Embedding Astronomical Computer Models into Complex Statistical Models

David A. van Dyk

Statistics Section, Imperial College London

UCLA, February 2012

Imperial College London

Outline

Outline

Imperial College London

Computer Models

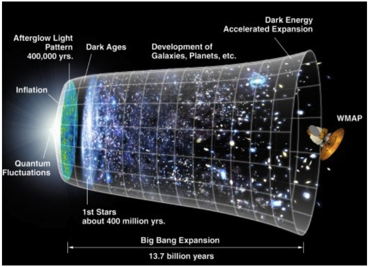
- Complex scientific phenomena can often only be modeled via simultaneous mathematical equations.
- E.g., Pharmacokinetics, Meteorology, Climatology, Seismology, Transportation, Immunology, Astronomy, etc.
- Like a statistical likelihood these *computer models* involve
  - unknown input parameters and
  - prediction or simulation of observations.

  —*Deterministic and Stochastic Computer Models.*
- Require sophisticated and time consuming computation.
- Goal: Use data to learn about parameters and models.

Computer Models in Astronomy

Computer Models are used to

- model stellar evolution,
- describe properties of planetary and stellar atmospheres,
- simulate chemical reactions in interstellar clouds,



- calculate the emergence of clusters and superclusters of galaxies in the early Universe,
- determine the yield of the elements during the Big Bang

Imperial College London

Computer Models in Statistics

Statistical literature focuses on Computer Models in isolation:

- **Emulation:** A statistical model (e.g., a Gaussian Process) is used for interpolation and extrapolation of the computer model.
- **Calibration:** Tuning/Fitting of the input parameters to observed data via a discrepancy measure.
- **Prediction:** Use of calibrated and/or emulated computer model for experiments in place of actual physical process.
- **Uncertainty:** Careful quantification of uncertainty is key: parameter uncertainty and variability, model inadequacy, residual variance, observation error.

Imperial College London

Embedding in a Fully Bayesian Analyses

We aim to use computer models

1. as a component of a statistical likelihood function or
2. to generate from a sampling distribution.

Build multi-level models

- Combine computer models with other model components via multi-level or hierarchical models,
- Combine multiple computer models via parametric models,
- Enable standard techniques for model fitting, checking, comparison, and improvement.

Computation becomes the real issue!

Strategy: *Combine sophisticated models with efficient emulation.*

Imperial College London

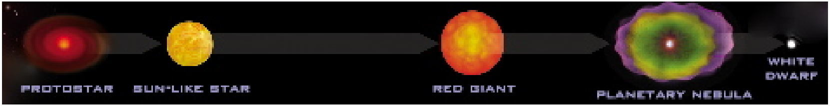
Outline

Imperial College London

Stellar Formation

Stars form when the dense parts of a molecular cloud collapse into a ball of plasma.

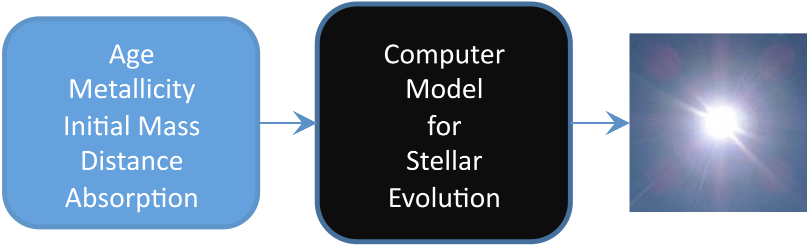
Evolution of a Sun-like Star



- Fusion of Hydrogen into Helium in the core can last millions or billions of years, depending on the initial stellar mass.
- When H is depleted, He may fuse into heavier elements.
- At the same time the star goes through dramatic physical changes, growing and cooling into a *red giant* star.
- The star undergoes mass loss forming a *planetary nebula*.
- Eventually only the core is left, a *white dwarf star*.
- White dwarf is a stellar ember and cools slowly via conduction, convection, and/or radiation.

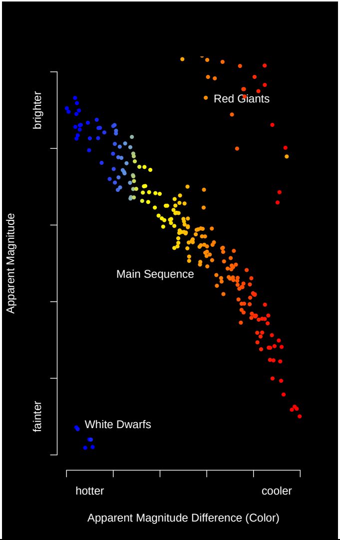
Imperial College London

Compter Model for Sun-Like Stellar Evolution



- Computer model predicts how the emergent and apparent spectra evolve as a function of input parameters.
- We observe photometric magnitudes, the apparent luminosity in each of several wide wavelength bands.

The Data: Color Magnitude Diagrams



Color-Magnitude Diagram

- Plot Magnitude Difference vs. Magnitude.
- Identifies stars at different stages of their lives.
- Evolution of a CMD.
- Facilitates physical intuition as to likely values of parameters.
- “Chi-by-eye” fitting.

Imperial College London

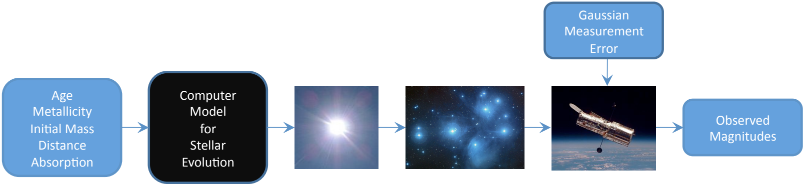
Computer Models for MS/RG Evolution

Computer Models Predict Magnitudes From Stellar Parameters

- Must iteratively solve set of coupled differential equations.
- This creates a static physical model of a star, which is how a star of a particular mass and *radial abundance profile* would appear in terms of its luminosity and color.
- Stars are evolved by updating the mass and abundance profile to account for the newly produced elements.
- Finally interstellar absorption and distance can be used to convert absolute magnitudes into apparent magnitudes.

Imperial College London

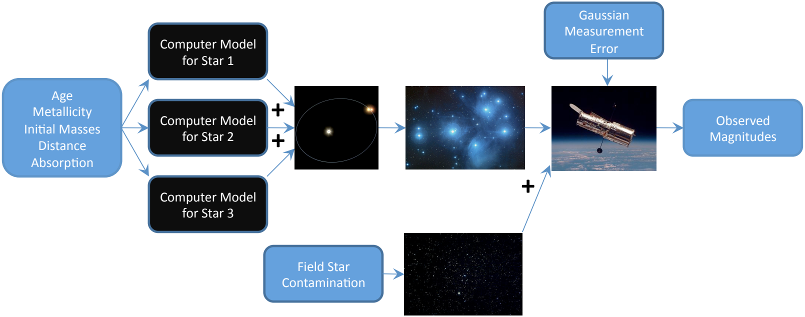
Embedding Computer Model intp Statistical Model



- Typically more parameters than measurements per star.
- We study stellar clusters with (nearly) common age, metallicity, distance, and absorption.
- Magnitudes observed with Gaussian measurement errors.

Imperial College London

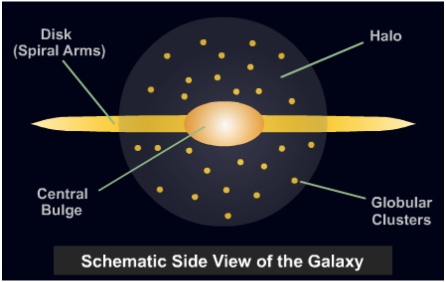
Multi-Star Systems and Field Star Contamination



- Between 1/3 and 1/2 of “stars” are unresolved binaries.
  - *Sum luminosities from multiple computer model runs.*
- Cluster data is contaminated with field stars.
  - *Finite mixture model.*

Study of WDs: Age of Galactic Structures

- Age of galactic *halo* or *disk* can only be estimated with older stars.
- Stellar clusters are pulled apart as they interact gravitationally with other stars and clusters.

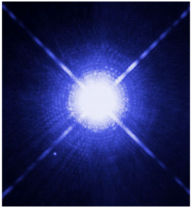


- Older stars tend to be “in the field” — not in clusters.
- The colors of a single white dwarf are much more informative as to its age than are the colors of a MS star.

We would like to model white dwarf colors.

Imperial College
London

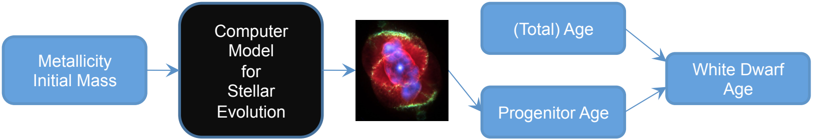
White Dwarfs Physics



- White dwarf spectra are not predicted from MS/RG models
- Different physical processes require different models:
  1. Computer Model for White Dwarf Cooling
  2. Computer Model for White Dwarf Atmosphere
  3. Initial Final Mass Relationship (IFMR)

Imperial College London

Computing the Progenitor Age

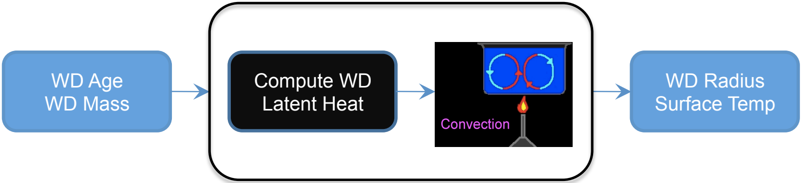


Begin with the MS / RG model:

- Rather than running the MS/RG model for a fixed *age*, we run it until the giant evolves into a white dwarf.
- This gives us the progenitor age of the MS / RG star.
- Subtract from total cluster age to get *White Dwarf Age*.

The White Dwarf Cooling Model

Computer Model for White Dwarf Cooling

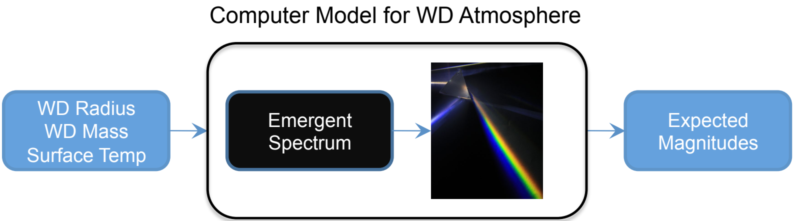


A White Dwarf is a Cooling Ember

- Heat passes to the surface via some combination of *conduction, convection, and/or radiation*.
- Depends on the local temperature.
- Numerically modeling these processes yields the *surface temperature and radius*.

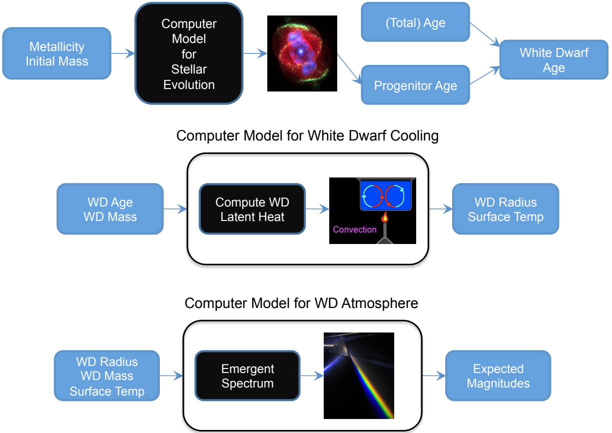
Imperial College London

The White Dwarf Atmosphere Model



- Predicts the distribution of the wavelength of emitted electromagnetic radiation.
- We account for the filters used in photometric magnitudes.
- We account for absorption and distance.

The Missing Link: White Dwarf Mass



We must model the white dwarf mass.

Imperial College London

A Simple Model for the WD Mass

The Initial Final Mass Relationship (IFMR):

- Predict the White Dwarf mass a function of the Initial Mass.
- With narrow range of mass, relation is approximately linear:

$$\text{White Dwarf Mass} = \alpha + \beta \text{ Initial Mass}$$

  - More massive stars evolve into white dwarfs sooner.
  - Progenitors of visible cluster white dwarfs had similar mass.
- Goals:
  - Account for IFMR uncertainty in a coherent model.
  - Fit IFMR over a wide range of masses using several clusters, each with (different) linear models.

Parametric Bridge between Computer Models.

Traditional Method for Fitting IFMR

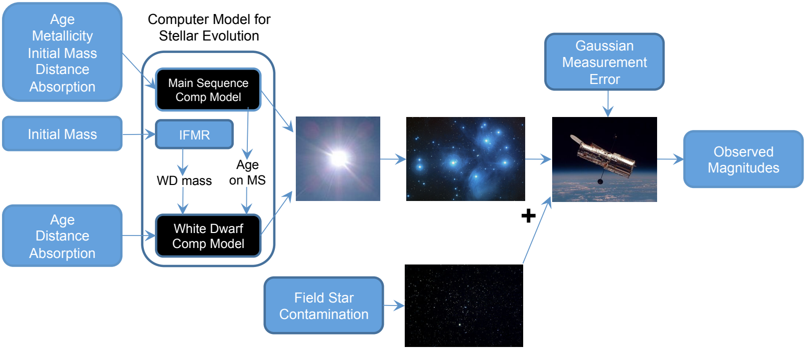
A Sequence of Independent Analyses:

- *Total (cluster) age:* Fit the color-magnitude diagram to the MS / RG model using chi-by-eye.
- *White Dwarf Mass and Age:* Spectroscopy to determine temperature and surface gravity. Age and mass are backed out using computer model for white dwarf cooling.
- *Progenitor Age:* Subtract white dwarf age from total age.
- *Initial Mass:* Progenitor age with fitted MS / RG model.
- ***Errors in fitted line are difficult to evaluate!!***

We aim to fit the IFMR with a coherent statistical model without expensive star-by-star spectrography.

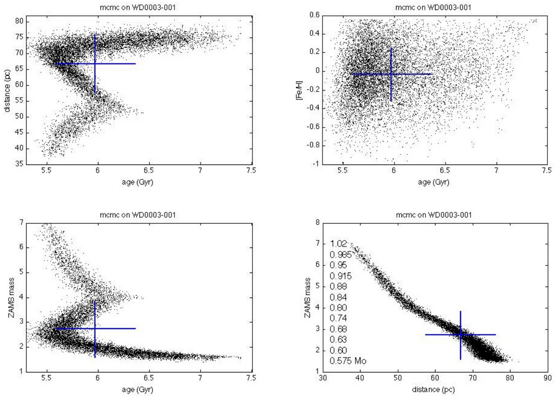
Imperial College London

Opening Up the Black Box: The Final Model



Imperial College London

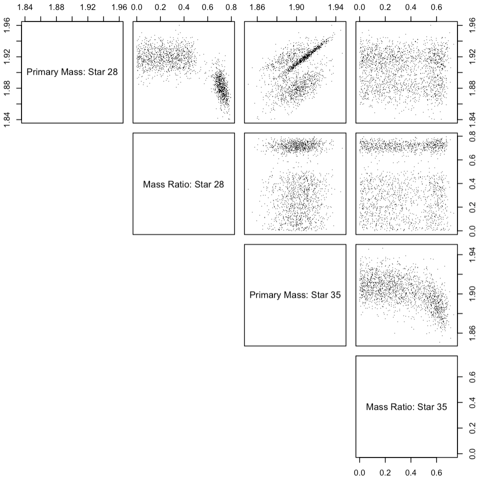
Model Fitting: Complex Posterior Distributions



Highly non-linear relationship among stellar parameters.

Imperial College London

Model Fitting: Complex Posterior Distributions



The classification of certain stars as field or cluster stars can cause multiple modes in the distributions of other parameters.

Imperial College London

Statistical Computation

- Hundreds of parameters

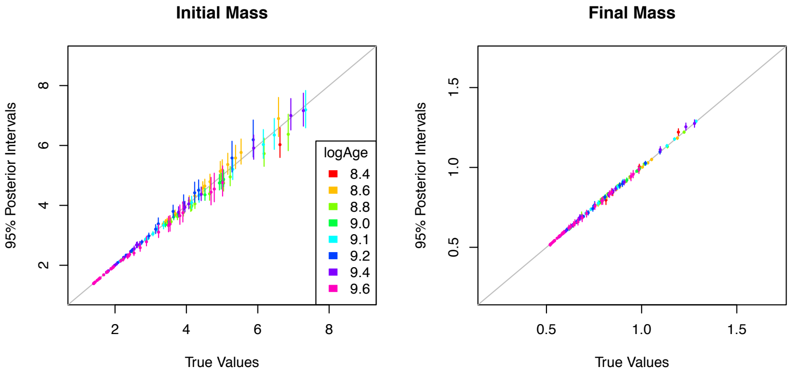
Stellar:	Mass, Mass Ratio, Cluster Membership
Cluster:	Age, Metallicity, Distance, Absorption
General:	IFMR slope, IFMR intercept

- Strategy: numerically integrate out stellar parameters and use Metropolis on remaining six parameters.
- Marginal posterior factors into N_{stars} 2D integrals.
- Computer code for MCMC is easy to parallelize.

Result: Fast Mixing but computationally expensive code.

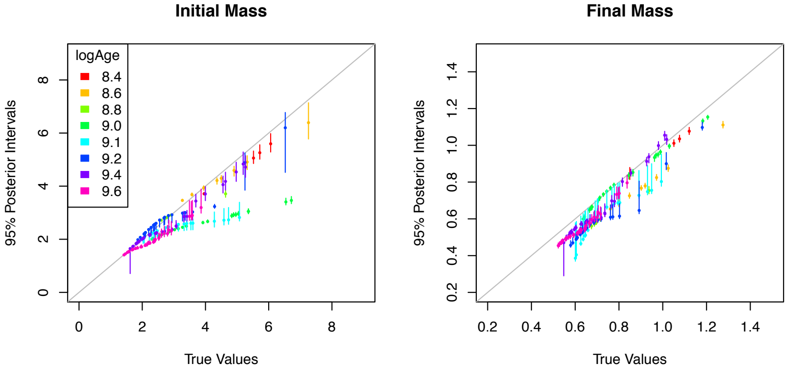
Imperial College London

Simulation: Recovering the Masses



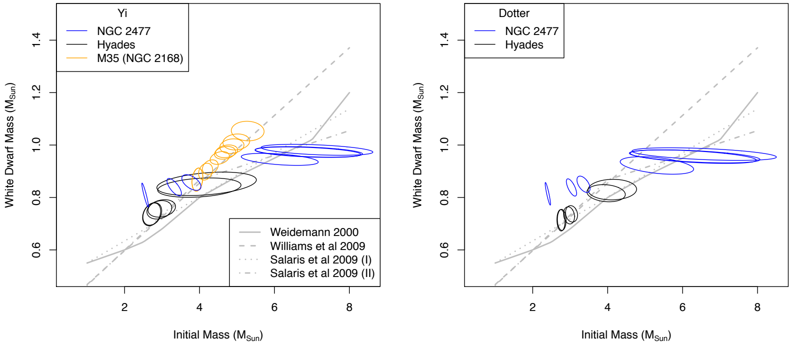
- Simulated 8 clusters of varying age–and white dwarf mass.
- Resulting fits recovery of the masses well.

Simulation: Sensitivity to MS / RG Model Choice



- Checking Computer Models: Use different computer models in simulation (YY) and fit (Dotter et al.).
- Measure of bias relative to "True Model"?

FItting the IFMR



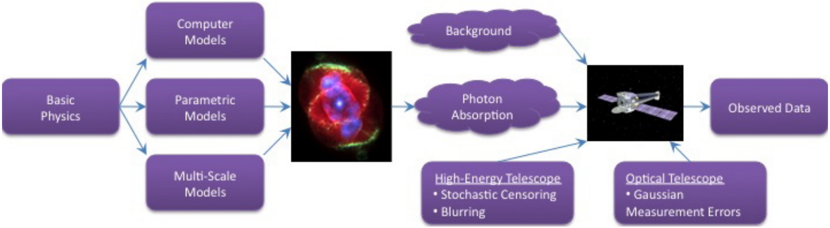
- How best to combine results from three clusters?
- Is there one relationship? Depend on other variables?

Imperial College London

Outline

Imperial College London

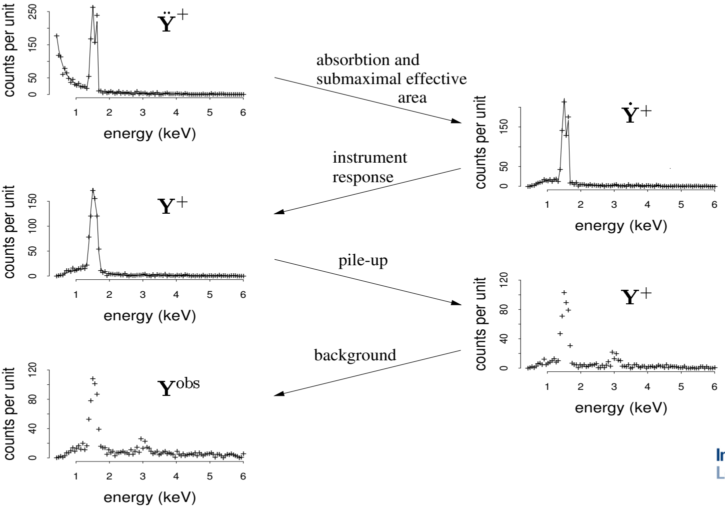
The Basic Statistical Model



- Embed physics models into multi-level statistical models.
- X-ray and γ-ray detectors count a typically *small number of photons* in each of a *large number of pixels*.
- Must account for complexities of data generation.
- State of the art data and computational techniques enable us to fit the resulting complex model.

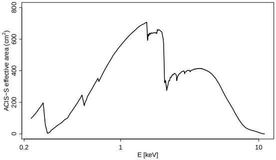
Imperial College London

Degradation of the Photon Counts

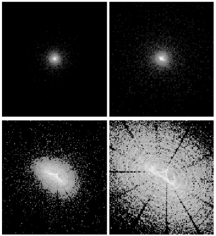


Calibration Products

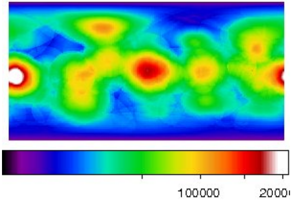
- Analysis is highly dependent on *Calibration Products:*
  - Effective area records sensitivity as a function of energy
  - Energy redistribution matrix can vary with energy/location
  - Point Spread Functions can vary with energy and location
  - Exposure Map shows how effective area varies in an image
- In this talk we focus on uncertainty in the effective area.



A *CHANDRA* effective area.



Sample Chandra psf's
(Karovska et al., ADASS X)

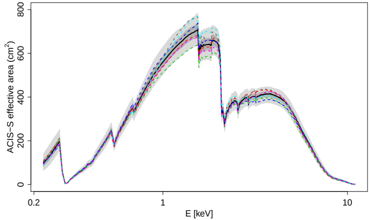


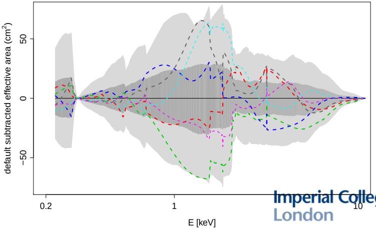
EGERT exposure map
(area $\times$ time)

Imperial College London

Derivation of Calibration Products

- Prelaunch ground-based and post-launch space-based empirical assessments.
- Aim to capture deterioration of detectors over time.
- Complex computer models of subassembly components.
- Calibration scientists provide a sample representing uncertainty
- *Calibration Sample* is typically of size $M \approx 1000$.



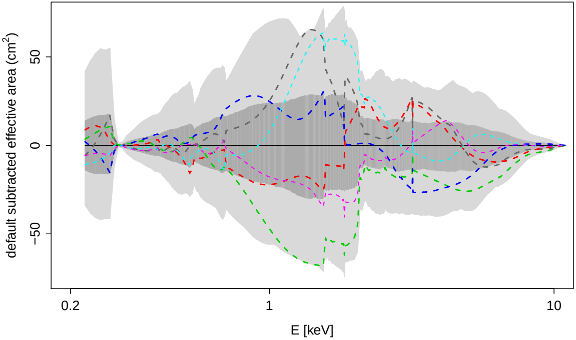


Imperial College London

Complex Variability

Computer model generated calibration sample requires:

1. *running the computer model on the fly, or*
2. *storing many high dimensional calibration products.*



Imperial College London

Simple Emulation of Computer Model

We use *Principal Component Analysis* to represent uncertainly:

$$A \sim A_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j,$$

A_0: default effective area,

$\bar{\delta}$: mean deviation from A_0,

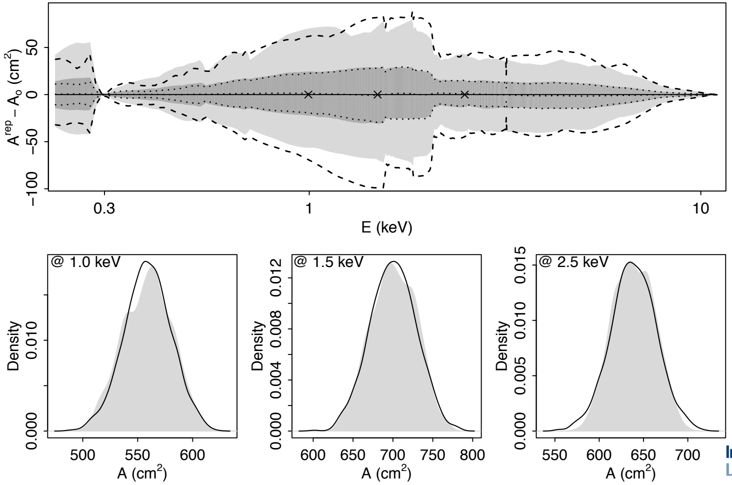
r_j and $\boldsymbol{v}_j$: first m principle component eigenvalues & vectors,

e_j: independent standard normal deviations.

Capture 95% of variability with $m = 6 - 9$.

Imperial College London

Checking the PCA Emulator



Imperial College London

Using Monte Carlo to Account for Uncertainty

1. Drake et al. (2006) propose a bootstrap-like method:
   - Simulate M spectra under fit model & default effective area.
   - Fit each spectra with effective area from calibration sample.
2. A simpler solution involves Multiple Imputation:
   - Treat $m \ll M$ effective areas from calibration sample as imputations and fit the model m times.
   - Use MI *Combining Rules* to compute estimates and errors.
3. When using MCMC in a Bayesian setting we can:
   - Sample a different effective area from calibration sample at each iteration according to its conditional distribution.
   - Effectively average over the calibration uncertainty.

Imperial College London

Two Possible Target Distributions

We consider inference under:

A PRAGMATIC BAYESIAN TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.

THE FULLY BAYESIAN POSTERIOR: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

Concerns:

- **Statistical** Fully Bayesian target is "correct".
- **Cultural** Astronomers have concerns about letting the current data influence calibration products.
- **Computational** Both targets pose challenges, but pragmatic Bayesian target is easier to sample.
- **Practical** How different are $p(A)$ and $p(A|Y)$?

Drake (2006) and MI are approximations to π_0.

Imperial College London

Sampling the Full Posterior Distribution

- Sampling $\pi(A, \theta) = p(A, \theta | Y)$ is complicated because we only have a computer-model generated sample of $p(A)$ rather than an analytic form.
- But PCA gives a *degenerate normal* approximation:

$$A \sim A_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j,$$

  where e_j are independent standard normals.
- PCA represents A as deterministic function of $e = (e_1, \ldots, e_m)$.
- We can construct an MCMC sampler of $p(e, \theta | Y)$.

A Prototype Fully Bayesian Sampler

An MH within Gibbs Sampler:

STEP 1: $e \sim \mathcal{K}(e|e', \theta')$ via MH with limiting dist'n $p(e|\theta, Y)$

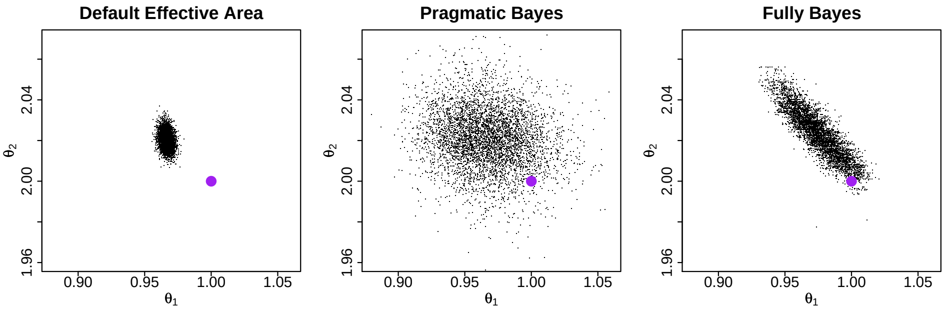
STEP 2: $\theta \sim \mathcal{K}(\theta|e', \theta')$ via MH with limiting dist'n $p(\theta|e, Y)$

- STEP 1: Gaussian Metropolis jumping rule centered at e'.
- STEP 2: Simplified pyBLoCXS (no rmf or background).

A Simulation.

- Sampled 10^5 counts from a power law spectrum: E^{-2}.
- A_{true} is 1.5σ from the center of the calibration sample.

Sampling From the Full Posterior



Spectral Model (purple bullet = truth):

$$f(E_j) = \theta_1 e^{-\theta_3 x(E_j)} E_j^{-\theta_2}$$

Pragmatic Bayes is clearly better than current practice, but a Fully Bayesian Method is the ultimate goal.

Implementing the Fully Bayesian Analysis

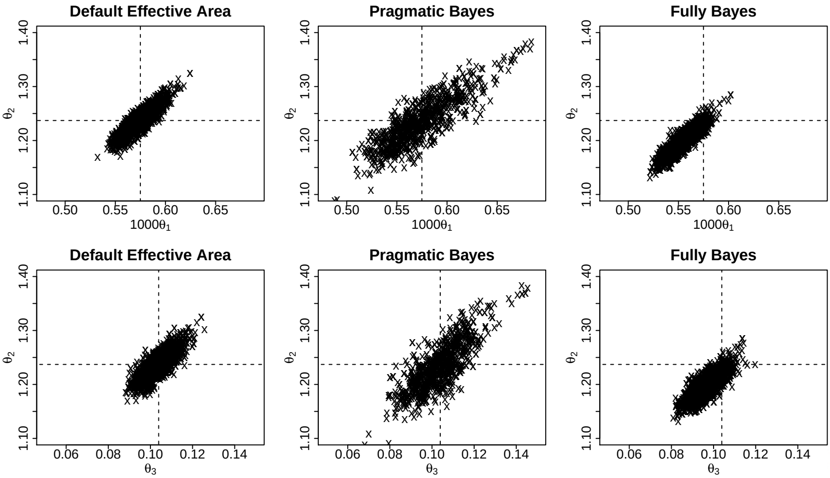
An MH within Gibbs Sampler:

STEP 1: $e \sim \mathcal{K}(e|e', \theta')$ via MH with limiting dist'n $p(e|\theta, Y)$

STEP 2: $\theta \sim \mathcal{K}(\theta|e', \theta')$ via MH with limiting dist'n $p(\theta|e, Y)$

- We use a mixture of two jumping rules in each step:
    - STEP 1: Gaussian Metropolis jump centered at e' and MH jump from the prior.
    - STEP 2: T Metropolis jump centered at θ' and MH jump from an approximation to the posterior.
- Six tuning parameters: three scale parameters, two proportions (M vs MH jump), and *df* for T.
- Tuning parameters must be adjusted.

The Effect in an Analysis of a Quasar Spectrum



Imperial College London

Thanks...

Stellar Evolution:

- Nathan Stein
- Steven DeGennaro
- Elizabeth Jeffery
- William H. Jefferys
- Ted von Hippel

Instrument Calibration

- Vinay Kashyap
- Jin Xu
- Alanna Connors
- Hyunsook Lee
- Aneta Siegminowska
- California-Harvard Astro-Statistics Collaboration

Imperial College London

For Further Reading I

- Stein, N, van Dyk, D., von Hippel, T., DeGennaro, S., Jeffery, E., Jeffreys, W. H.
  Combining Computer Models in a Principled Bayesian Analysis: From Normal Stars to White Dwarf Cinders. Submitted.
- Jeffery, E., von Hippel, T., DeGennaro, S., van Dyk, D., Stein, N., and Jefferys, W.
  The White Dwarf Age of NGC 2477. *Astrophysical Journal*, **730**, 35–43, 2011.
- Lee, H., Kashyap, V., van Dyk, D., Connors, A., Drake, J., Izem, R., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A., and Zezas, A.
  Accounting for Calibration Uncertainties in X-ray Analysis: Effective Area in Spectral Fitting. *The Astrophysical Journal*, **731**, 126–144, 2011.
- van Dyk, D. A., DeGennaro, S., Stein, N., Jefferys, W. H., von Hippel, T.
  Statistical Analysis of Stellar Evolution *The Annals of Applied Statistics* **3**, 117-143, 2009.
- DeGennaro, S., von Hippel, T., Jefferys, W., Stein, N., van Dyk, D., and Jeffery, E.
  Inverting Color-Magnitude Diagrams to Access Precise Cluster Parameters: A New White Dwarf Age for the Hyades. *Astrophysical Journal*, **696**, 12–23, 2009.

The Simulation Studies

Simulated Spectra

- Spectra were sampled using an absorbed power law,

$$f(E_j) = \alpha e^{-N_H x(E_j)} E_j^{-\Gamma},$$

  accounting for instrumental effects; E_j is the energy of bin j.
- Parameters (Γ and N_H) and sample size/exposure times:

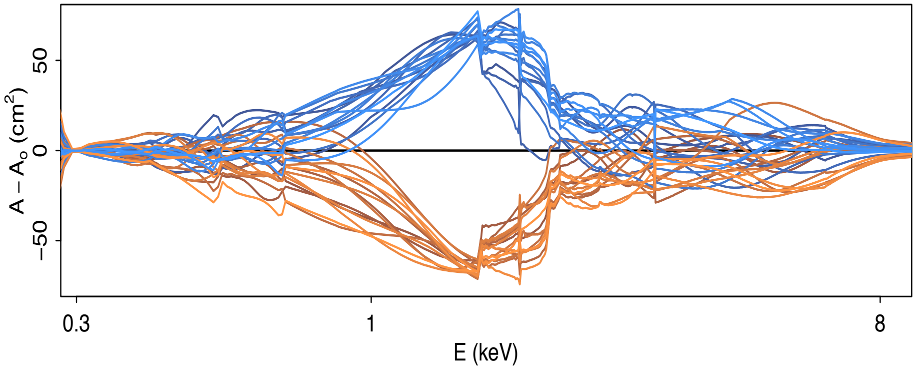
	Effective Area		Nominal Counts		Spectal Model	
	Default	Extreme	10^5	10^4	Hard†	Soft‡
SIM 1	X		X		X	
SIM 2	X		X			X
SIM 3	X			X	X	

†An absorbed powerlaw with $\Gamma = 2$, $N_{\mathrm{H}} = 10^{23}/\mathrm{cm}^2$

‡An absorbed powerlaw with $\Gamma = 1$, $N_{\mathrm{H}} = 10^{21}/\mathrm{cm}^2$

Imperial College London

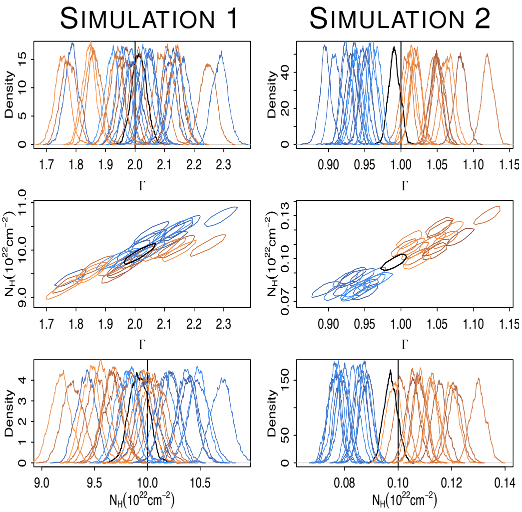
30 Most Extreme Effective Areas in Calibration Sample



15 largest and 15 smallest determined by maximum value

Imperial College London

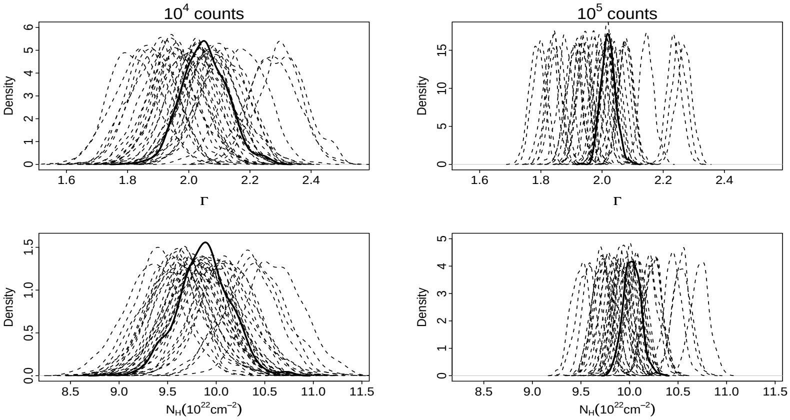
The Effect of Calibration Uncertainty



- Columns represent two simulated spectra.
- True parameters are horizontal lines.
- Posterior under default calibration is plotted in black.
- The posterior is highly sensitive to the choice of effective area!

Imperial College London

The Effect of Sample Size



The effect of Calibration Uncertainty is more pronounced with larger sample sizes.

Imperial College London