Astrostatistics: Complex Models and Complex Questions

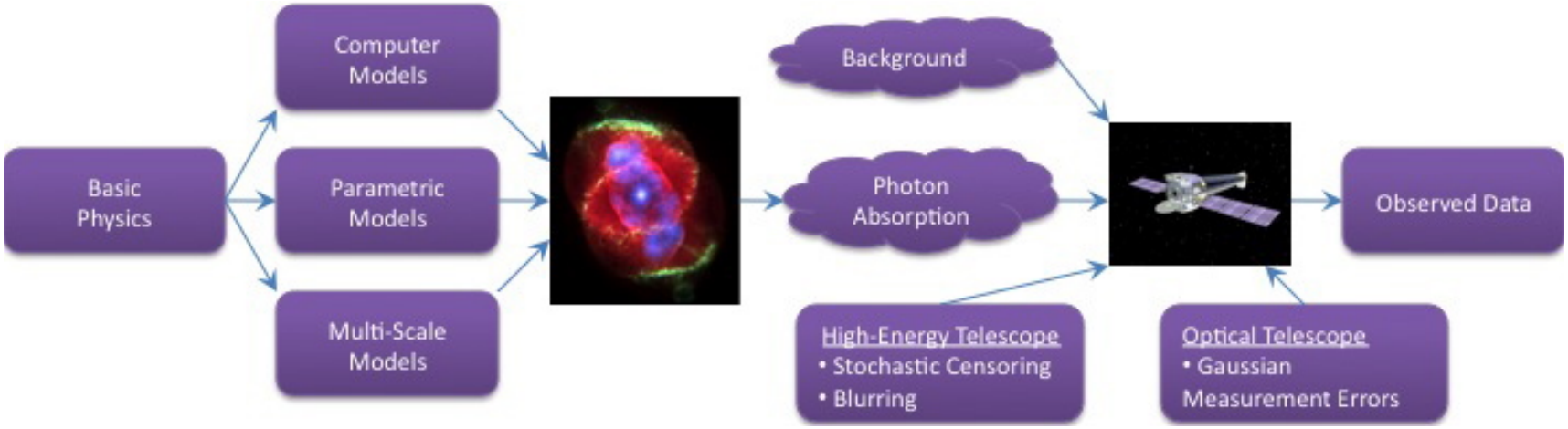
David A. van Dyk

University of California, Irvine

Workshop on Machine Learning and AI Applications
in Astrophysics and Cosmology, July 2009

UCIrvine

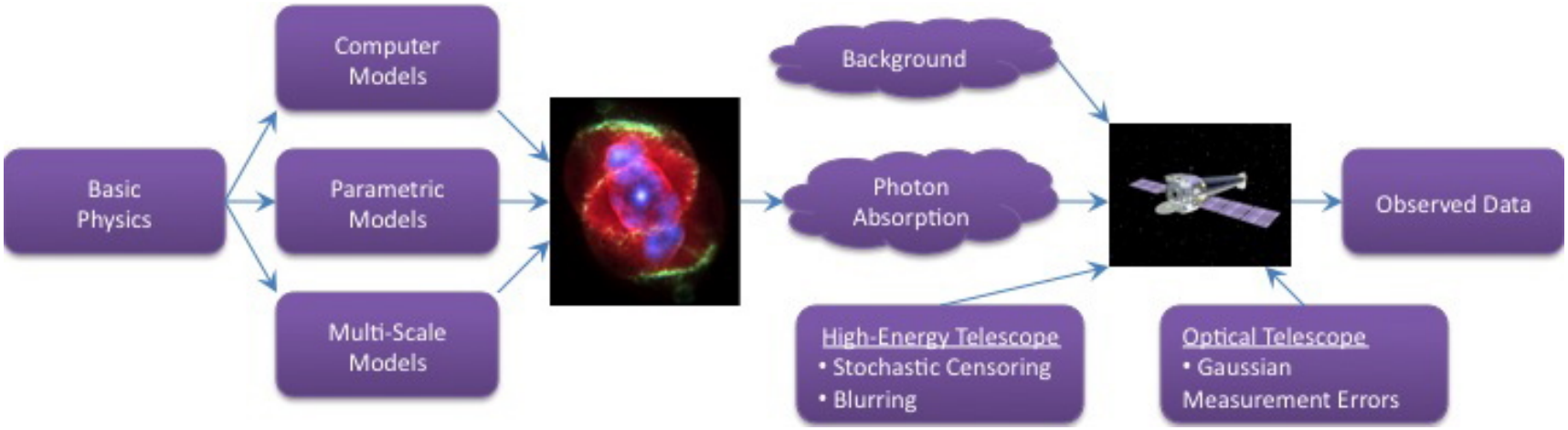
Physics-Based Generative Models



- Aim to formulate models in terms of specific questions of scientific interest.
- Must account for complexities of data generaration.
- Embed complex physics-based models into multi-level statistical models.

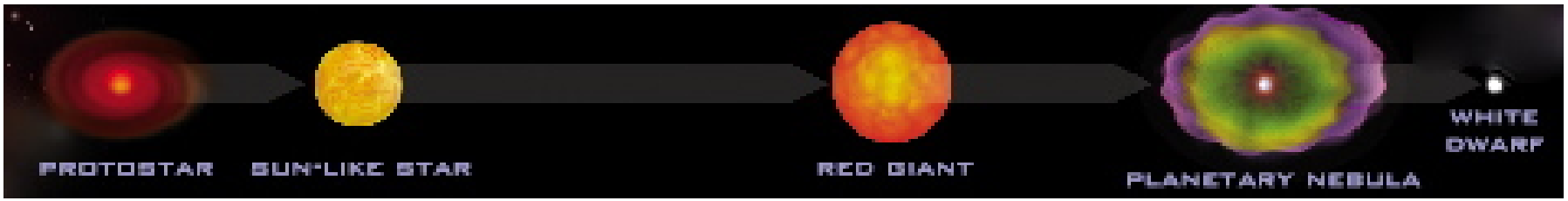
UCIrvine

Physics-Based Generative Models



- State of the art data enable us to fit the resulting complex model.
- This require sophisticated computational techniques.

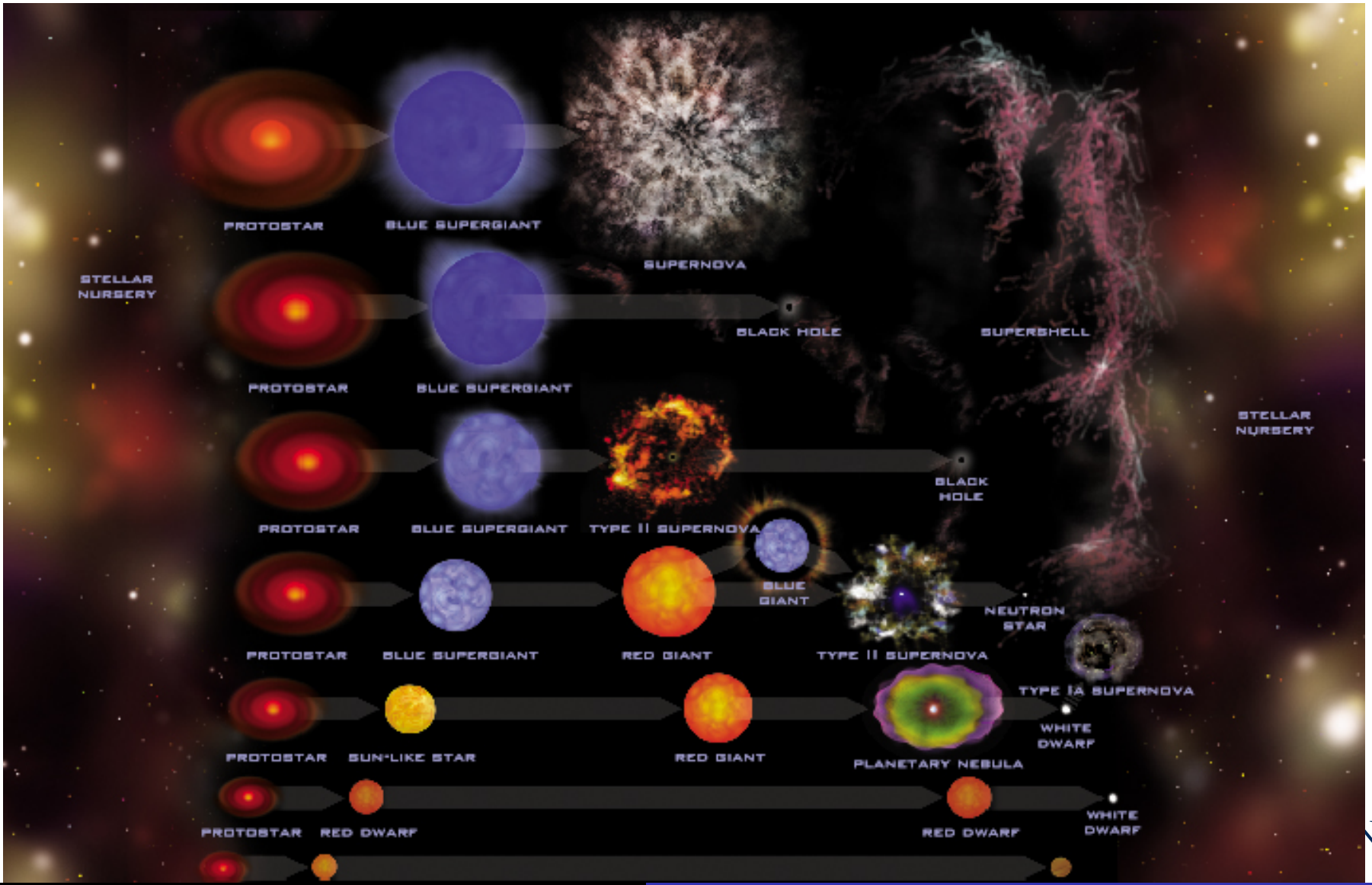
Stellar Evolution



- Sophisticated computationally-expensive physics-based computer models predict the magnitudes of a star given its age, composition, initial mass, distance, and absorption.
- Data are contaminated & include unresolved star systems.
- Misspecification of computer model complicates analysis.
- Highly non-linear correlations and multiple modes in a large dimensional parameter space pose significant computational challenges.

UCIrvine

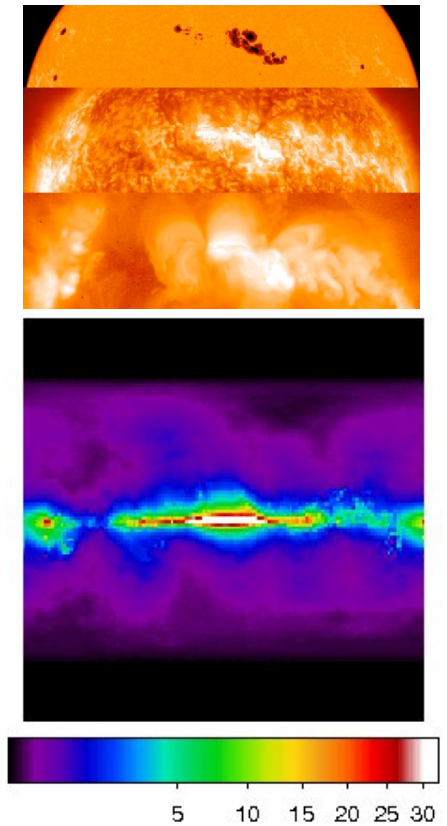
Stellar Evolution



Computer Models and External Information

Multi-level statistical models aim to directly model physical processes that generate observed data:

- Detailed quantum mechanical computations of the expected spectrum of a particular ion at a particular temperature,
- Measurements of the geometry, composition, and spectra of interstellar gas and its interaction with high-energy particles and low-energy light, or
- Measurement errors in high-dimensional instrument calibration.

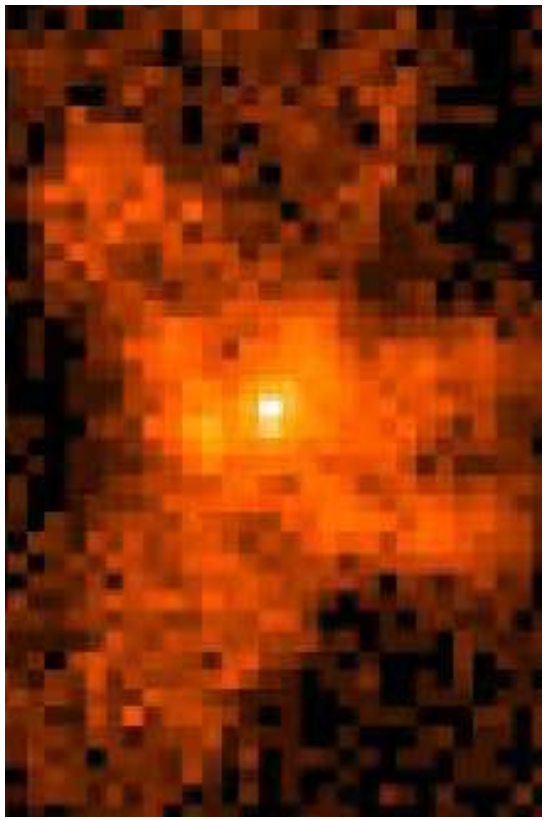


UCIrvine

Representing High Dimensional Uncertainty

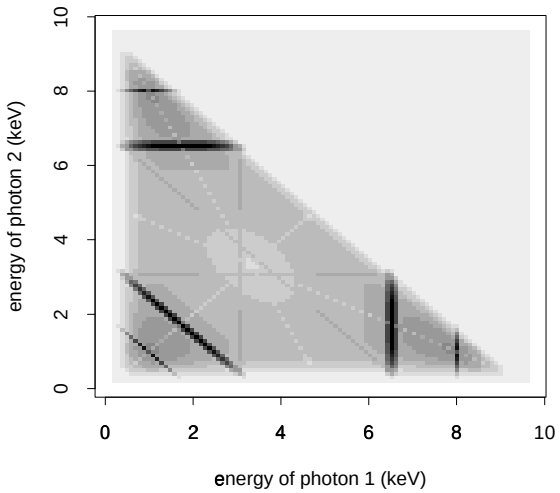
Representing and/or summarizing high-dimensional uncertainty in complex physics-based models poses real challenges:

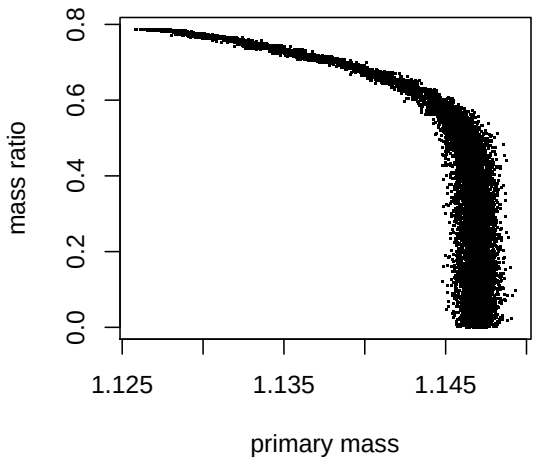
- Is a perceived structure in an image real?
- How can we quantify its statistical significance?
- Calibration quantities and quantum mechanical measurements are recorded with error, but correlations are unknown.

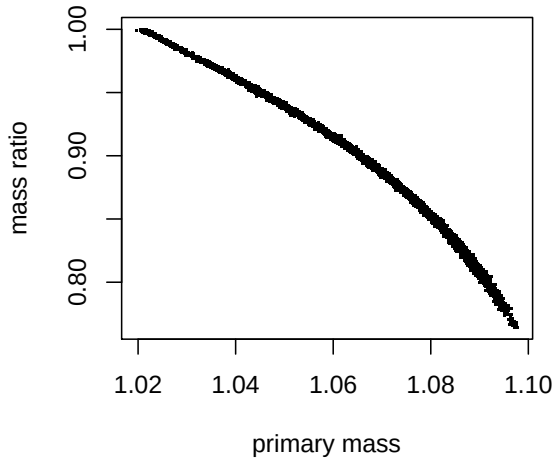


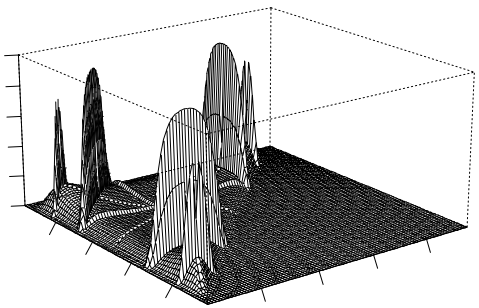
UCIrvine

Complex Likelihoods and Posterior Distributions









Difficult to explore, especially in high dimensions.

UCIrvine

The End

THANK YOU!

UCIrvine