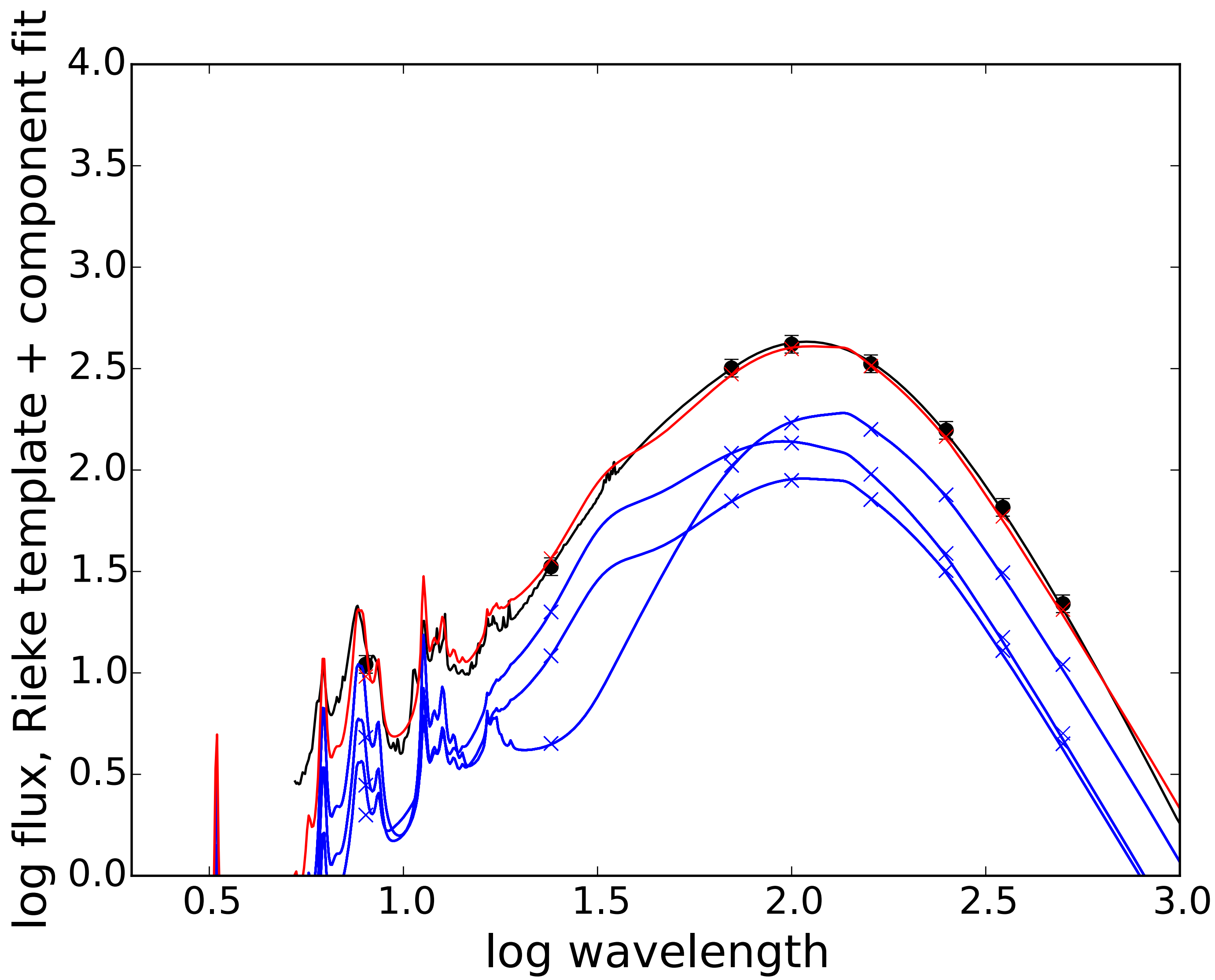
Sparse Composite Models of Galaxy Spectra (or Anything) using L1 Norm Regularization



Benjamin Weiner
Steward Observatory
University of Arizona

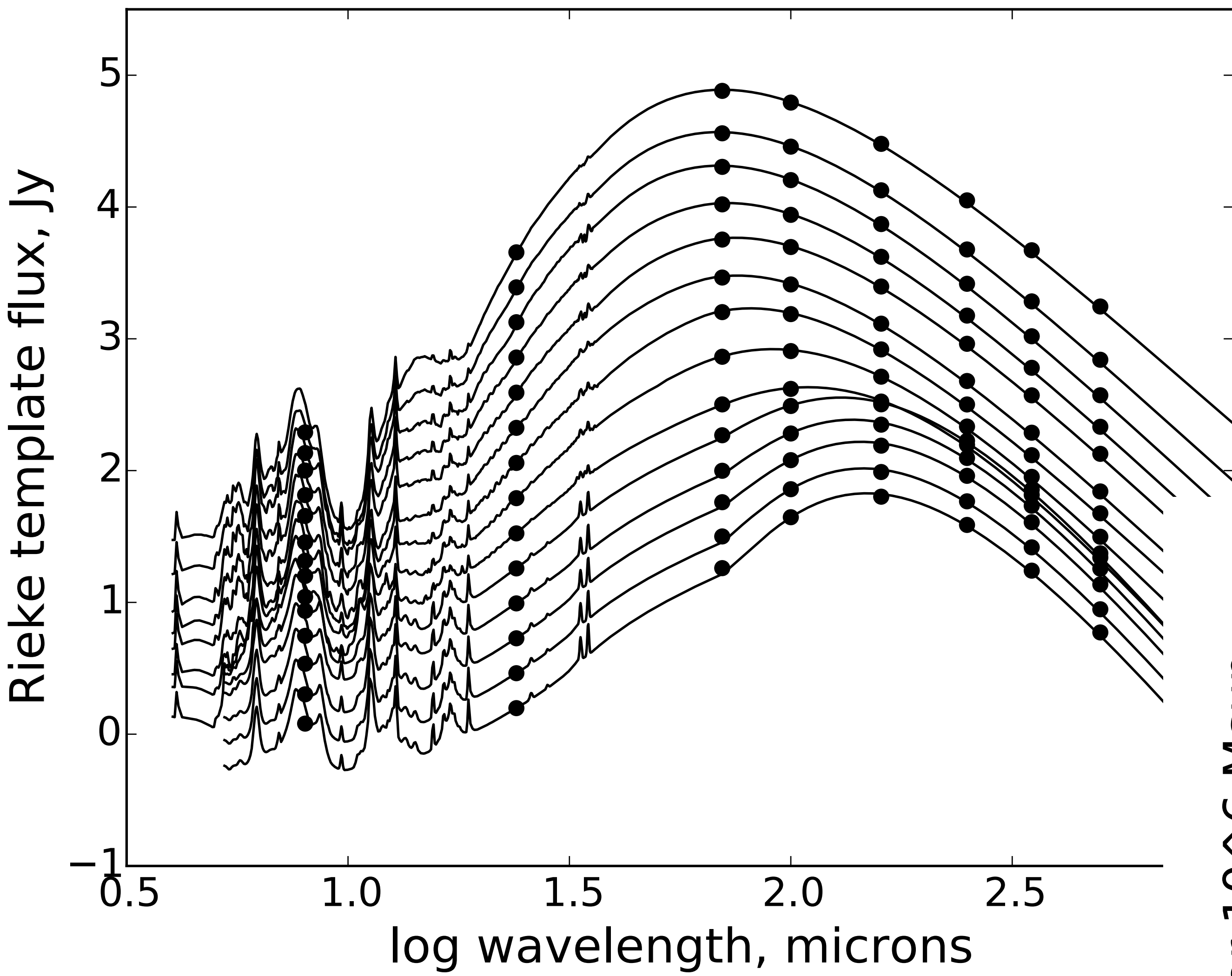
Inference of mass, redshift, etc. from physical models

Many astrophysical problems require fitting complex physical models to a limited set of observables, where the model parameters can be either the object of study or nuisance parameters. The parameter space is large and the generating process is typically non-linear and expensive.

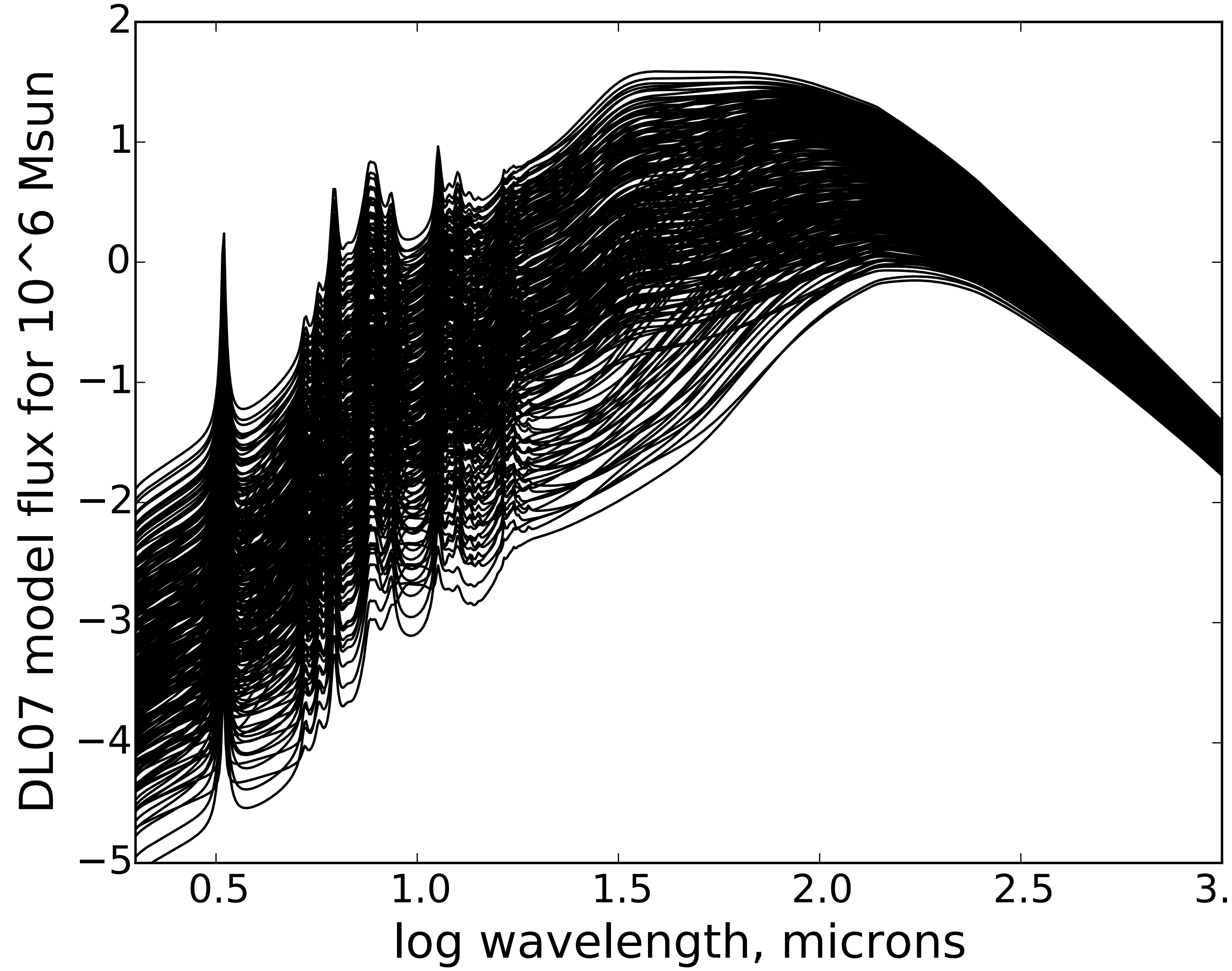
For a galaxy's stellar mass, redshift, star formation history, dust mass, etc., the usual approach is to take a very large set of possible models, e.g. star formation histories, generate stellar populations and thence colors from each model, and fit the models one at a time to the observed flux data, scaling in brightness (and sometimes dust extinction) only. This process has several problems:

- galaxies are intrinsically composite,
- systematics due to template mismatch,
- limited model space due to imposed forms of SF history, dust models, etc.

Galaxy dust mass from modeling far-infrared emission

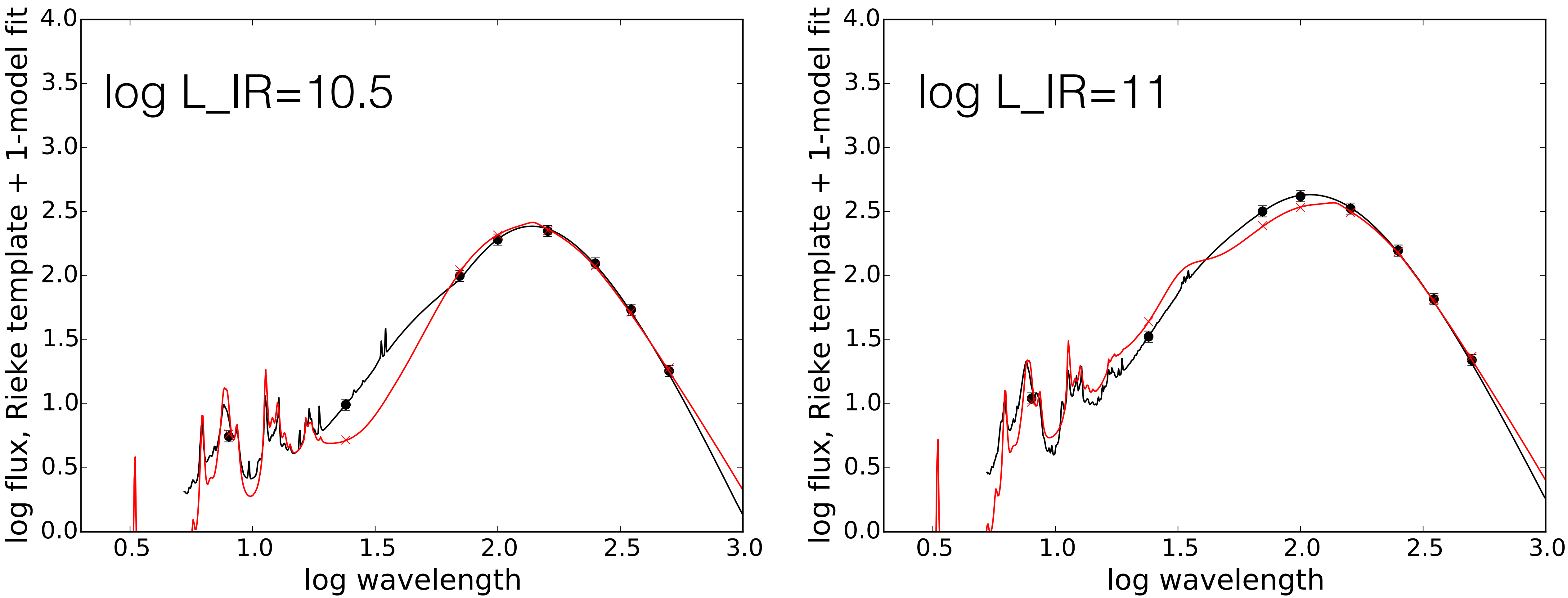


Data: Empirical galaxy far-IR spectra (Rieke et al 2009), which we “observe” at 8, 24, 70, 100, 160, 250, 350, 500 um



Models: 240 physically motivated dust model spectra, generated by different mixtures of radiation intensity, dust composition, etc (Draine & Li 2007)

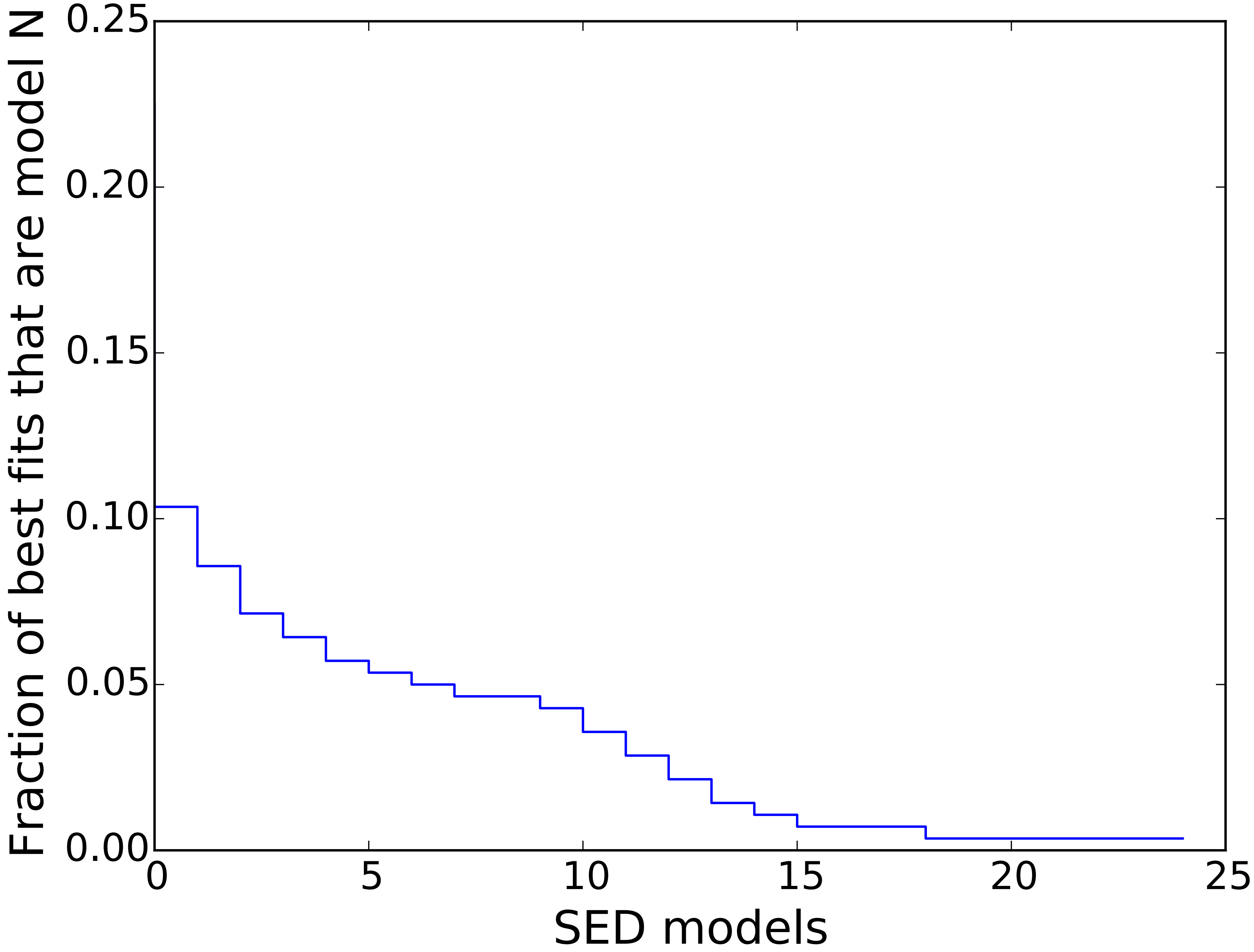
Fitting dust models one-at-a-time to simulated observations



Red line is best fit of the 240 Draine & Li dust models (themselves composites of a “diffuse” and “hot” component) to black points with 10% errors. Normalization -> dust mass (eg Magdis et al 2012, Magnelli et al 2012). Problems:

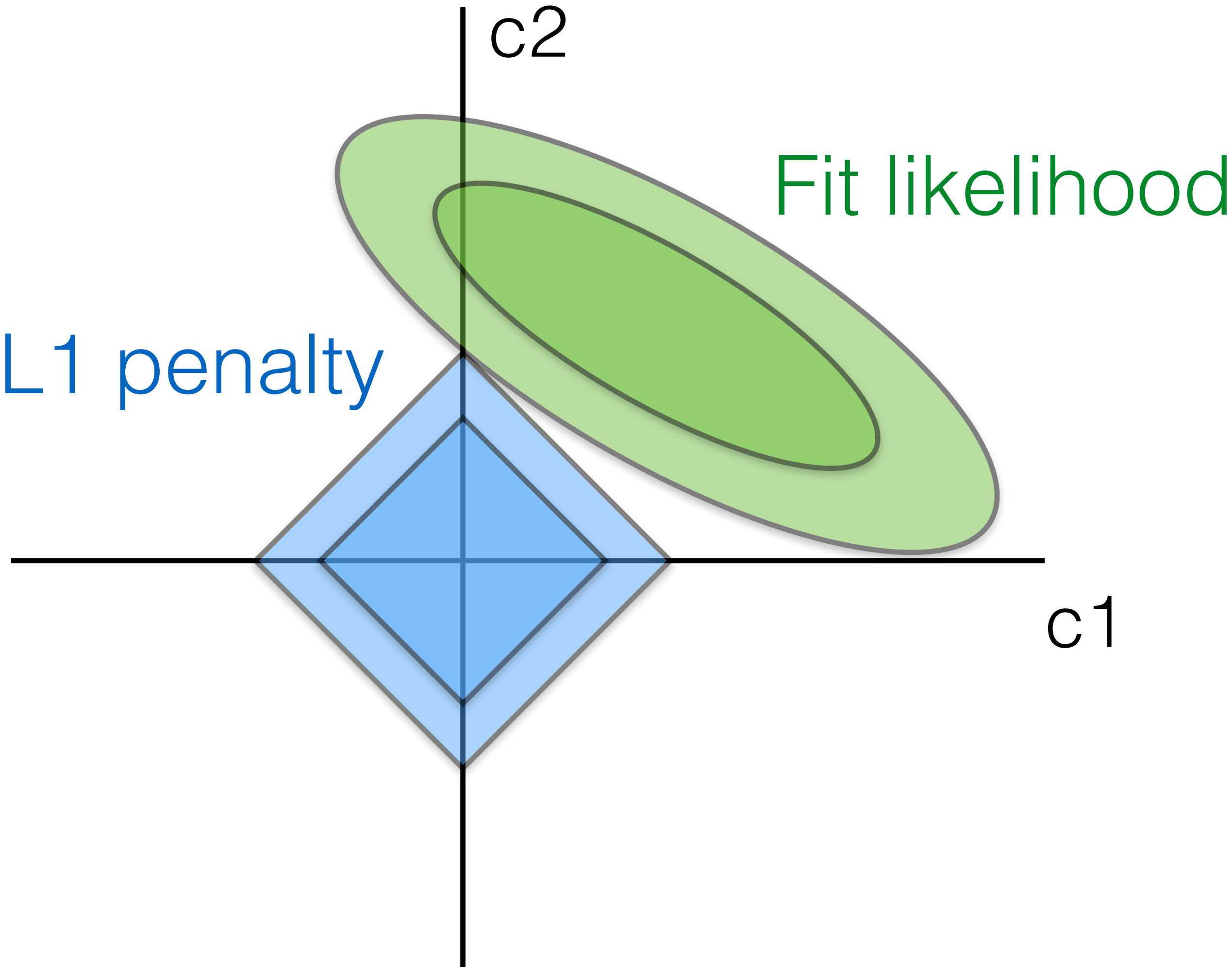
- Best fit still isn’t very good.
- Marginalizing over probability of each model is thus bogus.
- Marginalization and Monte Carlo realizations of error stick to a few models, and so underestimate the parameter uncertainties.

Lack of diversity in best-fit single model SEDs



Despite the seeming variety of 240 model SEDs, in the Monte Carlo realizations of 14 different far-IR spectra with 10% errors, nearly all of the best fits use only the 10 most common models.

How to fit linear combinations of models without going crazy



Minimize:

$$\sum_j \frac{(y_j - \sum_i c_i x_{i,j})^2}{\sigma_{y,j}^2}$$

$$+\lambda \sum_i c_i$$

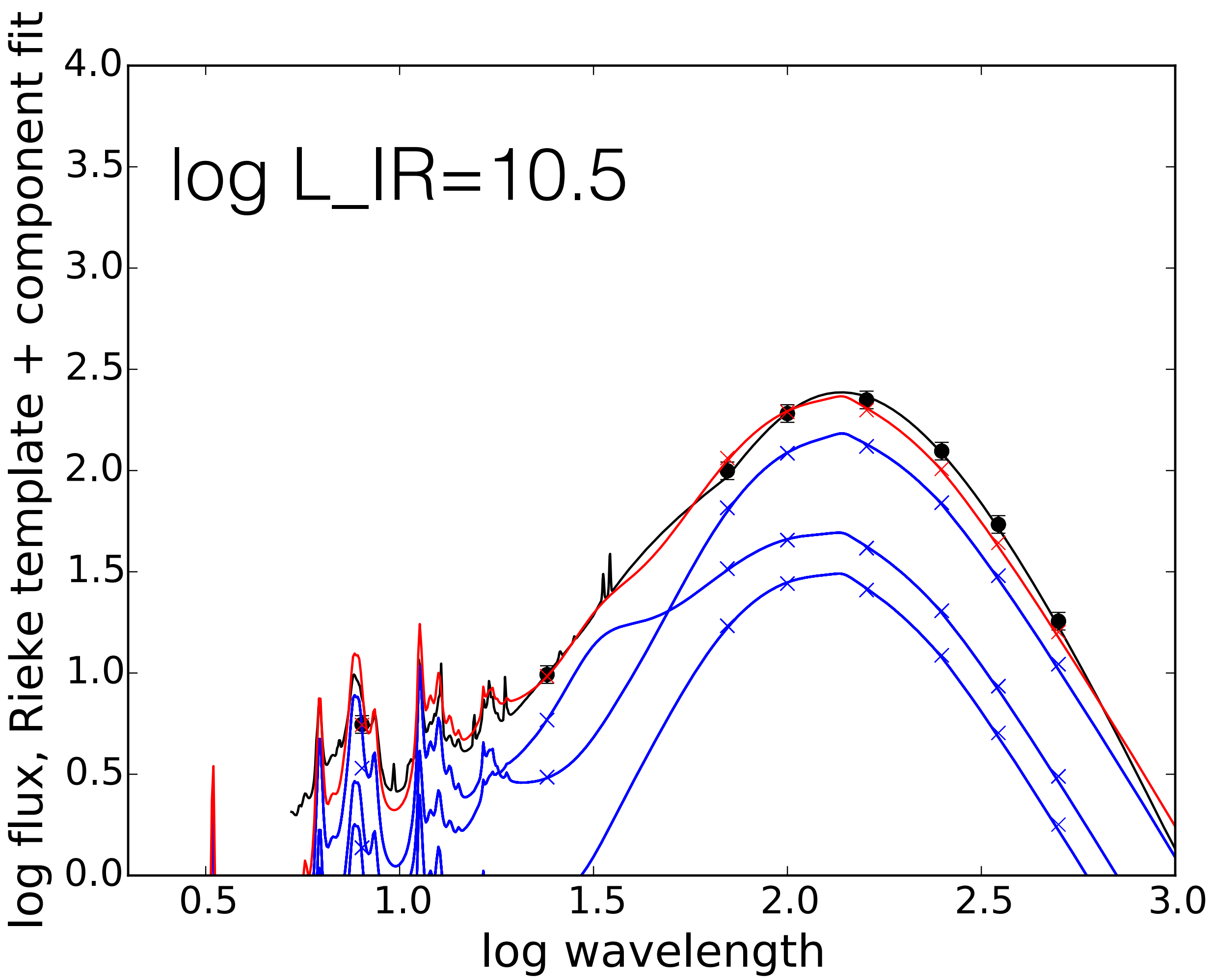
y_j are data

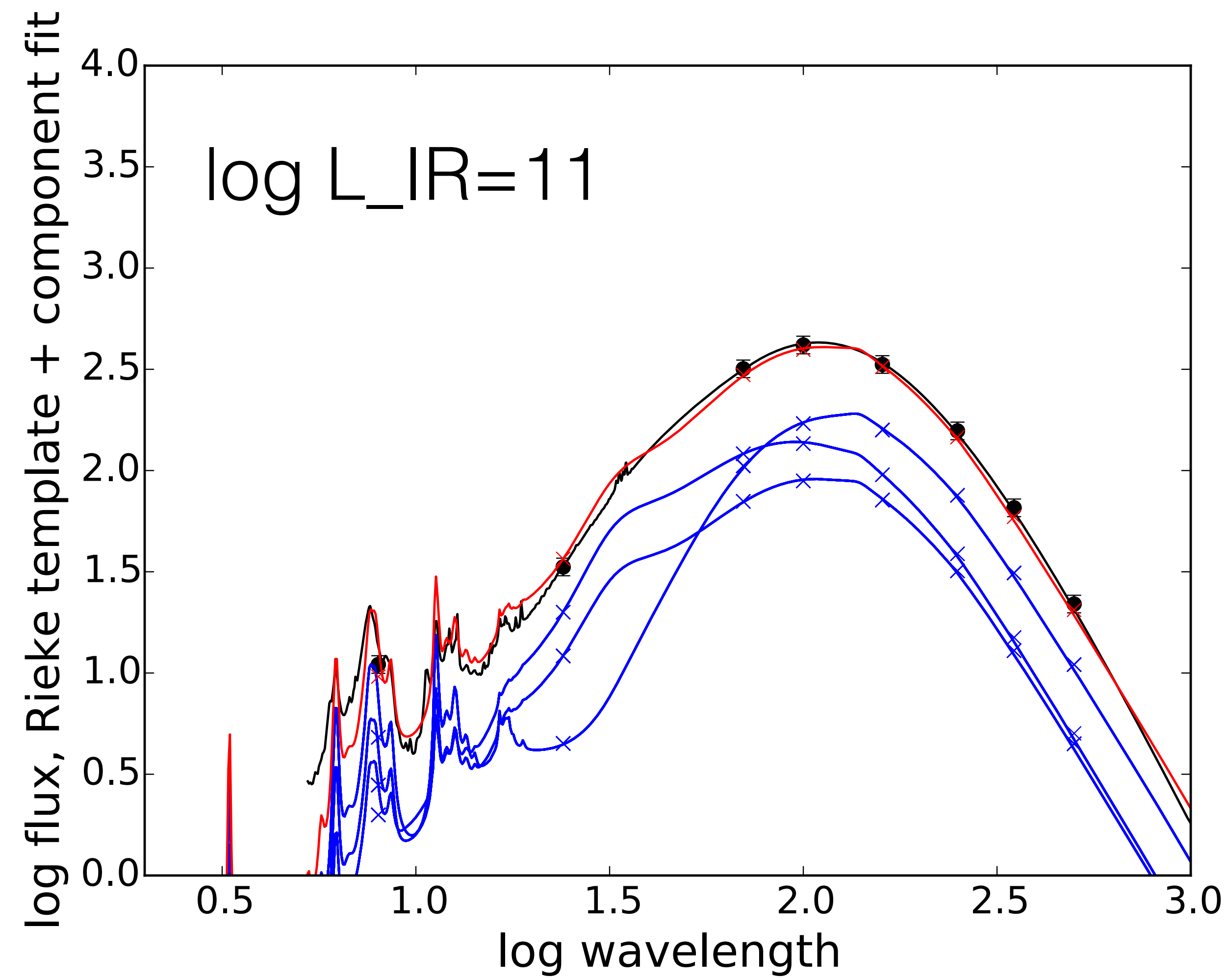
$x_{i,j}$ are i'th model

Penalize by sum of c's

Real galaxies are composites. A linear combination of models spans observable space better than one-at-a-time models, but combinations of 10s to 100s of models are highly degenerate. Applying an L1-norm penalty makes the fit sparse by keeping many coefficients at zero. (see LASSO, e.g. R. Tibshirani 1996, M. Schmidt 2005)

Fitting composite dust models with L1 regularization

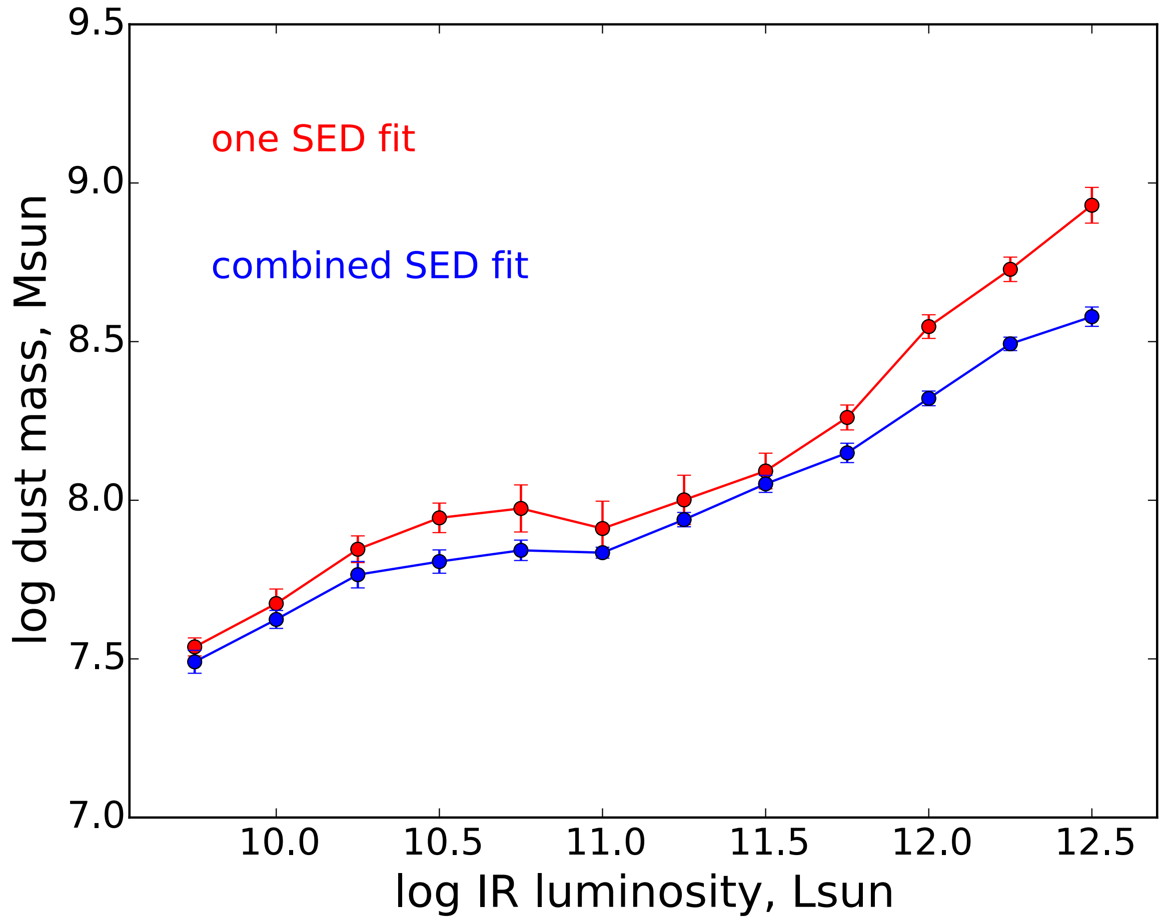




Blue lines are the fitted components. From an input set of the 9 most popular Draine & Li models, the L1-penalized fit left 6 at zero, and fit the data with only 3 components. Remaining issues:

- Overfitting: shallow minimum / fits are degenerate.
- This causes MC realizations to again underestimate the error on mass.
- Solution can be sensitive to the minimization algorithm; need to explore the acceptable region, e.g. with MCMC.

Composite models fit the data better, but their real value is showing the larger parameter space allowed by the data



The best-single-fit and composite models find similar trends of dust mass with far-IR luminosity, but with an offset. The offset is not huge, but is grossly larger than the conventional method of error estimate on the best-fits, because galaxies are not well described by single models (even though these DL models are already mixtures of diffuse+hot components).

The L1 regularization method is generally applicable to any data that is the sum of complex, physically meaningful models.

Thanks to David Hogg for suggesting the L1-norm as a technique to enforce sparsity, and George Rieke for discussions on infrared spectra and dust models.

References:

B. Draine & A. Li. 2007, ApJ, 657, 810
G. Rieke et al. 2009, ApJ, 692, 556
G. Magdis et al. 2012, ApJ, 760, 6
B. Magnelli et al. 2012, A&A, 548, 22
M. Schmidt. 2005, “Least squares optimization with L1-norm regularization,” https://www.cs.ubc.ca/~schmidtm/Documents/2005_Notes_Lasso.pdf
R. Tibshirani. 1996, “Regression shrinkage and selection via the lasso,” J. Roy. Stat. Soc. B, 58, 267