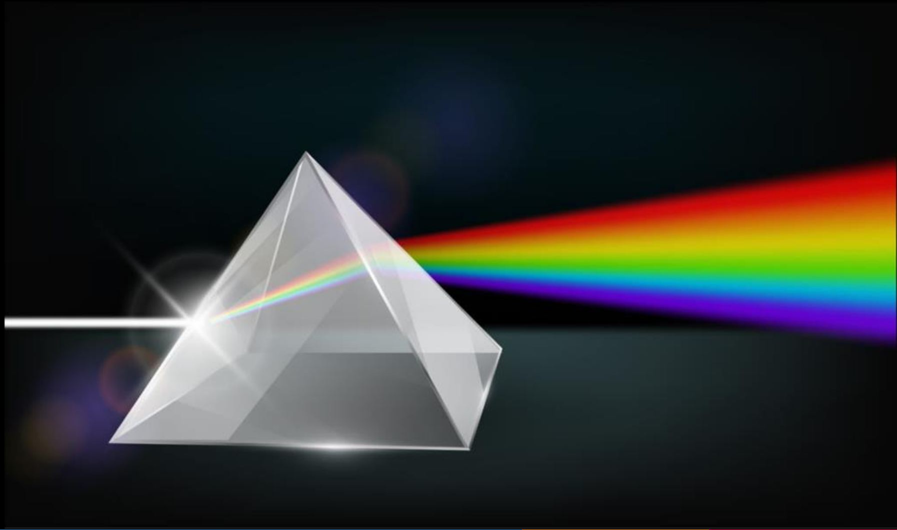
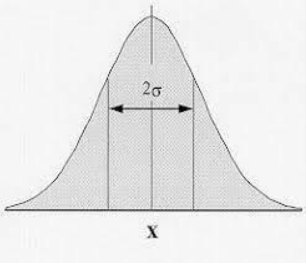
Challenges and opportunities in fitting high-resolution spectra

Anna Ogorzałek
NASA GSFC, UMD

Inspired by discussions with numerous colleagues at Goddard and elsewhere

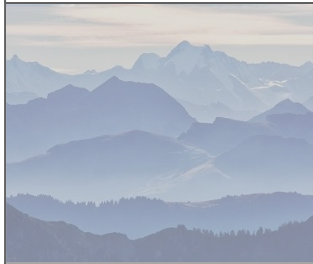


- Building a model
- Fitting / optimization problem
- Model selection
- Measurements and uncertainties



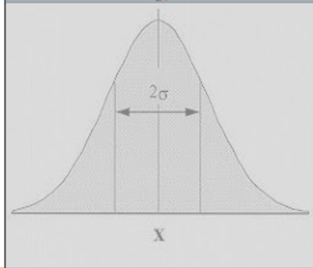
Building a model

Fitting / optimization problem



Model selection

Measurements and uncertainties



Explain **all** of the data

- If you are fitting one/two/three lines, you are throwing out a lot of information
  - If you infer physical conditions from individual lines, likely these conditions also result in the presence or lack of presence of other lines!
  - Using models over the whole band naturally takes into account all information in the data

Explain **all** of the data

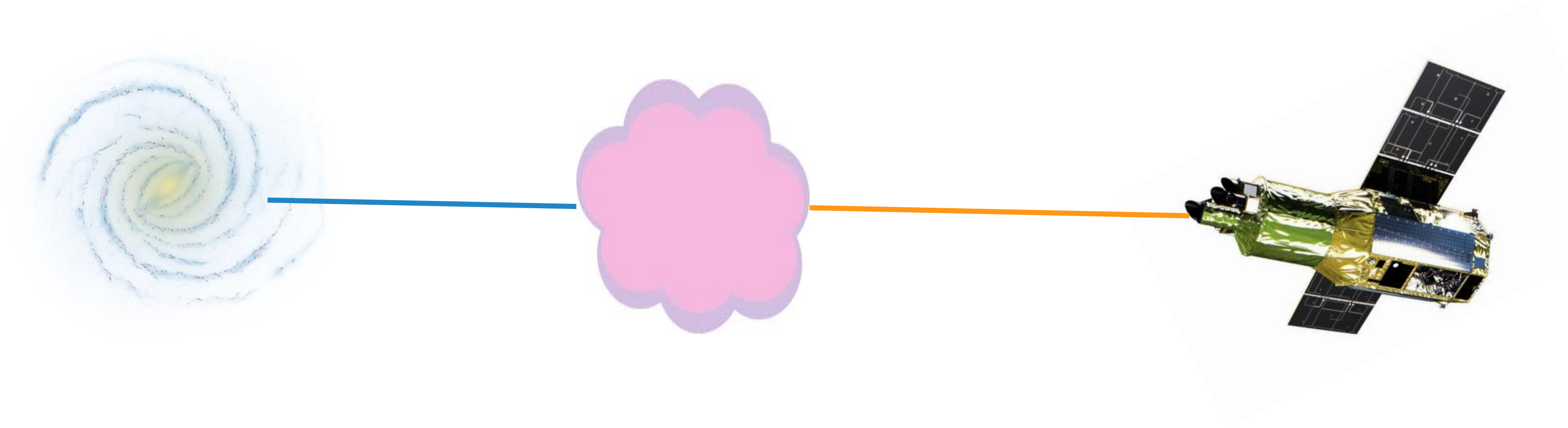
- Be open to new physics!
  - What process ionized the gas?
  - Is the gas in equilibrium?
  - Is the electron distribution Maxwellian?
  - Are lines Gaussian?
  - ...

Explain **all** of the data

- Think about all gas that is in your line of sight - be prepared for unexpected!

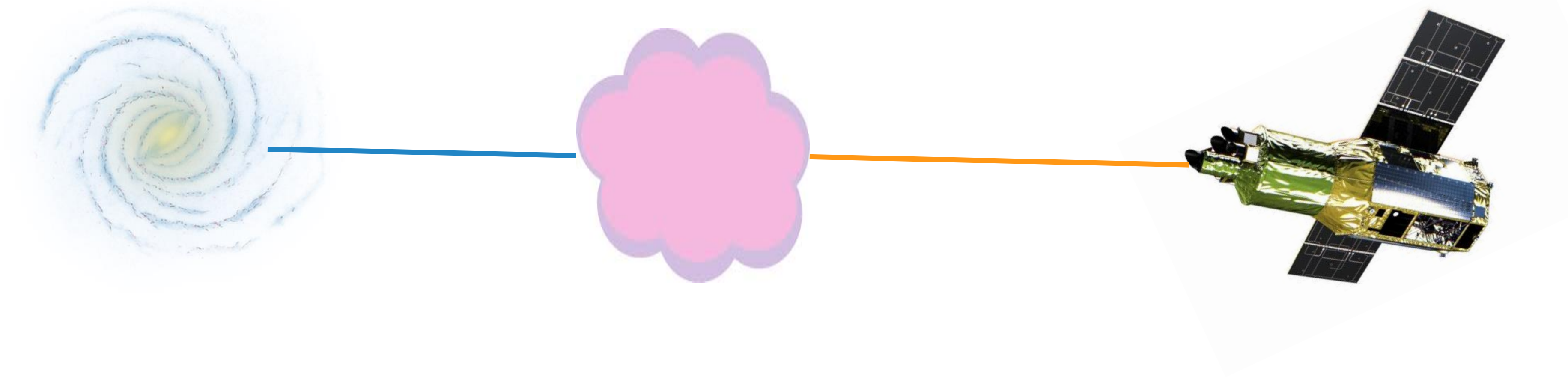
Explain **all** of the data

- Think about all gas that is in your line of sight - be prepared for unexpected!



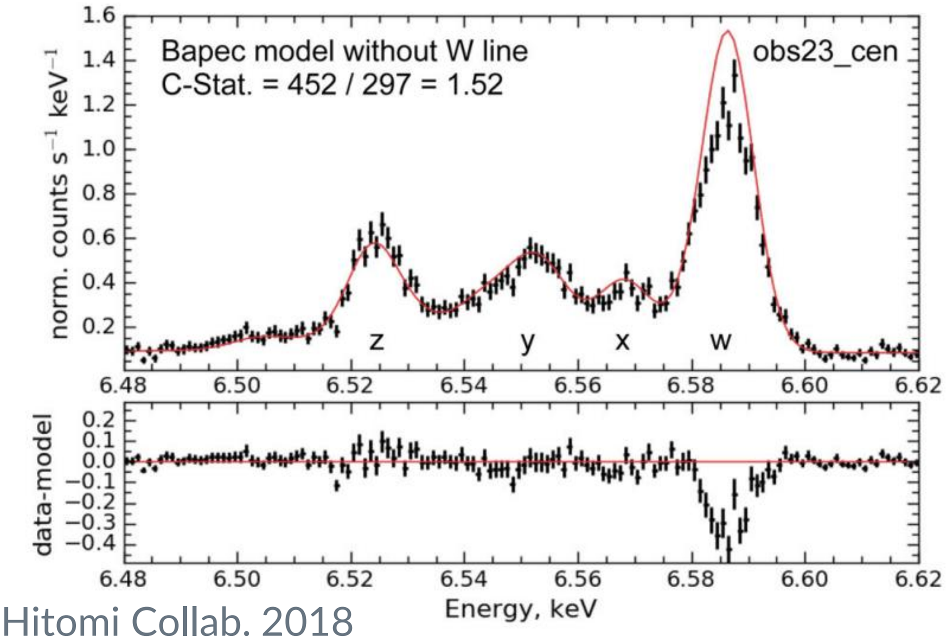
Explain **all** of the data

- Absorption by different phases of the Milky Way ISM
- Dust absorption (excellent new models!)
- Charge exchange
- ...



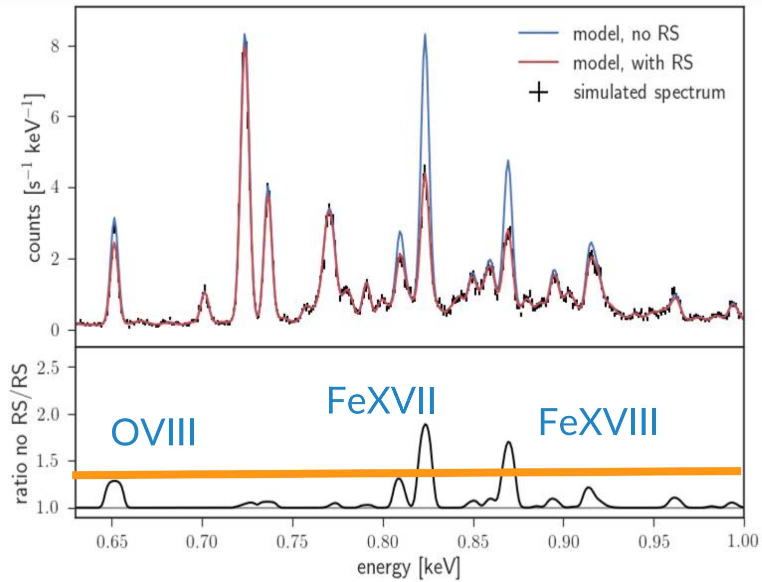
Explain **all** of the data

- Harder to model: radiative transfer

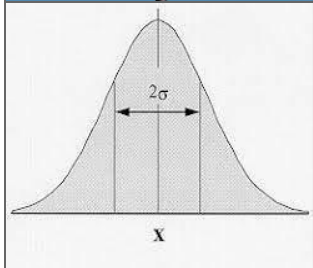


Hitomi Collab. 2018

NGC 4636, XRISM 100 ks



- Building a model
- Fitting / optimization problem
- Model selection
- Measurements and uncertainties

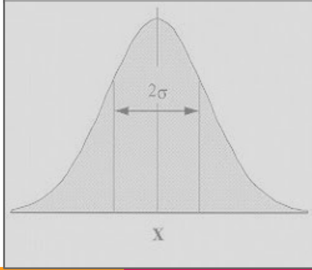


Building a model

Fitting / optimization problem

Model selection

Measurements and uncertainties

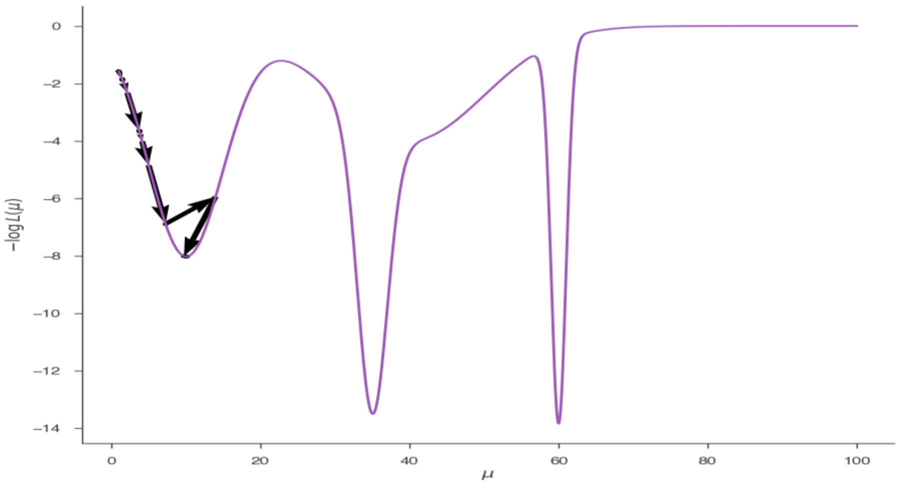


Towards high-dim spaces

- High spectral resolution data gives us more predictive power, so we can test more advanced models
- This means that models will have more parameters: ~10s to ~100s
- How do we find best fit?

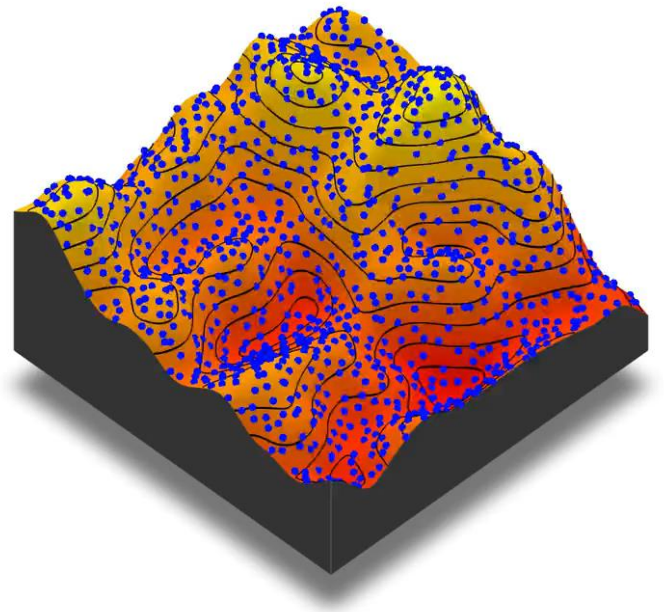
Optimization problem

- Finding the maximum likelihood (or minimum of a fit statistic)



Optimization problem

- If using steepest descent
  - Cover the parameters space well in all dimensions
  - Need to repeat fitting many, many times -> high performance computing



Video by T. Morgan-Wall

Optimization problem

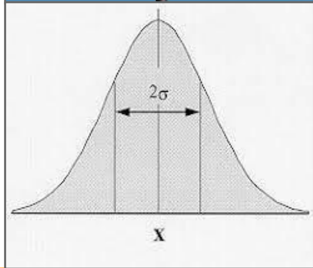
- Can we have better optimization algorithms?
  - *Note: MCMC is NOT an optimization algorithm*

Building a model

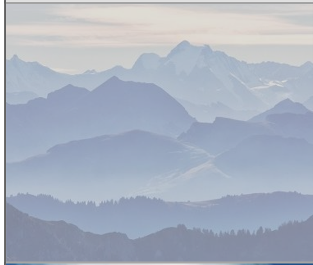
Fitting / optimization problem

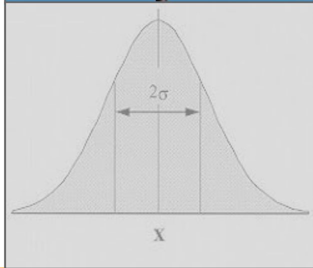
Model selection

Measurements and uncertainties



- Building a model
- Fitting / optimization problem
- **Model selection**
- Measurements and uncertainties





Example 1

Emission model with Solar abundances as a parameter

Example 1

Emission model with Solar abundances as a parameter

Same model with individual elemental abundances free to vary

Example 1

Emission model with Solar abundances as a parameter

Same model with individual elemental abundances free to vary

Adding another emission model

Example 1

Emission model with Solar abundances as a parameter

Which step is more justified given the data?

Same model with individual elemental abundances free to vary

Adding another emission model

Example 2

1 emission component

↓

2 emission components

↓

3 emission components

↓

....

Example 2

At which point am I overfitting the data?

1 emission component

↓

2 emission components

↓

3 emission components

↓

....

How to select models

- We are biased creatures - cannot see this by eye
- Likelihood/C-stat/chi^2 DO NOT answer any of these questions!
    - *These fit statistic should be only compared for the same data and the same model to find the best fitting parameters*
    - *Lower C-stat/higher likelihood does not mean that the model is better - parameter spaces are different!*

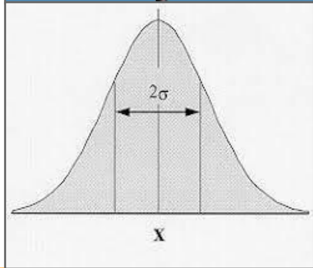
How to select models

- Frequentist approach
  - Likelihood ratio test
- Bayesian approach
  - Bayes factor

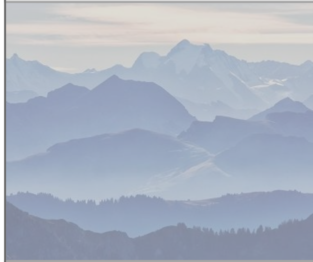
How to select models

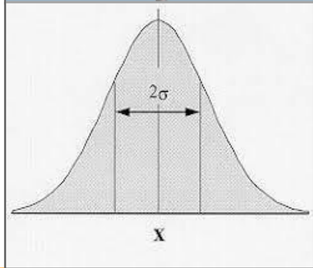
- Frequentist approach
  - - need to apply proper statistic for a given situation, sometimes this is hard
  - + computationally efficient
- Bayesian approach
  - + easier conceptually
  - + does not depend on the model
  - +works for non-nested models
  - - more computationally expensive

- Building a model
- Fitting / optimization problem
- Model selection
- Measurements and uncertainties



- Building a model
- Fitting / optimization problem
- Model selection
- **Measurements and uncertainties**

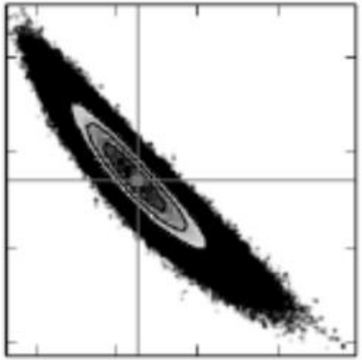



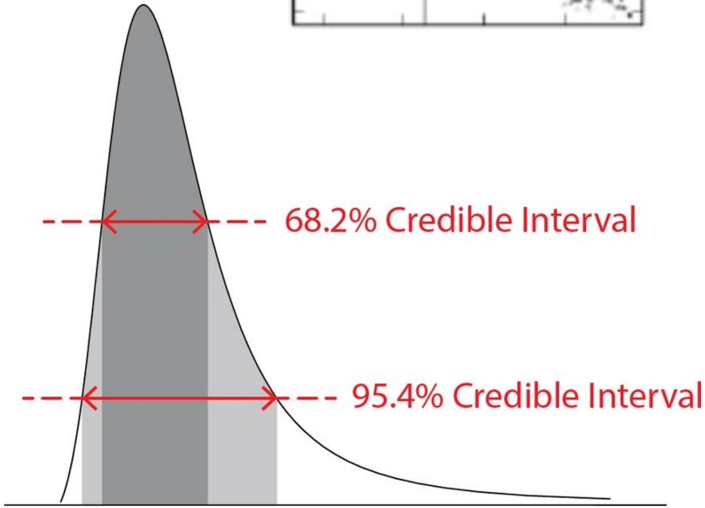
Statistical uncertainties

- Errors may be difficult to compute in highly dimensional parameters spaces
- Monte Carlo Markov Chains (MCMC) are excellent at calculating **posterior distributions**
  - But use MCMC after you already found your best fit!

Statistical uncertainties

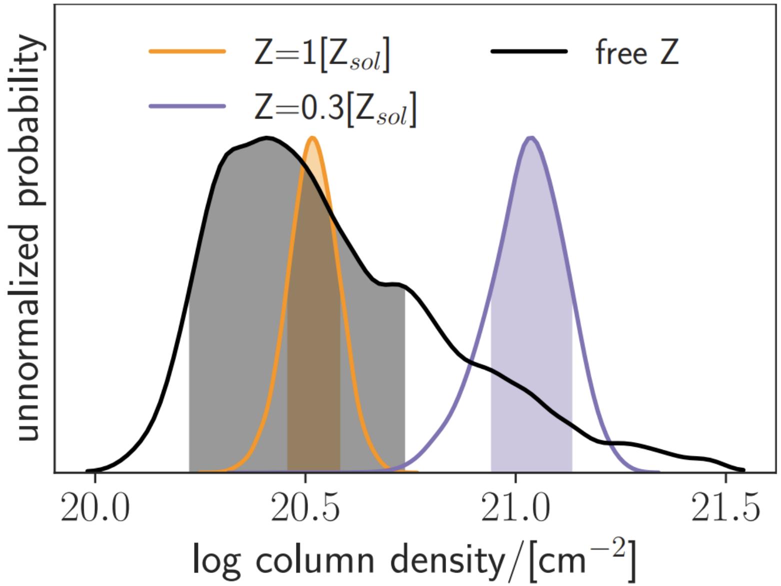
- Advantages of using posterior distributions over simple errors:
  - More meaningful credible intervals - not all measurements are Gaussian!
  - Correct error propagation, especially upper/lower limits
  - Straightforward treatment of correlated parameters - can reduce uncertainties on derived measurements!





Statistical uncertainties

- Fixing parameters biases your measurements and artificially lowers your uncertainties
    - Try to never do this, even if the parameter is unconstrained
    - If you are fixing a parameter, take into account the uncertainty



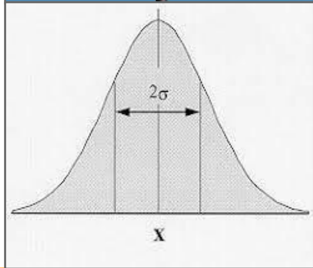
Systematic uncertainties

- Calibration accuracy
- Atomic data underlying models
    - Ideally, we know distributions on these and take them into account in our analysis

Summary

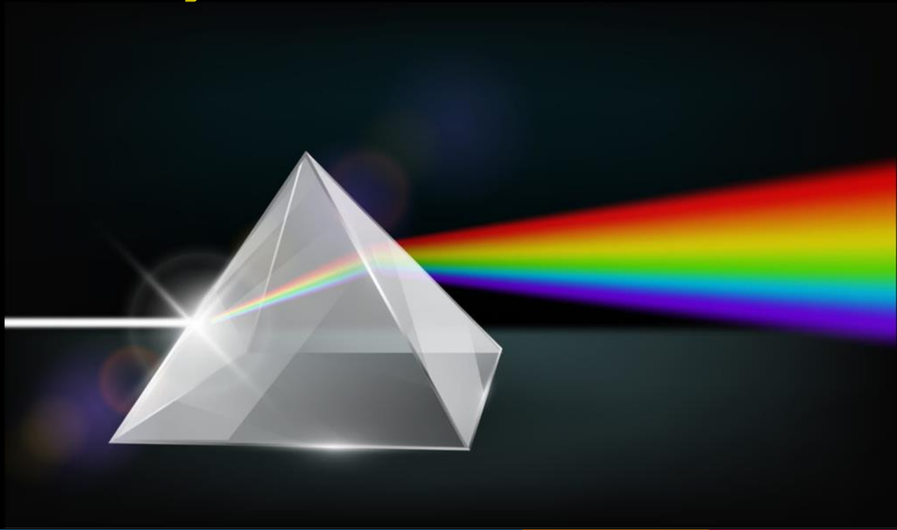
ogoann@umd.edu

- Build good models: take advantage of all data, be open to new physics, think about all line of sight contributors
- Make sure to find the true best fit
- Robustly select models, be agnostic
- Consider MCMC/posteriors for error computation
- Don’t fix parameters
- Acknowledge and estimate systematic effects
- **Make high performance computing part of your workflow (for optimization, model selection, computing errors)**
- **Make sure our software is parallelizable**
- **Be open to using software you’re not used to!**



Questions?
(Find me during coffee breaks!)

What else should we worry about?



ogoann@umd.edu