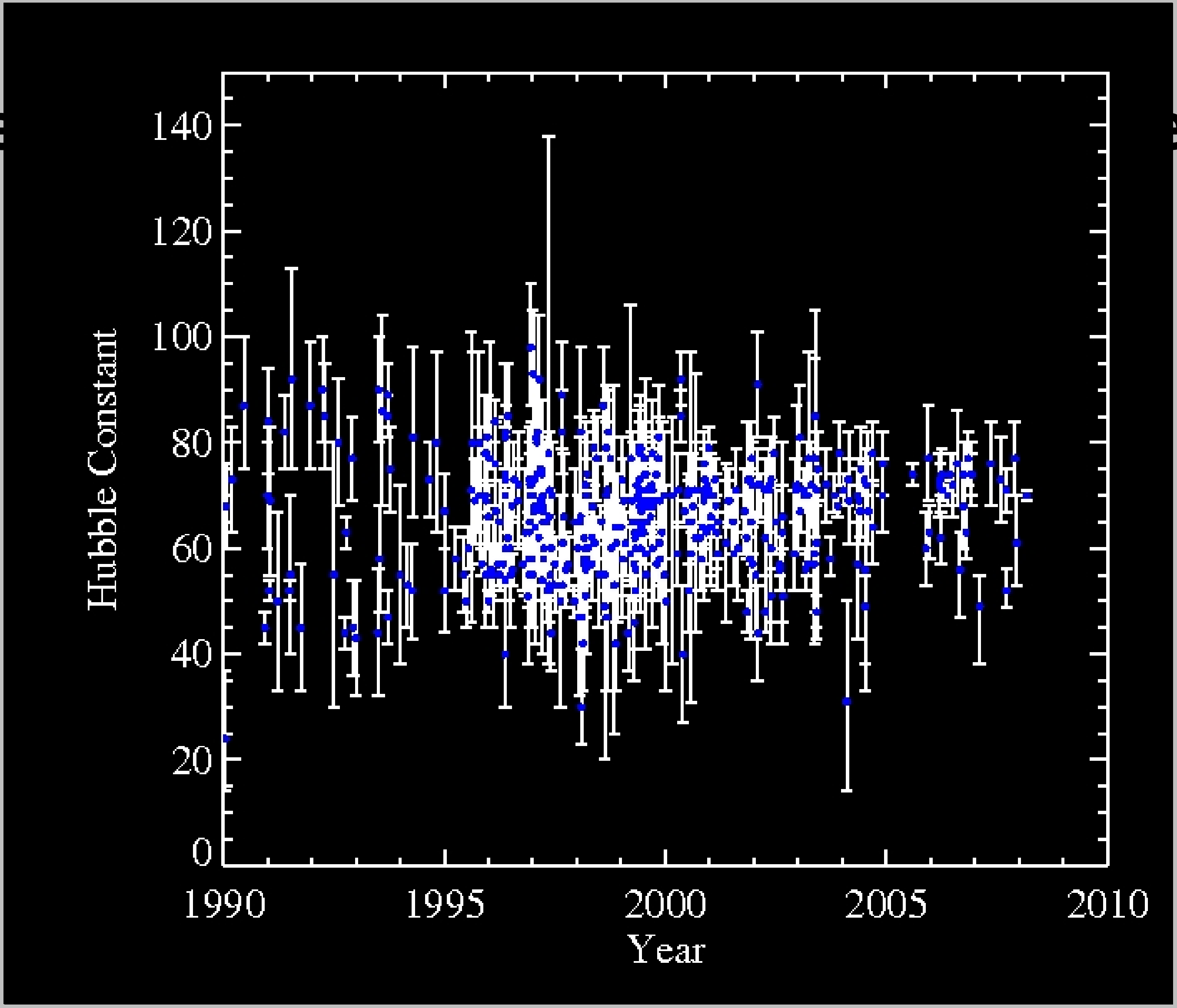
The Challenges

- **The De** ... **ence**
  - The



The Challenges

- **The De** ... **ence**
  - The



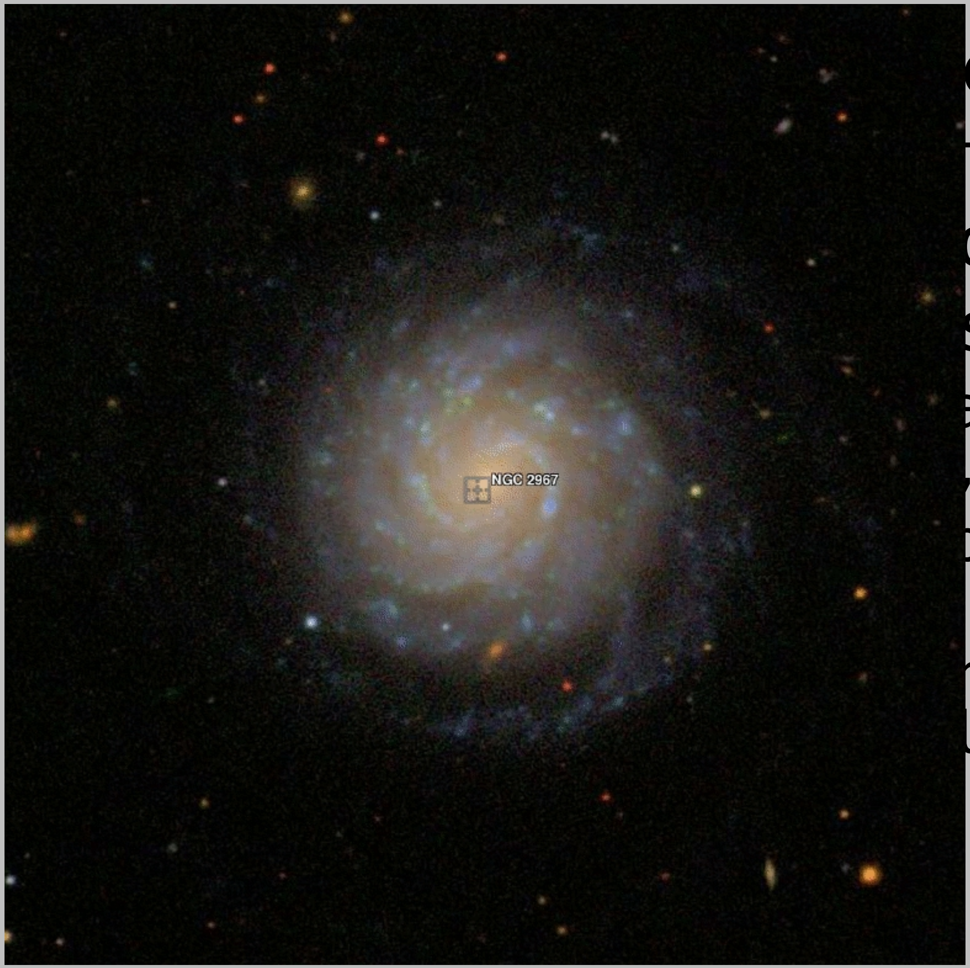
The Challenges

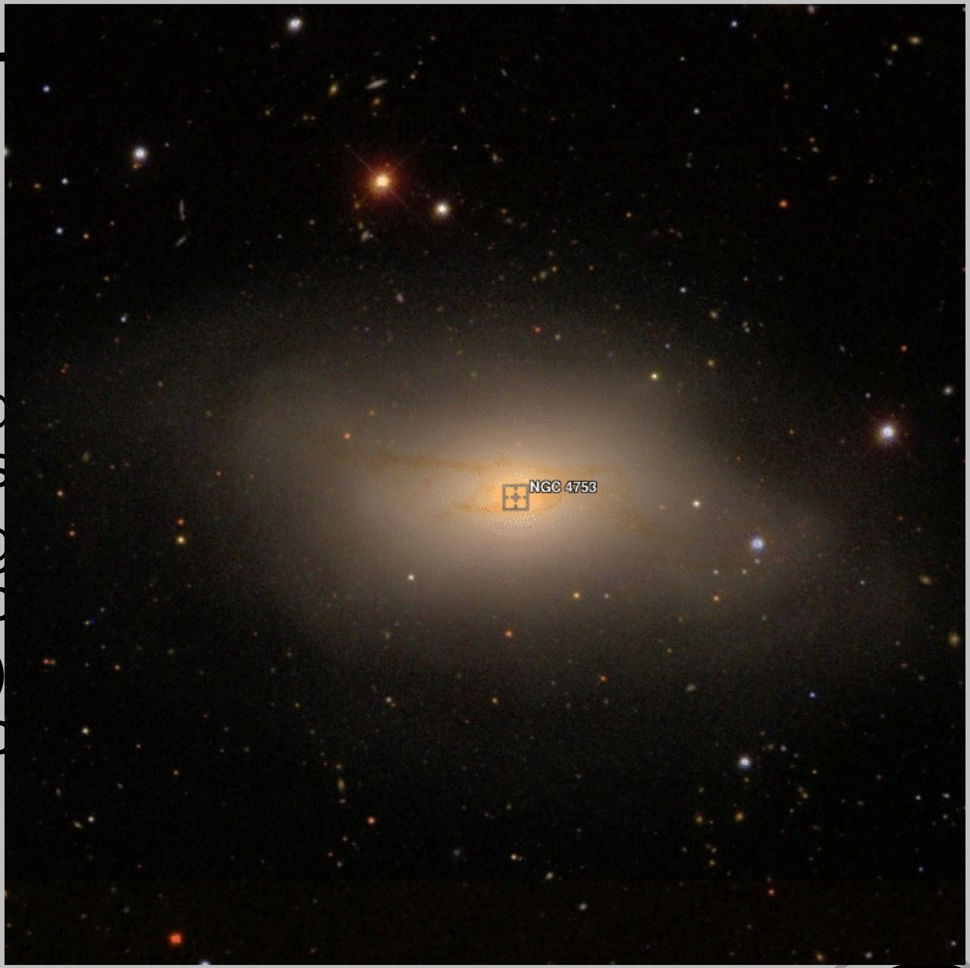
- **The Demand for Higher Precision Science**
  - The Hubble Constant
  - One of dozens of Cosmological Parameter

The Challenges

- **The Demand for Higher Precision Science**
  - The Hubble Constant
  - One of dozens of Cosmological Parameters
  - There is more to astronomy than cosmology
    - Galaxies have mass and stellar populations with ages, metallicities, star formation histories

The Challenges

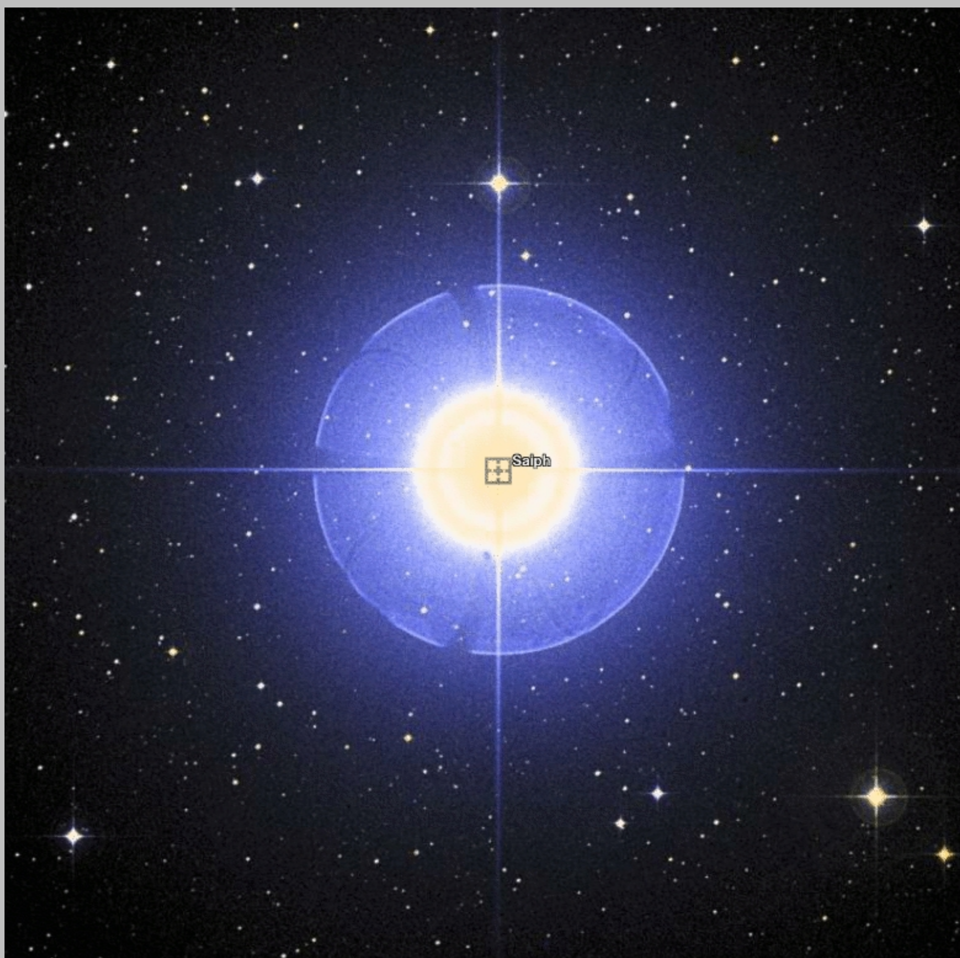


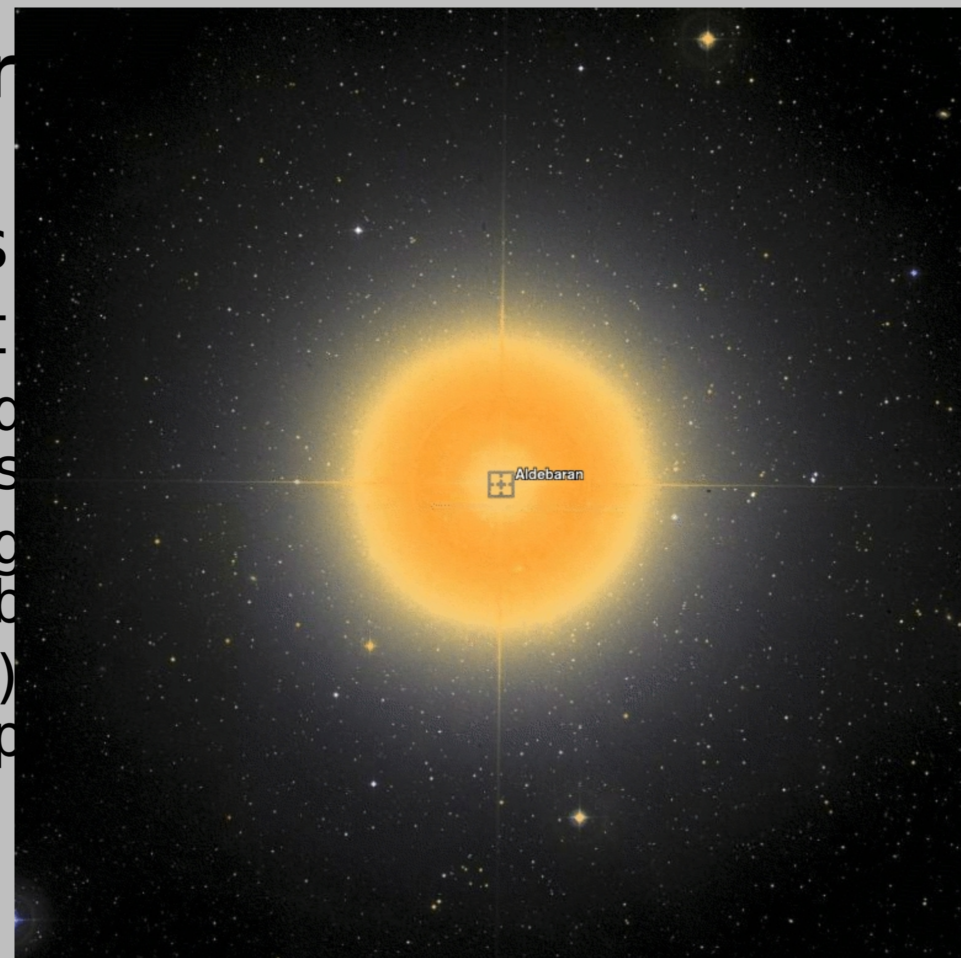


The Challenges

- **The Demand for Higher Precision Science**
  - The Hubble Constant
  - One of dozens of Cosmological Parameters
  - There is more to astronomy than cosmology
    - Galaxies have mass and stellar populations with ages, metallicities, star formation histories
    - Stars (which make up galaxies) have mass, temperatures, ages, abundances

The Challenges



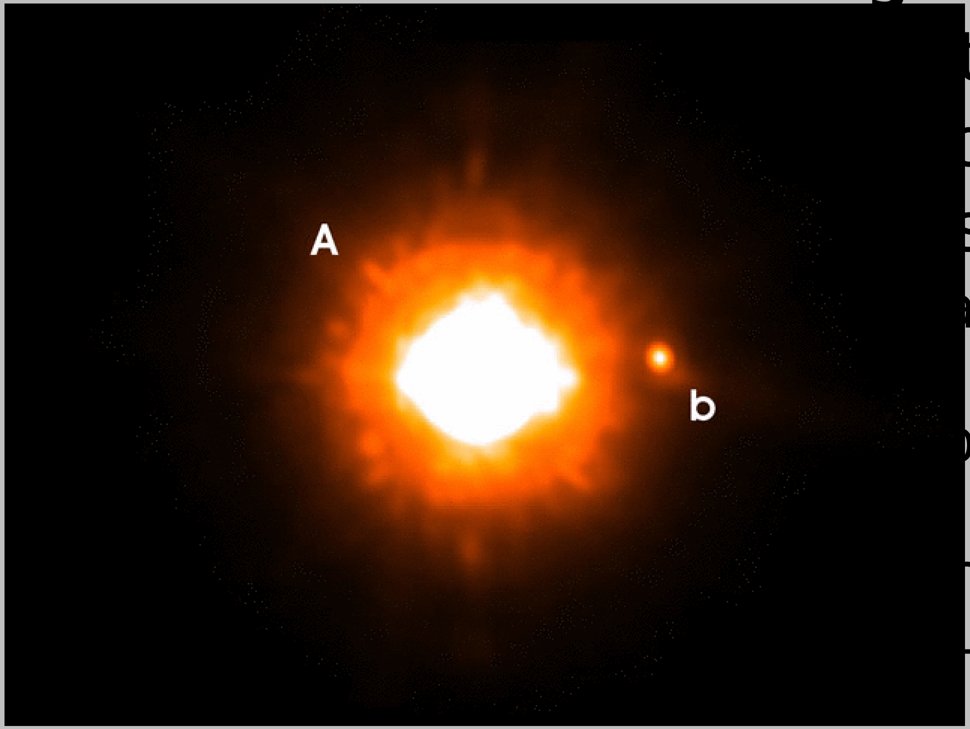


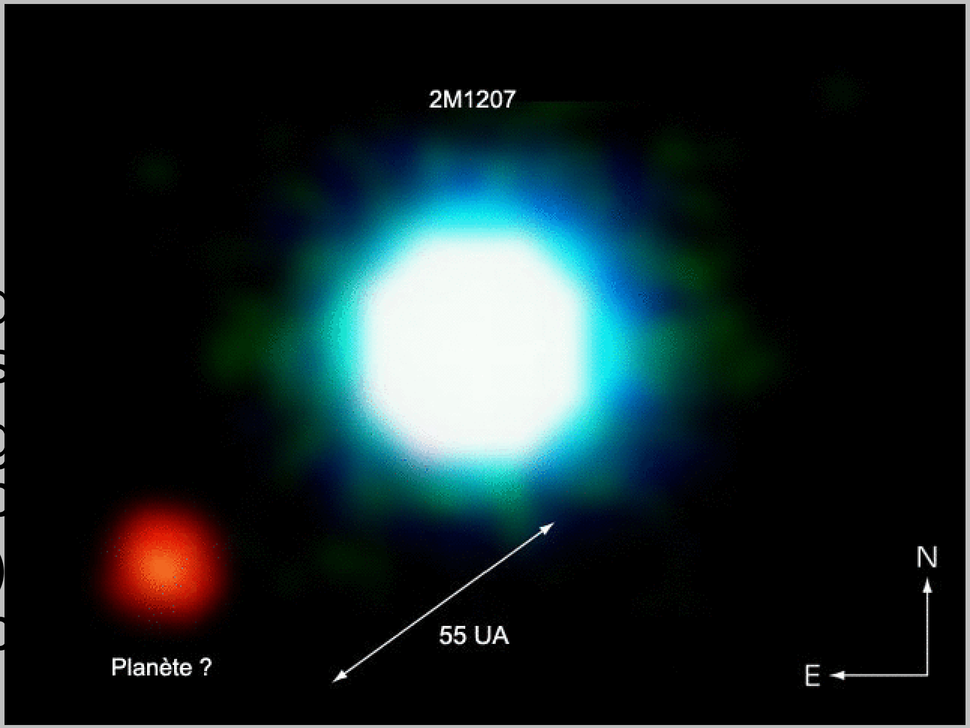
The Challenges

- **The Demand for Higher Precision Science**
  - The Hubble Constant
  - One of dozens of Cosmological Parameters
  - There is more to astronomy than cosmology
    - Galaxies have mass and stellar populations with ages, metallicities, star formation histories
    - Stars (which make up galaxies) have mass, temperatures, ages, abundances
    - Planets (around stars) have masses, compositions, atmospheres, orbital parameters

The Challenges

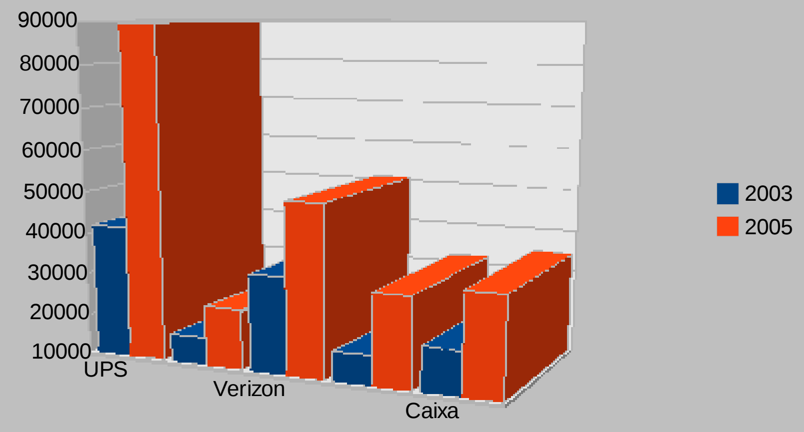
- **The Demand for Higher Precision Science**





The Challenges

- **The “Data Flood”**
  - Astronomical catalogs today contain about 1.5 billion objects (SDSS is ~300 million, IRSA is ~1 billion).
    - A factor of 20 smaller than the largest commercial DBs
  - LSST (~2020) will have ~50 billion objects
    - Large by today's standards. But average (or even small) by 2015
  - Astronomy has “Real World” DB challenges



Data volumes grow as well.....20 times increase from 2003-2005

The Challenges

- Astronomy Data is Distributed

The Challenges

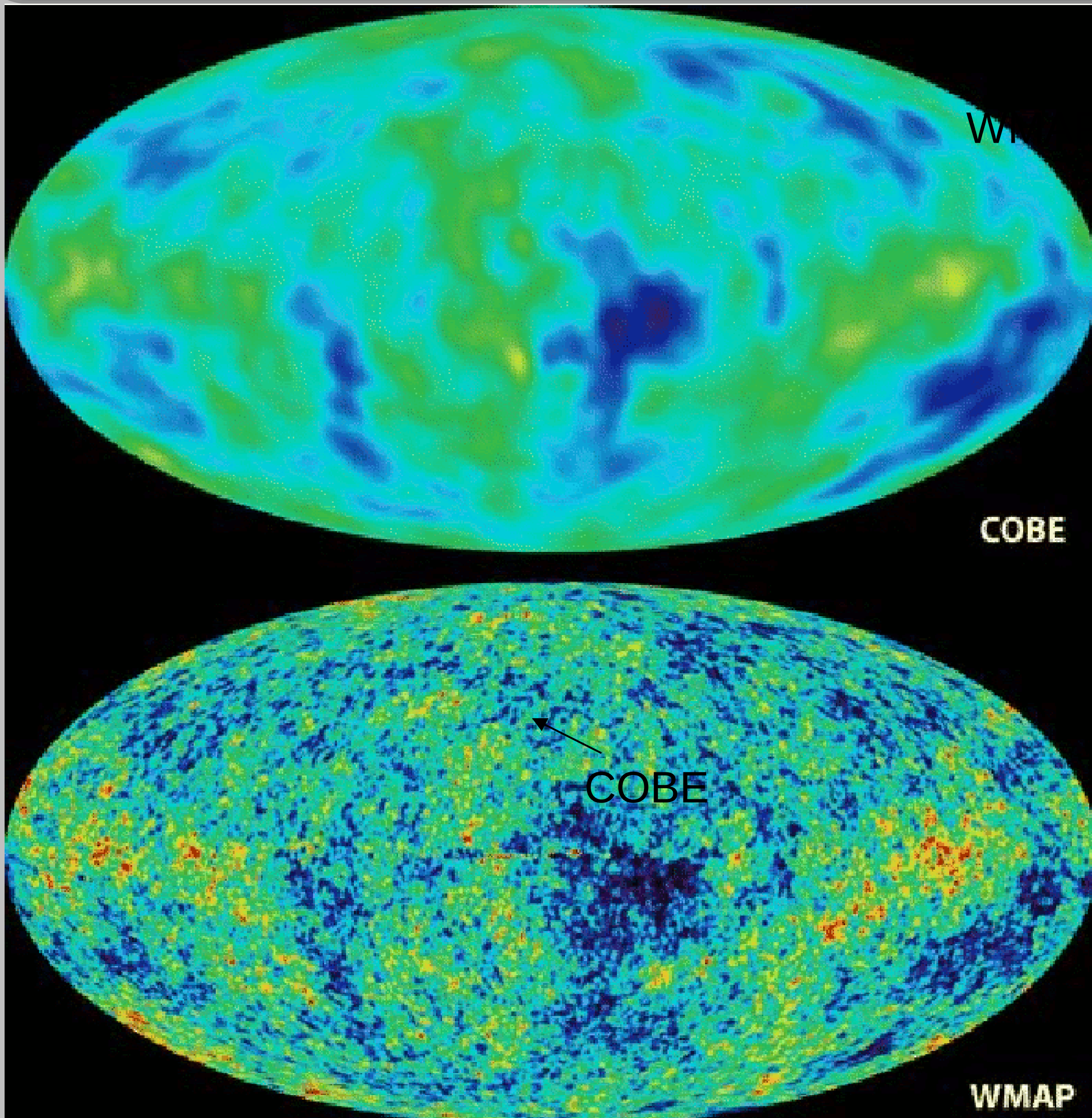
The Solutions

- Astronomical data creates opportunities for Computer Science, Information Technologies, and Statistics
  - Astronomy provides the (interesting) datasets, the distributed network, and the scientific questions
  - IT connects the network
  - CS handles the datasets and algorithms
  - MATHEMATICS and STATISTICS quantifies the answers

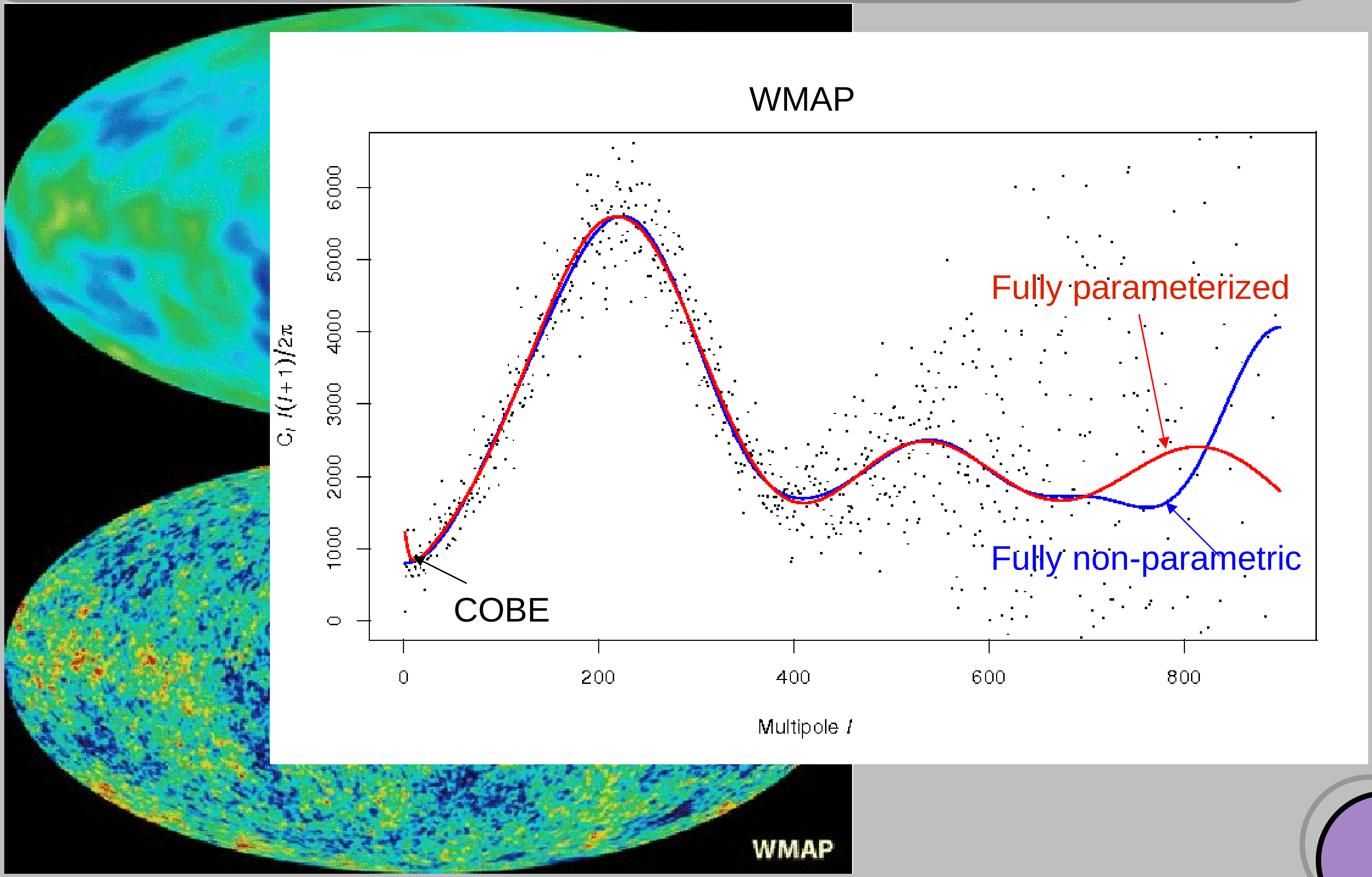
The Solutions

- Possible Detection of Baryonic Fluctuations in the Large-Scale Structure Power Spectrum: Miller, Nichol, Batuski 2001, ApJ
- Acoustic Oscillations in the Early Universe and Today: Miller, Nichol, Batuski 2001, Science
- Controlling the False Discovery Rate in Astrophysical Data Analysis: Miller, Genovese, Nichol, Wasserman, Connolly Reichart, Hopkins, Schneider, Moore, 2001 AJ
- A new source detection algorithm using FDR Hopkins, Miller, Connolly, Genovese, Nichol, Wasserman, 2002 AJ
- A non-parametric analyss of the CMB Power Spectrum Miller, Genovese, Nichol Wasserman, ApJ
- Non–parametric Inference in Astrophysics, Wasserman, Miller, Nichol, Genovese, Jang, Connolly, Moore, Schneider, 2002
- Detecting the Baryons in Matter Power Spectra Miler, Nichol, Chen 2002 ApJ
- Galaxy ecology: groups and low-density environments in the SDSS and 2dFGRS Balogh, Eke, Miller, Gray et al. 2002, MNRAS
- The Clustering of AGN in the SDSS Wake, Miller, Di Matteo, Nichol, Pope, Szalay, Gray, Schnieder, York 2004 ApJ
- Nonparametric Inference for the Cosmic Microwave Background Genovese, Miller, Nichol, Arjunwadkar, Wasserman. 2004 Annals of Statistics
- The C4 Clustering Algorithm: Clusters of Galaxies in the SDSS Miller et al. 2005 AJ
- The Effect of Large-Scale Structure on the SDSS Galaxy Three–Point Correlation Function Nichol et al. 2006, MNRAS
- Mapping the Cosmological Confidence Ball Surface Bryan, Schneider, Miller, Nichol, Genovese, Wasserman, 2007 ApJ
- Inference for the Dark Energy Equation of State Using Type Ia SN data Genovese, Freeman, Wasserman, Nichol, Miller 2008, Annals of Statistics

Example: Non-parametric fits of the CMB

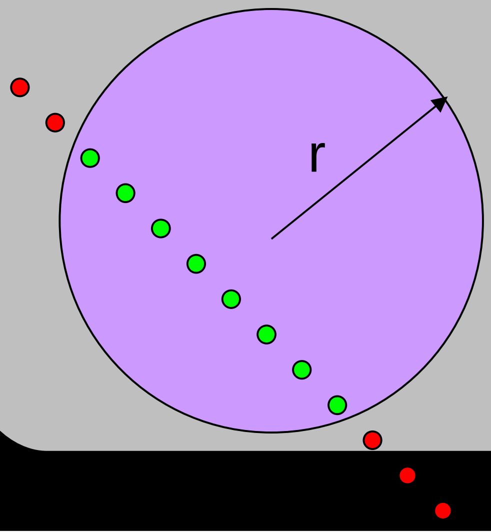


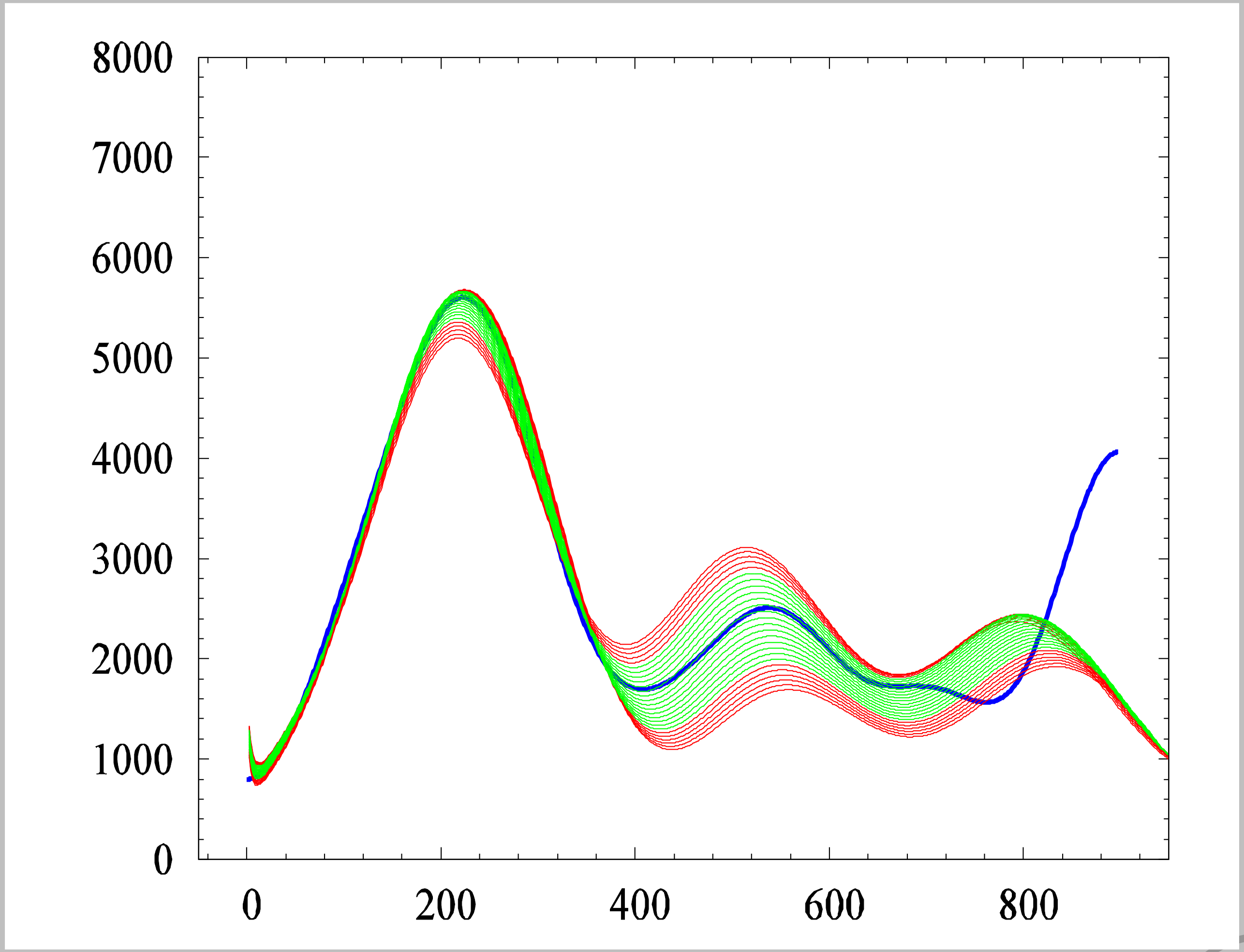
Example: Non-parametric fits of the CMB



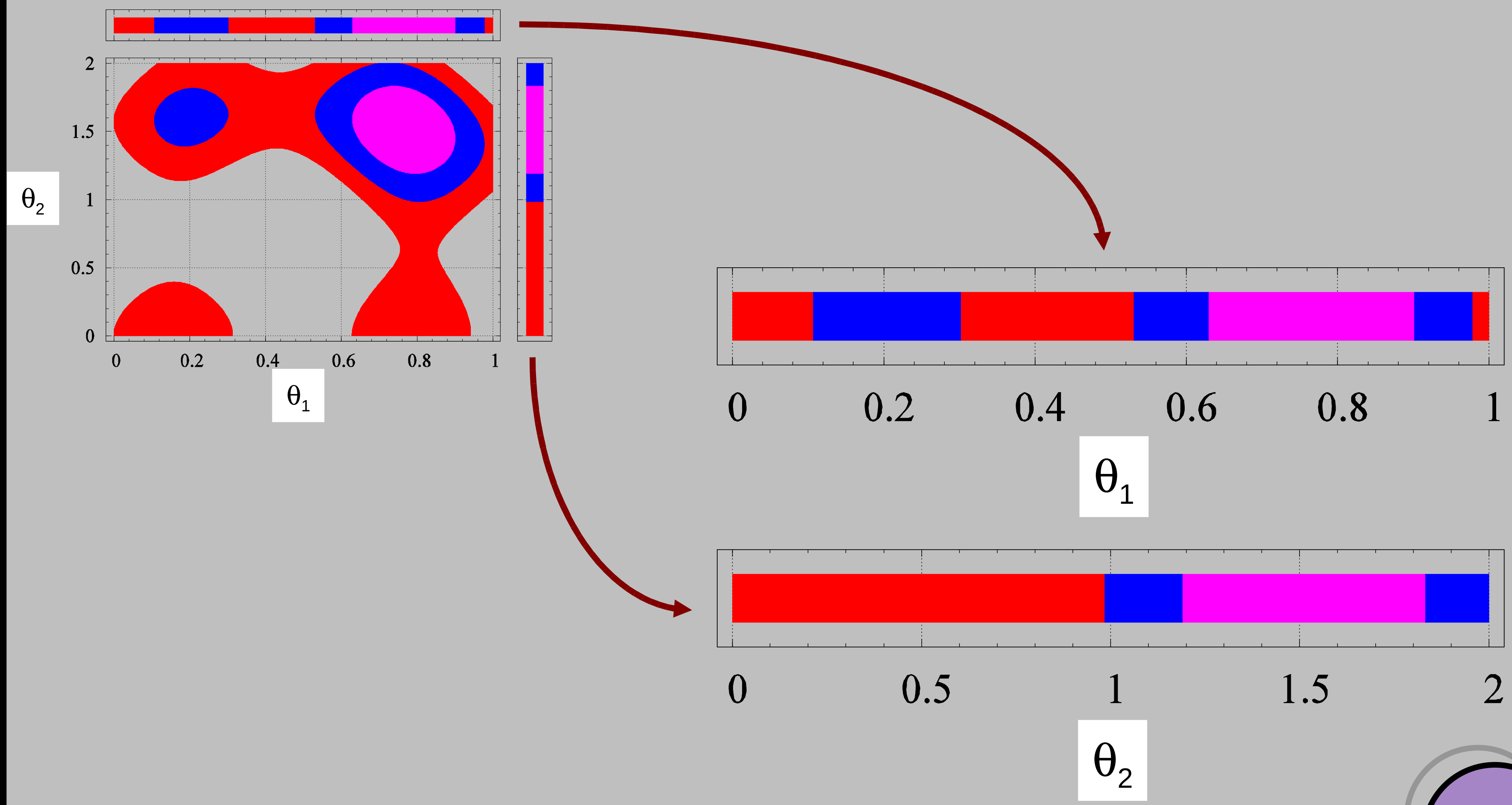
Confidence Balls: Pictorially

1. obtain experimental data
2. compute non-parametric fit
3. compute confidence ball
4. Iterate through parameters to determine confidence.

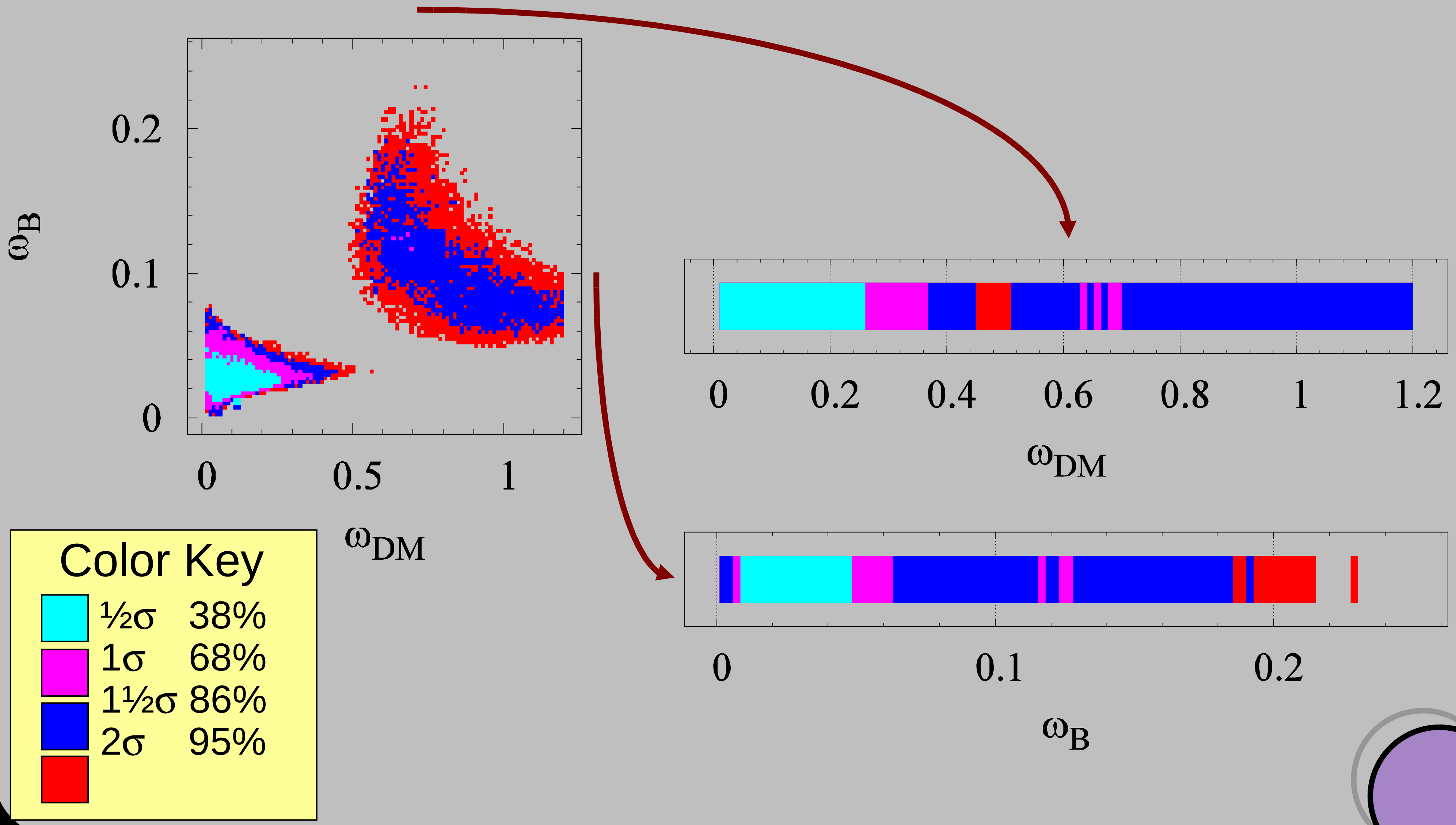




Deriving Confidence Intervals

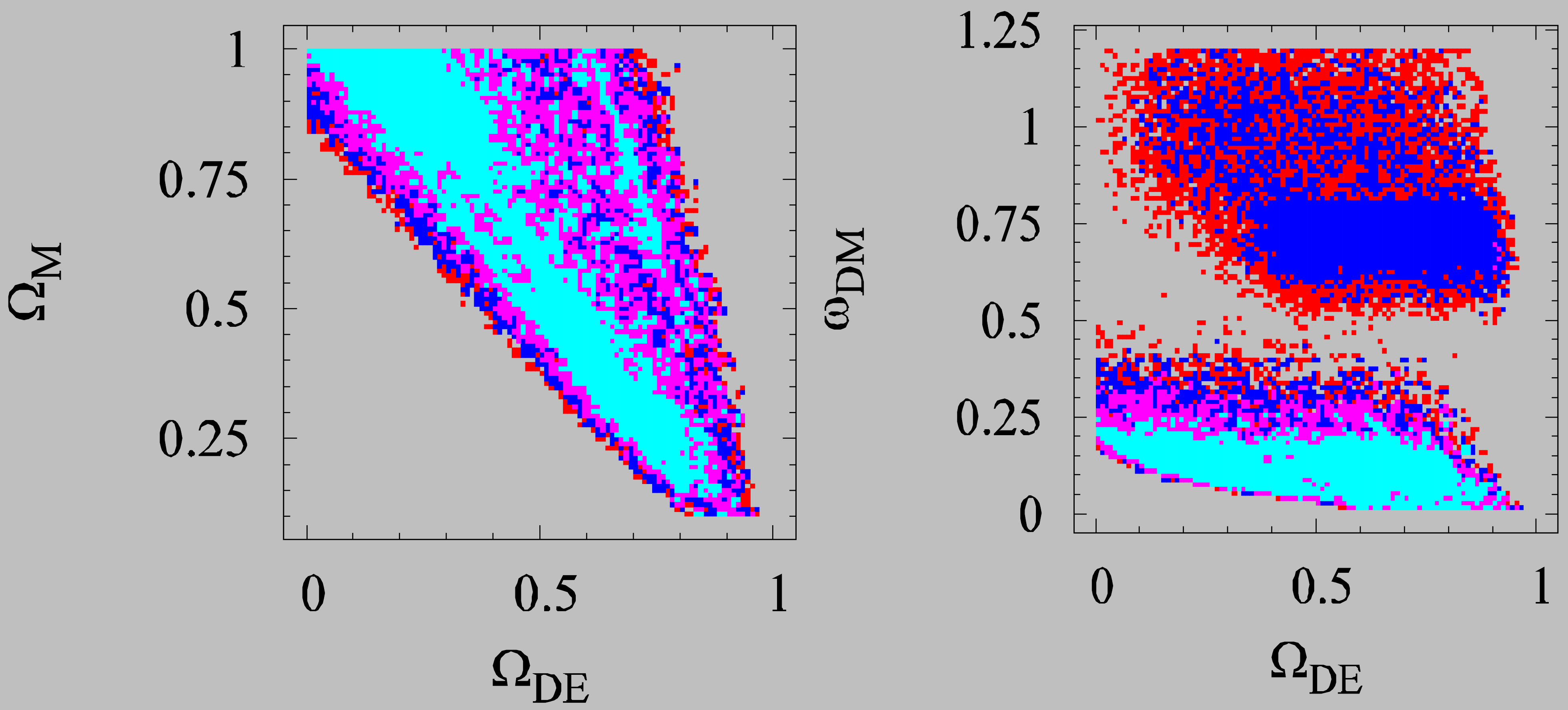


Cosmological Confidence Intervals



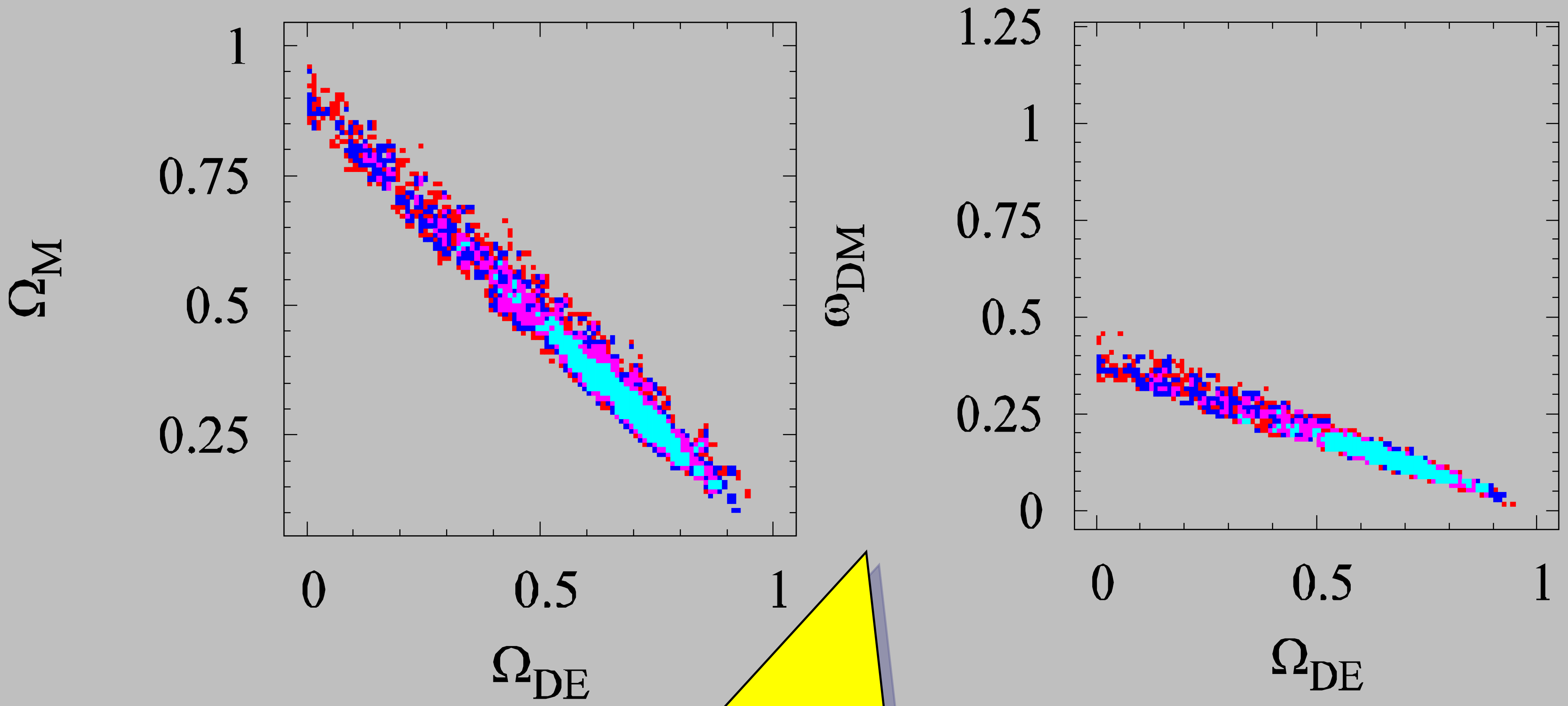
Including Assumptions

(Bryan et al., ApJ 2007)



Including Assumptions

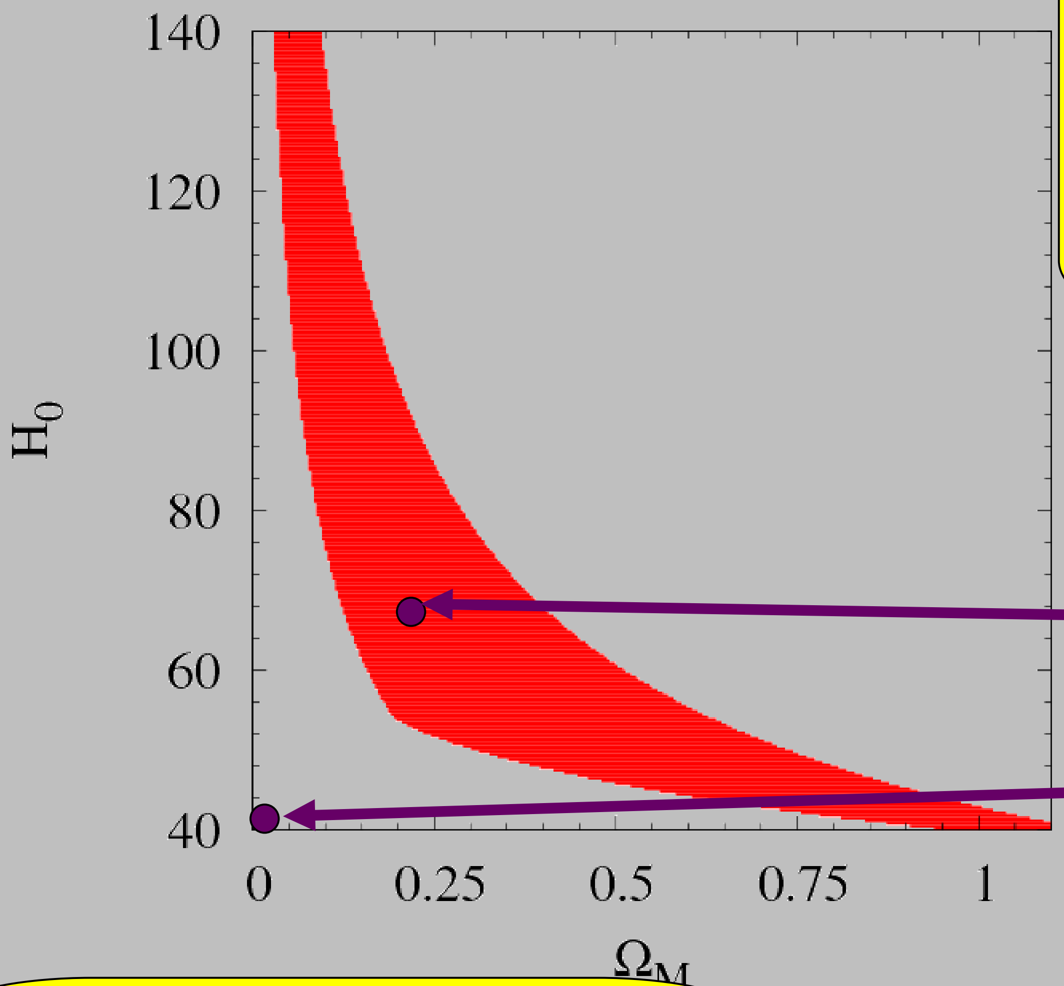
(Bryan et al., ApJ 2007)



Assumes $60 \leq H_0 \leq 75$ (km/s)/Mpc

What Does “Convergence” Mean?

To prove p_1 is within the confidence region:
$\exists\ \Omega_\Lambda$ such that $m(\theta)$ is accepted



- $\theta = \{H_0, \Omega_M, \Omega_\Lambda\}$
- $p_1 = \{65, 0.23, ?\}$
- $p_2 = \{0.02, 42, ?\}$

can’t check all Ω_Λ

s from based

To prove p is not within the region:
$\forall\ \Omega_\Lambda$, $m(\theta)$ is rejected

A Summary of the INCA Group Activities in Astronomy

- Originated at Carnegie Mellon University and the University of Pittsburgh (PiCA Group).
  - Membership expanded and members moved so that we changed the name to the INternational Computational Astrostatistics Group).
- A loose group of committed researchers (no formal structure)
- Astronomy provides the data and drives the science
  - Real work is done in developing, proving, and applying novel statistical methods and computational algorithms to astronomical datasets
  - Success is shared equally amongst the domains
- What are we interested in?

A Summary of the INCA Group Activities in Astronomy

- Originated at Carnegie Mellon University and the University of Pittsburgh (PiCA Group).
  - Membership expanded and members moved so that we changed the name to the INternational Computational Astrostatistics Group).
- A loose group of committed researchers (no formal structure)
- Astronomy provides the data and drives the science
  - Real work is done in developing, proving, and applying novel statistical methods and computational algorithms to astronomical datasets
  - Success is shared equally amongst the domains
- What are we interested in?
  - Parametrics
    - Dis-entangling multiple -components via Expectation Maximization
  - Nonparametrics
    - Reducing the size of the error ellipse
  - Non-linear SVM-like spaces
    - Focusing the available model space
  - High-dimensional searches and surface fitting
    - Constraining (as opposed to finding) the truth