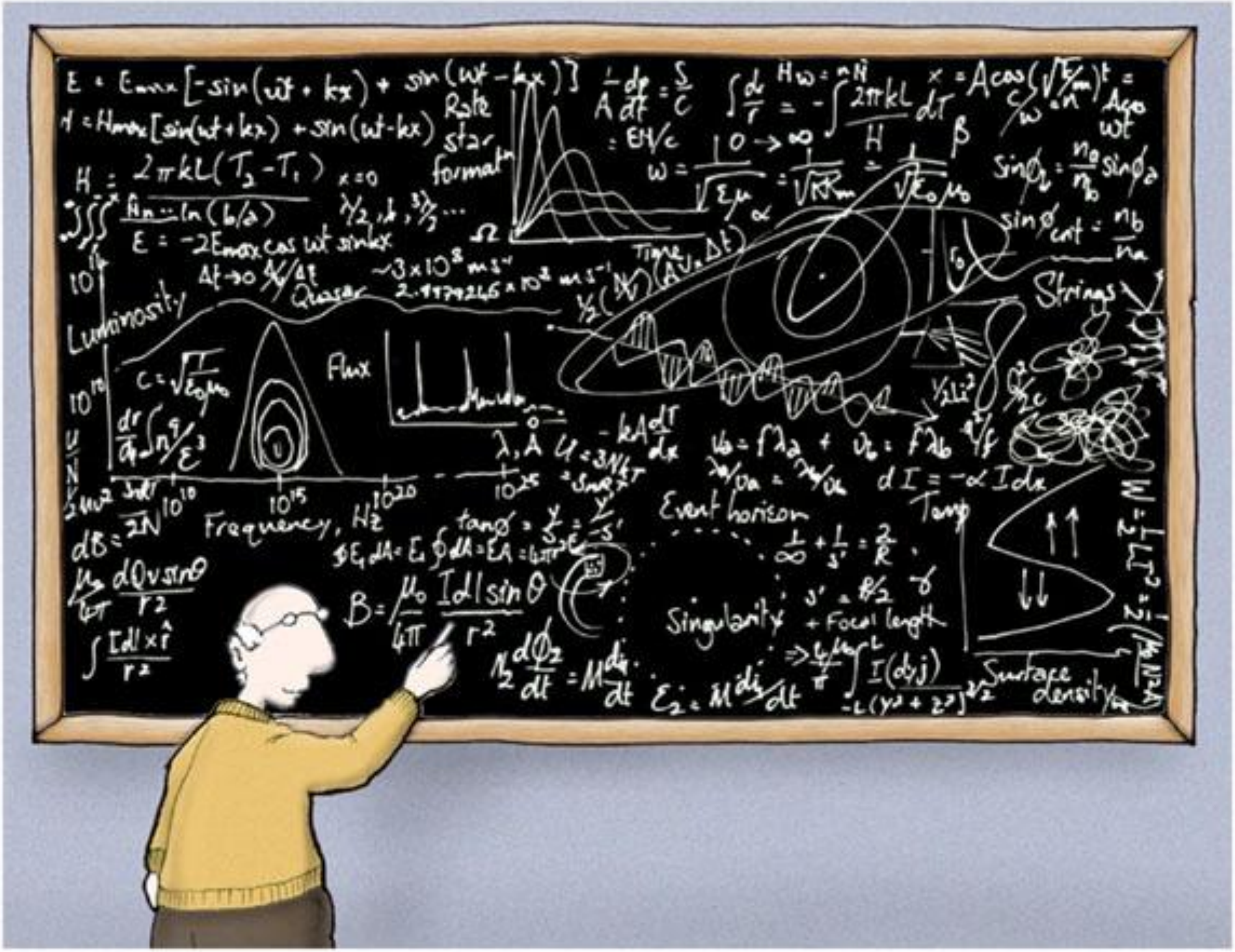
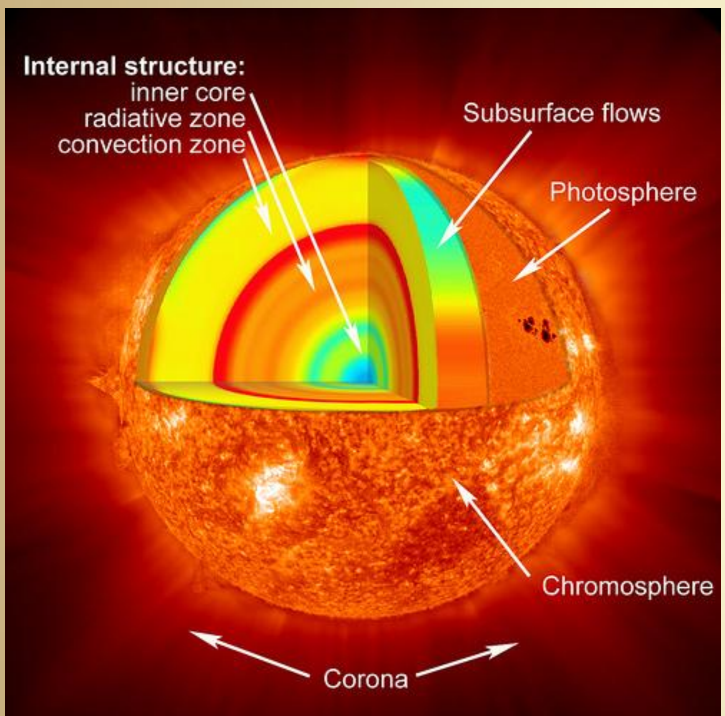
Differential Emission Measure analysis with high-resolution X-ray Spectra

Victoria Liublinska
Harvard Statistics Department



Astrophysics made simple

Have you ever thought why stars shine so brightly?

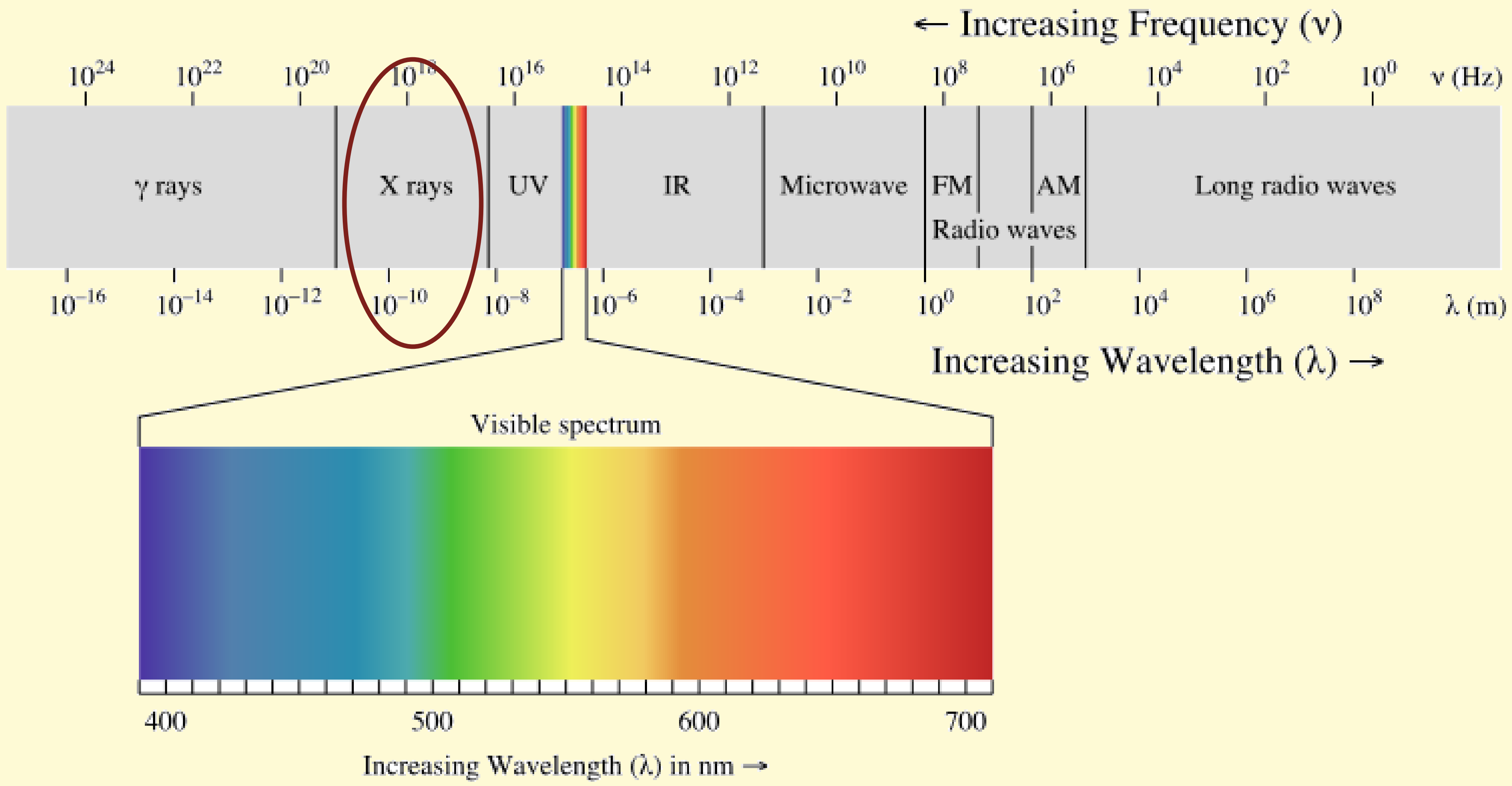


- A **star** is a massive, luminous ball of plasma that is held together by gravity

- It has certain layers, in particular, a hot solid core in the middle and a very hot gas of low density (corona) at the outermost layer.

- The energy produced by stars, as a by-product of free-flying radiation interacting with atoms, escapes into space as both electromagnetic radiation and particle radiation

Electromagnetic radiation



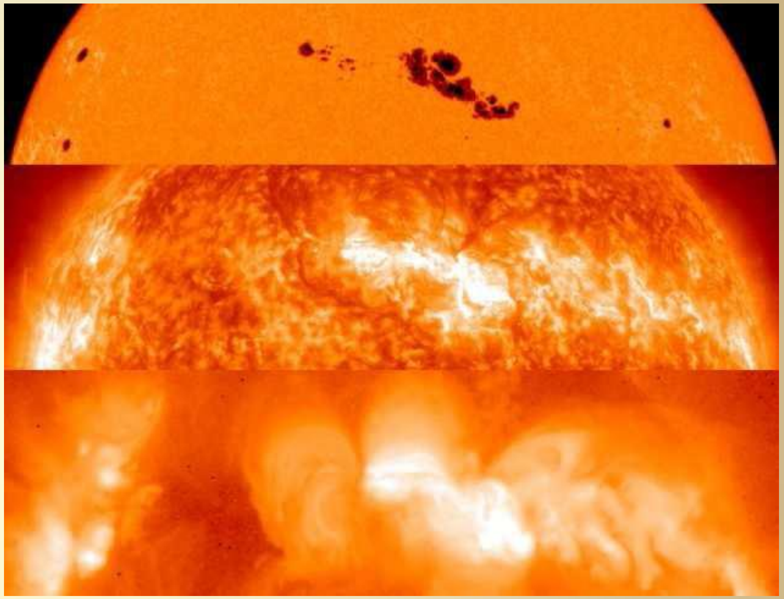
More then one "face" of the star

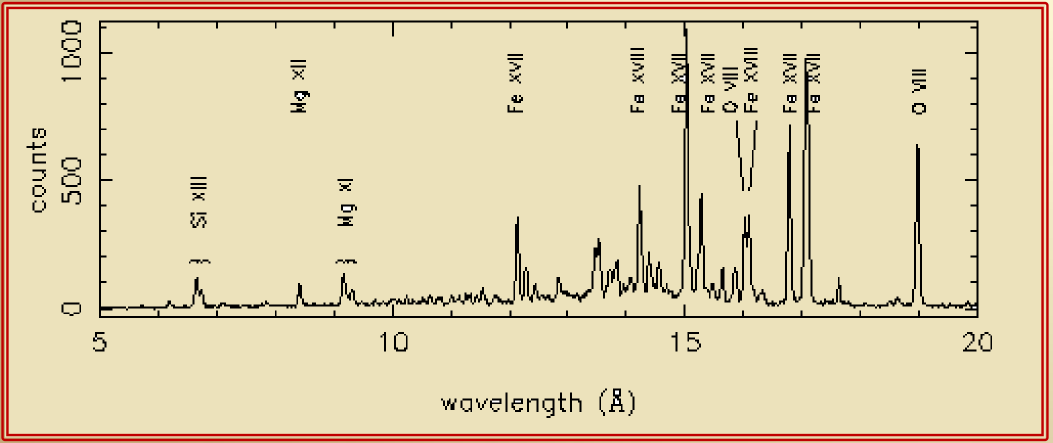
- As we can see, the star can be viewed through different filters that will reveal the spatial distribution of intensity of different types of radiation.

Optical

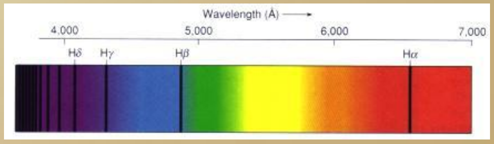
Extreme UV

X-ray image





For a better spectral resolution we turn to stellar emission spectrum (i.e. the photon counts corresponding to different energies)



Part of it can be represented as a "rainbow" with dark lines, or absorption spectrum

What's up with those lines?

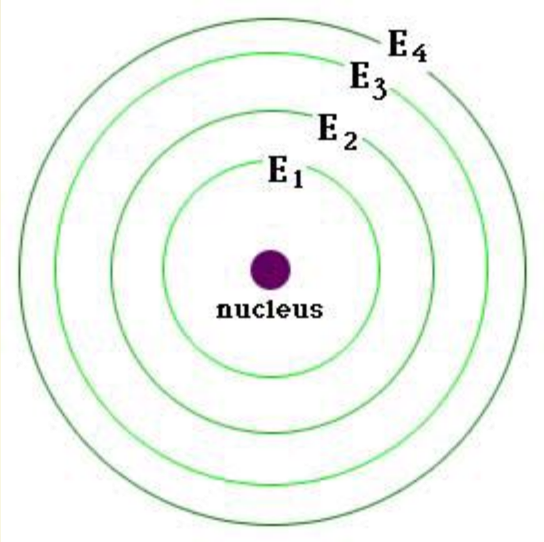
- Every atom or ion of a chemical element is characterized by a certain number of protons and neutrons (that make up a nucleus) and electrons that can occupy several energy levels around the nucleus.

- Electrons can be moved from one energy level to another by collisions among atoms or *by radiating photons.*

- Only photons with particular energies, those that correspond to differences between the various energy levels (E_{i+1}-E_i), can be emitted and their wavelength is simply calculated as

$$\lambda_i \approx \frac{hc}{E_{i+1} - E_i}$$

- Therefore, each atom or ion has a unique set of emission lines and we observe the convolution of all of them in a stellar spectrum.



Actually, a spectral line extends over a range of frequencies due to Heisenberg **uncertainty principle**

More complications…

- Atomic physicists continue to discover and calculate energy levels for atoms of different chemical elements and the corresponding wavelength of photons. For example, *H* has only 4 lines in the visual spectrum , but *Fe* has thousands but those calculations are prone to errors… (and current analysis can help correct those errors!)

- Higher temperature of gas (corona) broadens lines (Doppler effect) and change their intensities, so that instead of an array of lines, we get a contribution function $G_l^{L,k}(t)$ of line-to-plasma temperature correspondence.

- Besides , a so-called Bremsstrahlung continuum ,that refers to *any radiation due to the acceleration of a charged particle,* contribute to the total count of photons observed at any wavelength.

Model for High-Energy Spectra

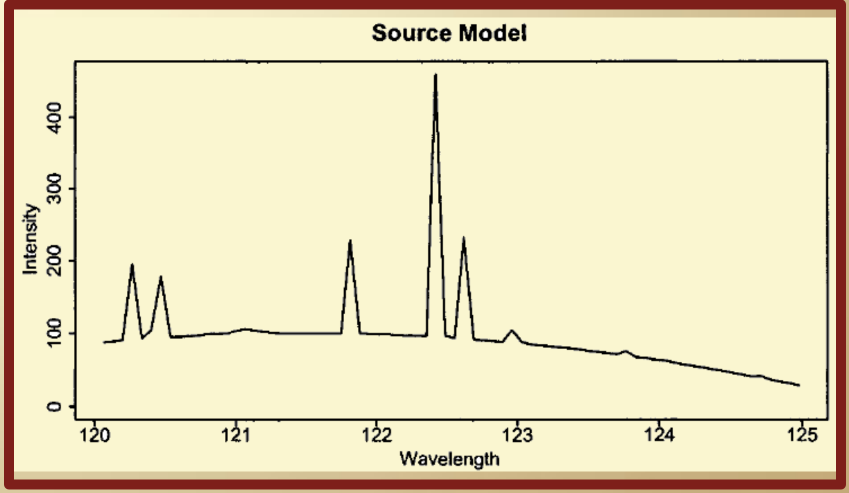
To summarize, the components of high-energy spectrum can be split into two groups:

- Continuum terms, which are quite smooth and describe emission distribution over the entire range of wavelength.
The information about them is provided by atomic physicist in the form of another contribution function $G_j^{C,k}(t)$.

- Emission lines, which are local positive aberrations from the continuum, represented, as we noted, by $G_l^{L,k}(t)$.

So the *total emissivity function* can be calculated as

$G_j(t) = \Sigma \gamma_k (G_j^{C,k}(t) + binned\{G_l^{L,k}(t)\})$,
where γ_k is the abundance of element *k*.



What are we studying?

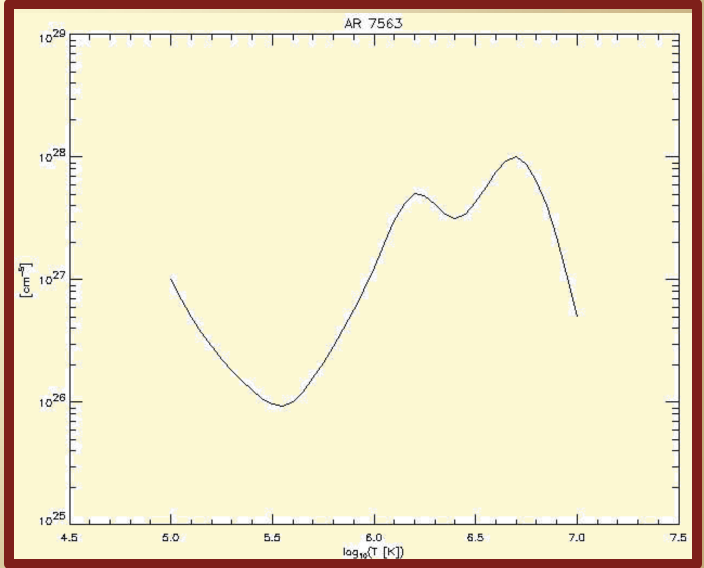
As we can see, using the stellar spectrum and a fair amount of modeling, we can potentially determine both:

- Coronal temperature
- metallicity (elemental abundance or composition) of a star.

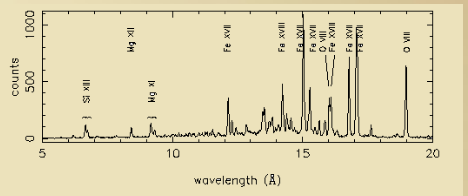
A way to characterize the temperature structure of stellar corona is Differential Emission Measure (DEM).

In simple terms, DEM is a *distribution of the amount of emission at different temperatures in a stellar corona.*

For example, this picture features the Solar DEM in an Active Region (many sunspots). It has log10(T) on x-axis and a measure of emission on y-axis (relative abundance of matter). Notice, that Sun's active regions have a lot of very hot plasma.



Building the spectral model



- Emission line spectral model:
  - Average photon intensity corresponding to spectral line l generated by element k is

$$\lambda_l^{L,k} = A\int \gamma_k G_l^{L,k}(T)DEM(T)d\log(T) \propto (\Delta\log(T))\sum_{t=1}^{T}\gamma_k G_{l,t}^{L,k}\mu_t$$

Note, that this model will only compute *relative values*, so the absolute magnitude of the values are *meaningless* (in particular, this fact lets us ignore the exposure time!).

- Continuum spectral model:
  - In the similar manner we get: $\lambda_j^{C,k} \propto (\Delta\log(T))\sum_{t=1}^{T}\gamma_k G_{j,t}^{C,k}\mu_t$

Now j corresponds to a certain energy bin (or range of wavelength) out of J bins in total.

Mixture of emission lines and continuum

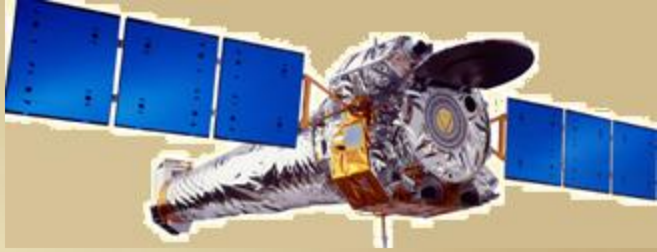
- Spectrum is divided into J (equispaced) bins (the range depends on the actual dataset)
- Temperature is divided into T (equispaced) intervals ranging from 10^4 to 10^8 K (so $\Delta\log(T)$ is constant and can be dropped)
- K is a total number of chemical elements that we include in the model (K=14)

Then the total expected photon counts at energy bin j will be:

$$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)' = \sum_{k=1}^{K} \left\{ \boldsymbol{\lambda}^{C,k} + binned(\boldsymbol{\lambda}^{L,k}) \right\} \propto \left[\sum_{k=1}^{K} \gamma_k \left(\mathbf{G}^{C,k} + binned(\mathbf{G}^{L,k}) \right) \right] \boldsymbol{\mu}$$

Actual counts can be modeled as $Pois(\lambda_j)$

Note, that these counts are “true” counts, the ones that we can’t really observe. Let’s look at what we get to observe…



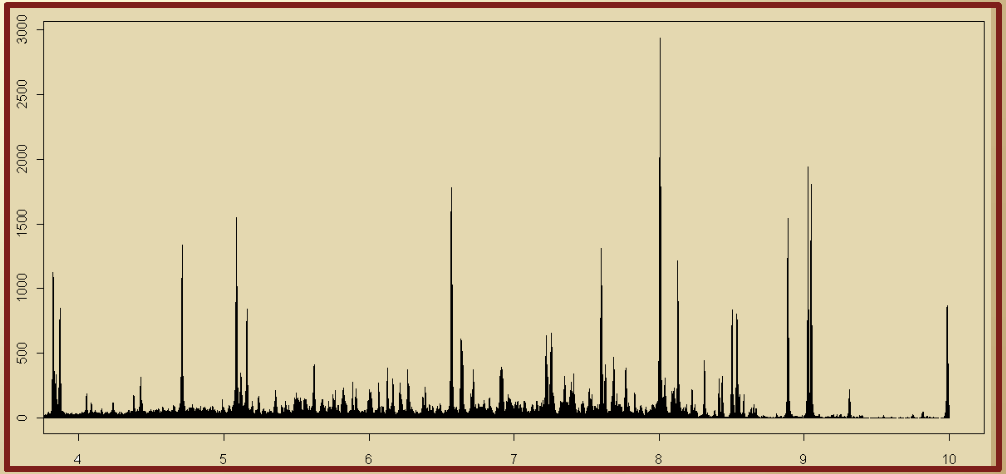
Data source

- This project uses data that are collected from a yellow giant star Capella by Chandra X-ray Telescope

- Detectors register photon arrival time, its energy and 2-dimentional direction of arrival – *and we only focus on energy variable!...*

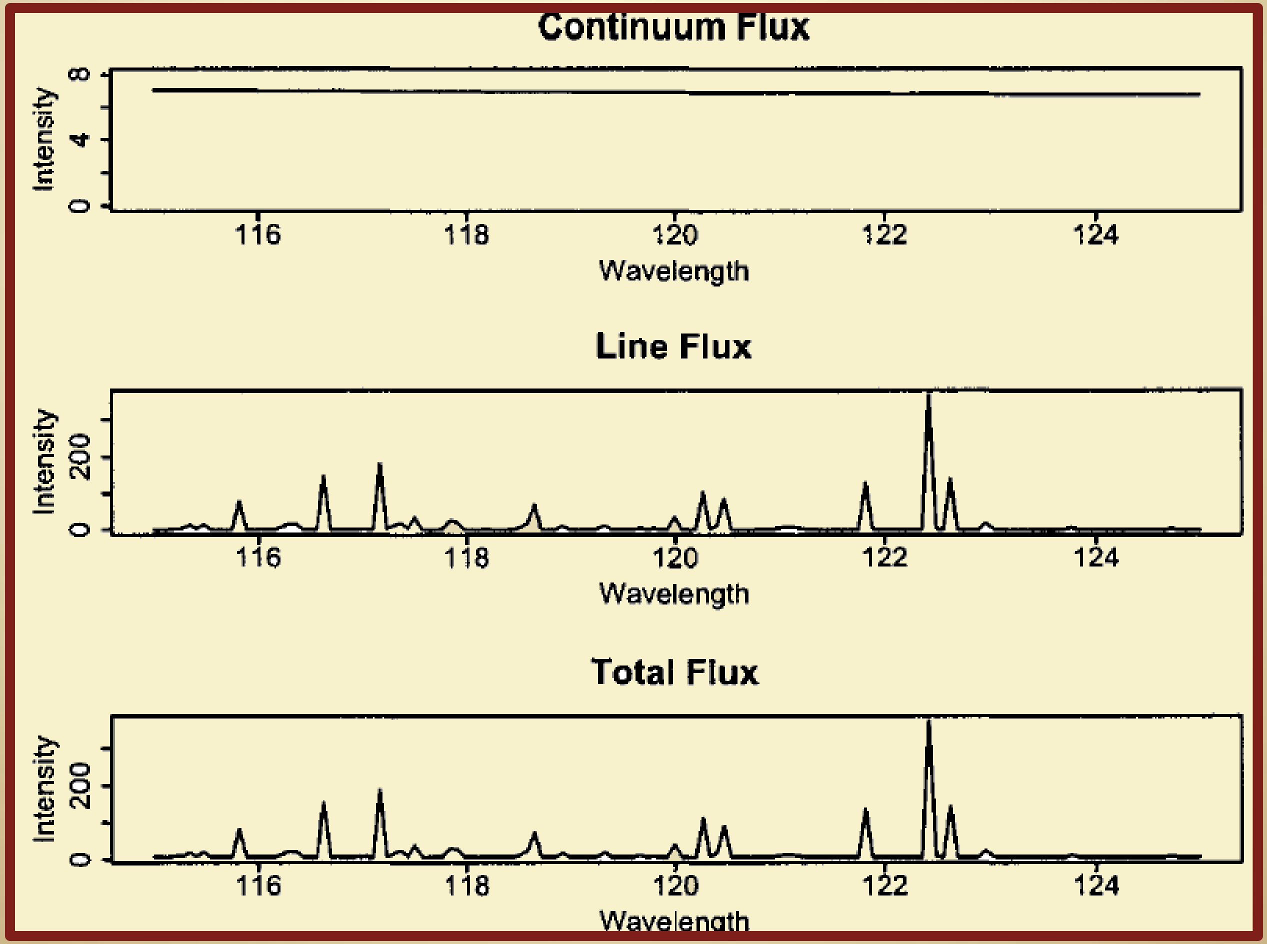
- Due to its digital nature, *Chandra* records the energy in a certain prespecified number of channels

So the data looks like this:



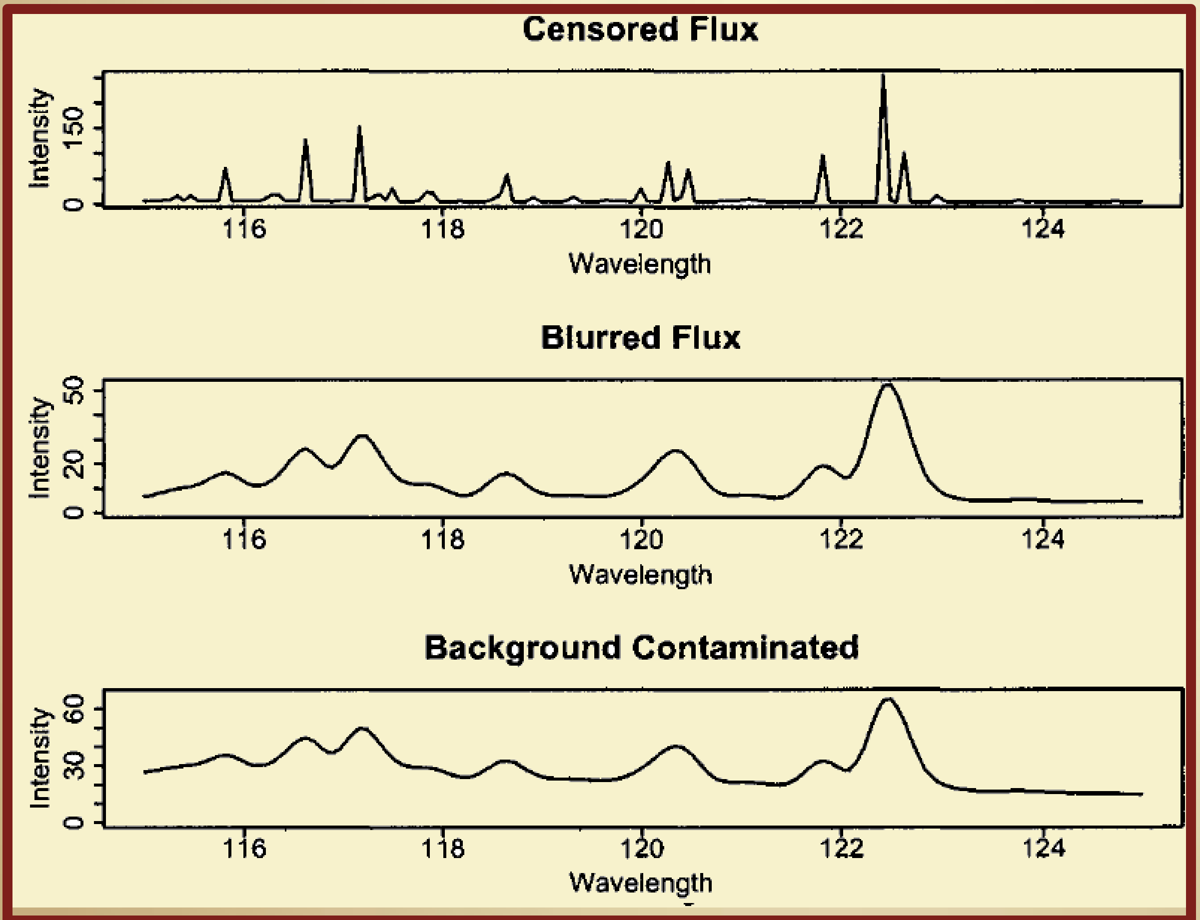
Degradation of true intensity

As we noted, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)'$ are true counts that are coming from the star:



Degradation of true intensity

After that, various stochastic processes significantly degrade the source model and result in Poison intensities for the observed counts:



Stochastic censoring *(effective area)*

Measurement error *(blurring)*

Extraneous photons, coming from other objects

Modeling observed counts

- Two major data distortion effects, introduced by the instrument itself, are modeled as follows:
  - *Effective area* of bin j is reflected by d_j, a probability that the X-ray is recorded by detector. It depends on the photon energy and is provided by calibration scientists.

  - *Blurring* of arriving photons can be described by Redistribution Matrix File (RMF) M_{ij}, that provides parameters for a multinomial model of redistribution of incoming photons of energy j into channels 1 … I. (That is exactly what Li Zhu is working on)

  It turns out that for this high-resolution data RMF can be very well represented by scaled *t*-distribution with df=4.

- After that we add the rate of background noise

Final model for observed counts

Observed data in channel i is $Y_i^{obs} \sim Pois(\xi_i)$, where

$$\xi_i = \sum_{j=1}^{J} M_{ij} d_j \lambda_j + \lambda_i^B$$

In matrix notation we get

$$\boldsymbol{\xi} = \boldsymbol{MD\lambda} + \boldsymbol{\lambda}^B = \boldsymbol{MD}\left(\sum_{k=1}^{K} \gamma_k \{G^{C,k} + binned(G^{L,k})\}\right)\boldsymbol{\mu} + \boldsymbol{\lambda}^B$$

The **goal** is to draw inference about **μ, γ** and **ξ.**

Next, we will briefly introduce the fitting procedure. The idea is to impute back all photons that were blurred and censored.

Hierarchical Missing Data Imputation

- Let's introduce levels of missing photon counts:
  - Channel level **Y**, ranges from 1 … I (the one we observe)
  - Bin level **Z**, ranges from 1 … J (data coming from the star) (usually J=I)
  - Temperature level **U**, ranges from 1 … T

- We will use superscript "**Z^-**" to denote a lower level of augmentation within each level
  - **Z^-** are bin level counts after censoring, where $Z \sim Pois(\lambda)$

It turns out that there is censoring on *temperature level*! Here is why:

Remember the *total emissivity function :* $G(t) = \Sigma\gamma_k(G_j^{C,k}(t) + binned\{G_l^{L,k}(t)\})$
It's column sums ARE NOT EQUAL, so photons originated from different temperatures have different rates to be emitted this "censoring" doesn't occur unfirmly across the temperature.

Imputation Algorithm

Given current estimates of abundance $\gamma^{(i)}$ and DEM $\mu^{(i)}$:

- Compute line and continuous intensities

$$\lambda^{L} = binned\{\sum_{k=1}^{K}\gamma_k \mathbf{G}^{L,k}\boldsymbol{\mu}\} \quad \text{and} \quad \lambda^{C} = \sum_{k=1}^{K}\gamma_k \mathbf{G}^{C,k}\boldsymbol{\mu}$$

- Estimate bin intensity (expected detector counts) using data distortion model (no background contamination):

$$\xi^{*} = MD(\lambda^{L} + \lambda^{C})$$

- Split Y^{obs} into background and channel photon counts:

$$Y \sim Binom(Y^{obs}, \frac{\xi^{*}}{\xi^{*} + \lambda^{B}})$$

- "Deblurr" Y (restore the blurred photons – note that we get to bin levels after that)

$$Z_i^{-} \mid Y \propto \sum_{i=1}^{I} Multinom(Y_i, \frac{(M_{1i}d_1\lambda_1, \ldots, M_{Ji}d_J\lambda_J)'}{\sum_j M_{ji}d_j\lambda_j})$$

Imputation Algorithm

- "Desensor" Z^-_i (restore absorbed counts due to effective area) to get actual bin counts

$$Z_j \mid Z_j^- \sim Z_j^- + Pois((1-d_j)\lambda_j)$$

- Separate bin counts into counts coming from line emission and countinuum

$$Z_j^L \sim Binom(Z_j, \frac{\lambda_j^L}{\lambda_j^C + \lambda_j^L})$$

- Distribute line emission counts between actual lines that fall in bin j (same multinomial principle) and reestimate elemental abundance $\gamma^{(i+1)}$

- In the mean time, we use **Z** again to restore temperature counts

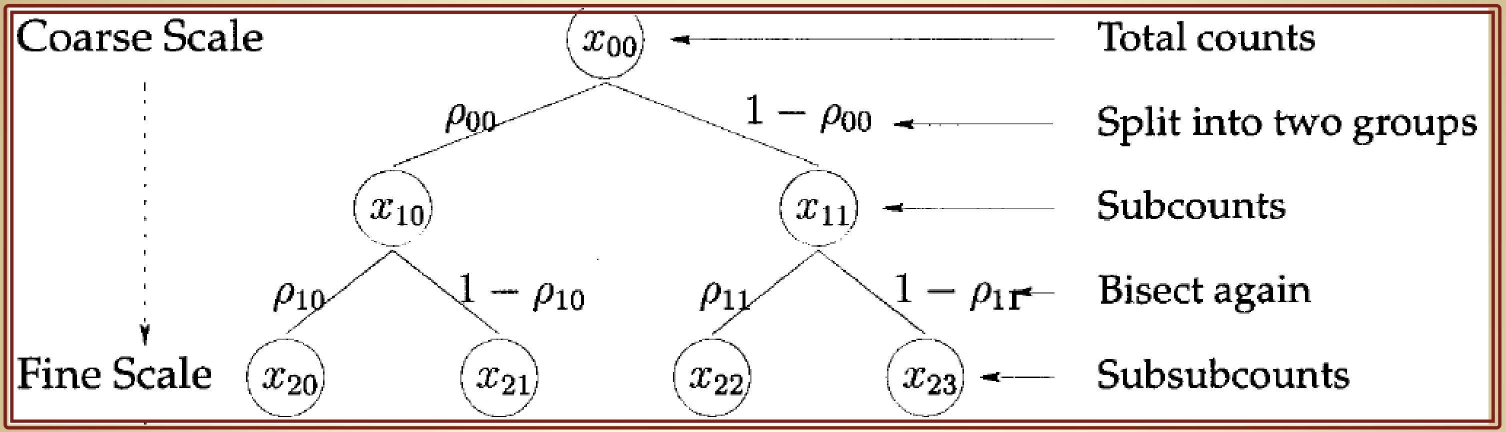
$$\boldsymbol{U}^- \mid \boldsymbol{Z} \sim Multinom(Z_j, \frac{(G_{j1}^* \mu_1, \ldots, G_{jT}^* \mu_T)'}{\sum_t G_{jt}^* \mu_t})$$

- Restore censored temperature counts

$$U_t \mid U_t^- \sim U_t^- + Pois((1-g_{+,t}^*)\mu_t)$$

Multiscale smoothing

After obtaining ***U*** we reestimate $\mu^{(i+1)}$ as follows:



- Considering $U_t \sim Pois(\mu_t)$, we choose R=6 scales so that T=2^6=64.
- At the finest scale $x_{R,n} := U_{n+1}$ for n=0,…,2^R-1

This produces the same binary tree for DEM parameters:

$$\mu_{R,n} := \mu_{n+1} \text{ and } \mu_{r,n} = \mu_{r+1,2n} + \mu_{r+1,2n+1} = \rho_{r,n}\mu_{r,n} + (1-\rho_{r,n})\mu_{r,n}$$

Multiscale smoothing

- Then a smoothing conjugate prior is introduced for weights

$\rho_{r,n} \sim Beta(\alpha_r, \alpha_r)$ and $\mu_{0,0} \sim Gamma(a,b)$

Note, that by varying α_r we can control the amount of smoothing that we need (the larger - the smoother)

- Then we either sample from posterior distributions or find MAP estimates of weights and mean total counts

- Mean level counts can be restored using recursion from the most coarse scale to the finest:

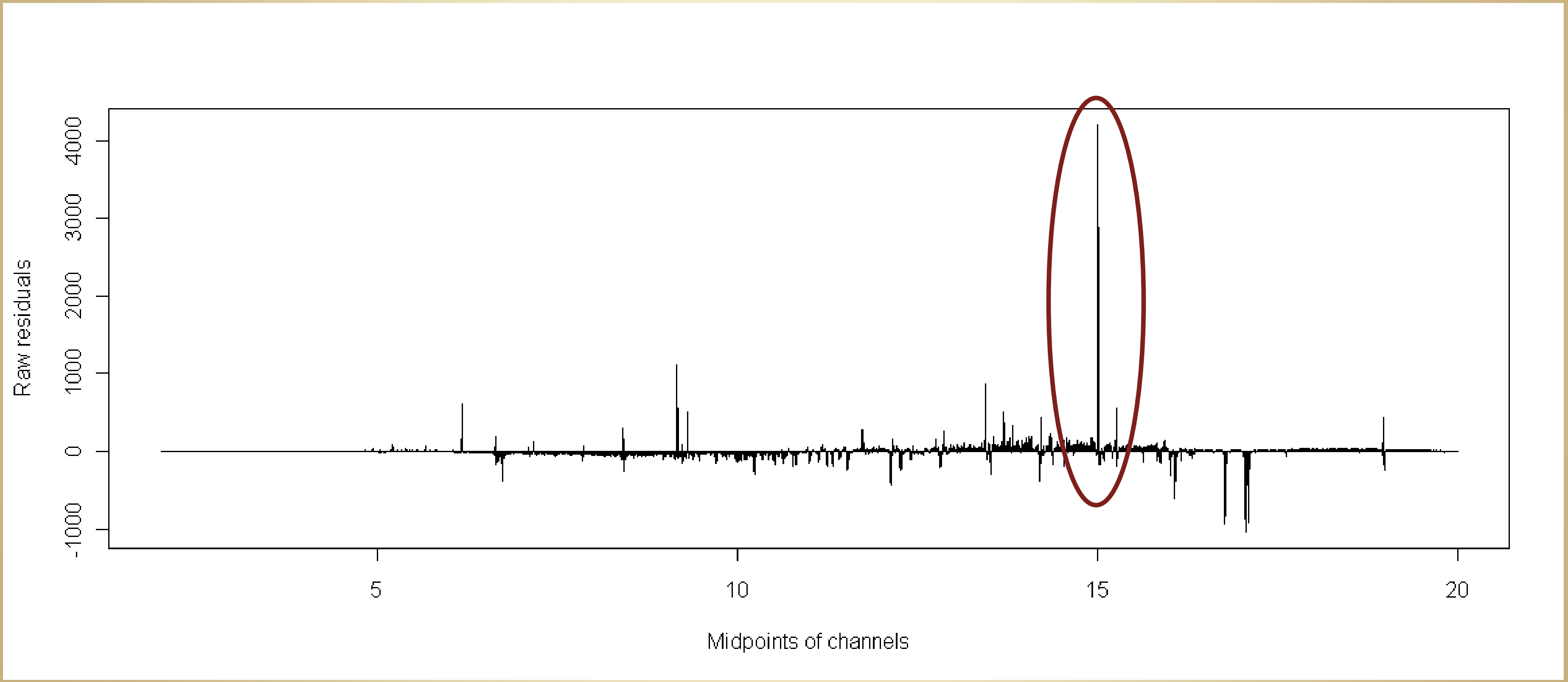
$\mu_{r+1,2n} = \rho_{r,n}\mu_{r,n}$ and $\mu_{r+1,2n+1} = (1-\rho_{r,n})\mu_{r,n}$

$r = 0 \ldots R$; $n = 0, \ldots, 2^R - 1$

Ref: Nowak and Kolaczyk (2000)

Results: Raw residuals (difference between MLE for expected counts and observed counts) for Capella

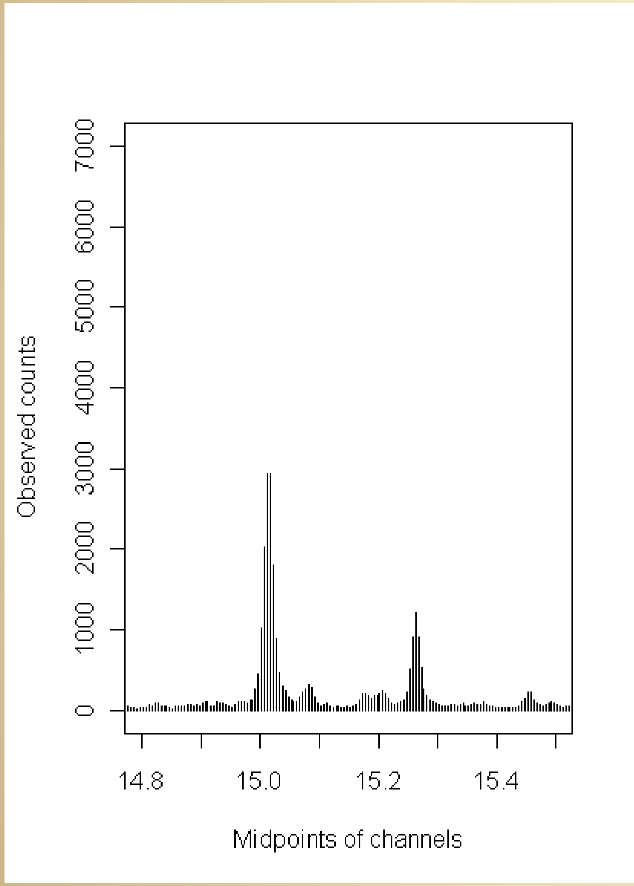
We use data obtained from MEG spectrum of Chandra's HETGS



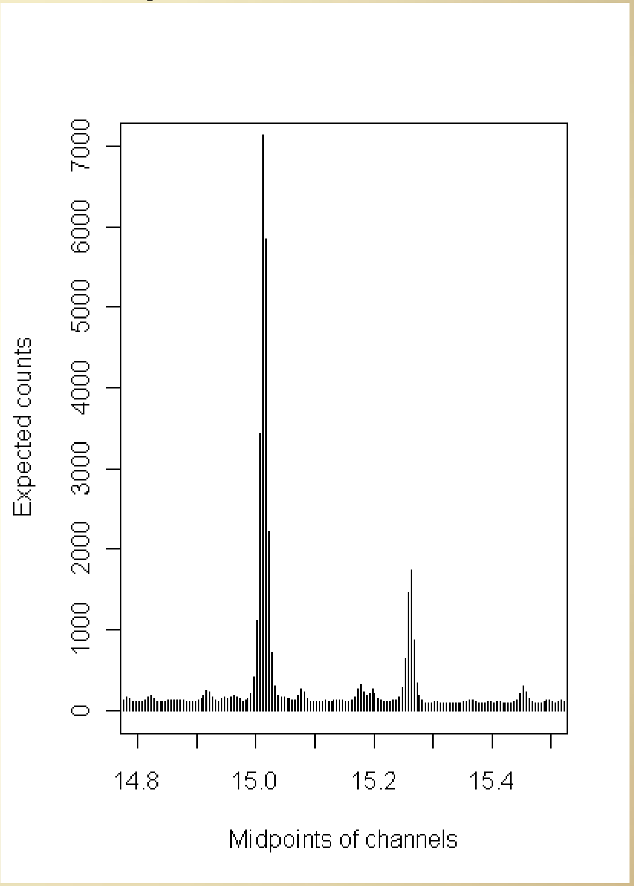
- Huge discrepancy around ~15 A
- We can interpret *positive* residual as potential miscalculated line and *negative* residual as missing line.

Magnified discrepancy

Observed counts

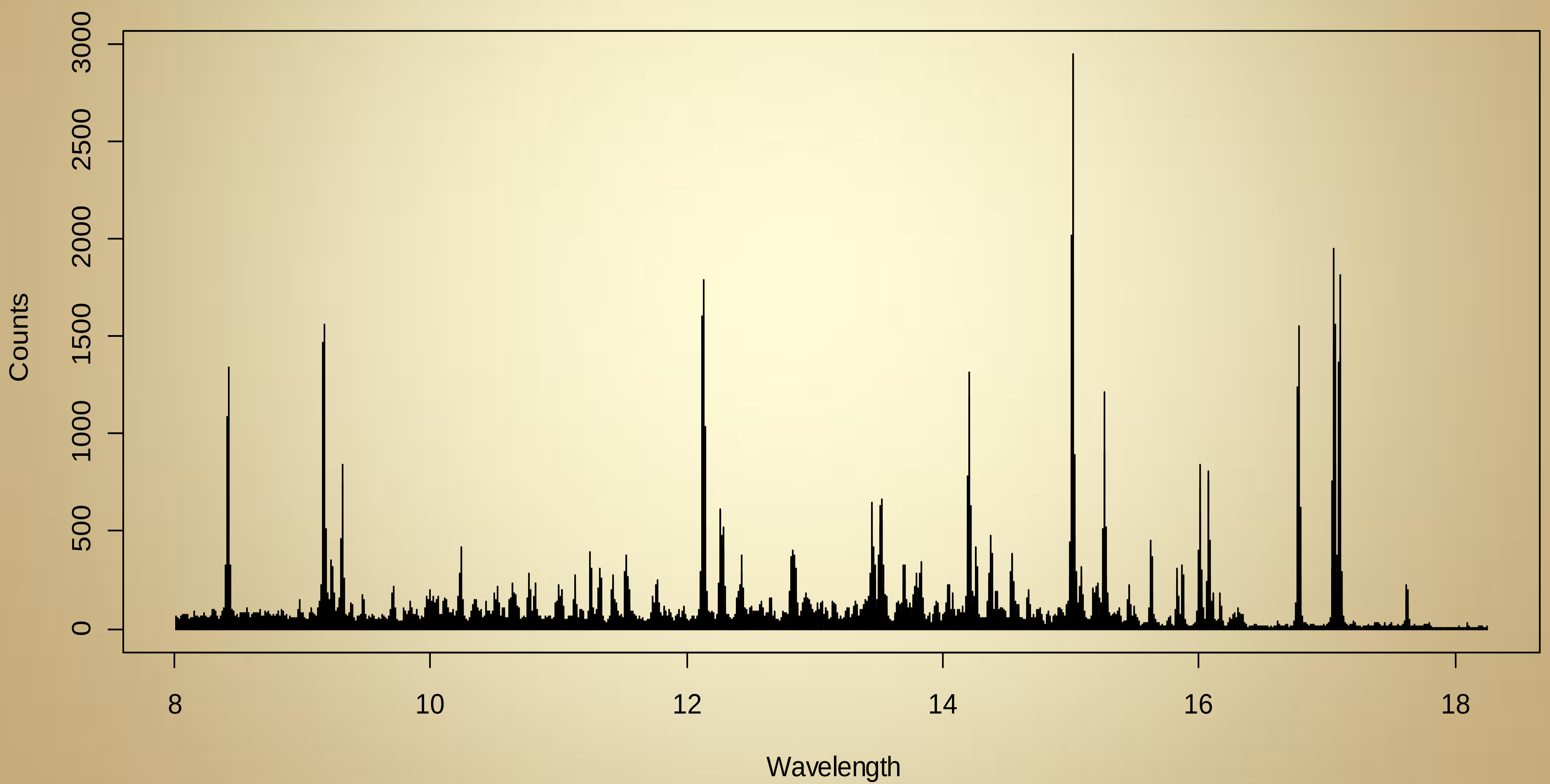


Expected counts

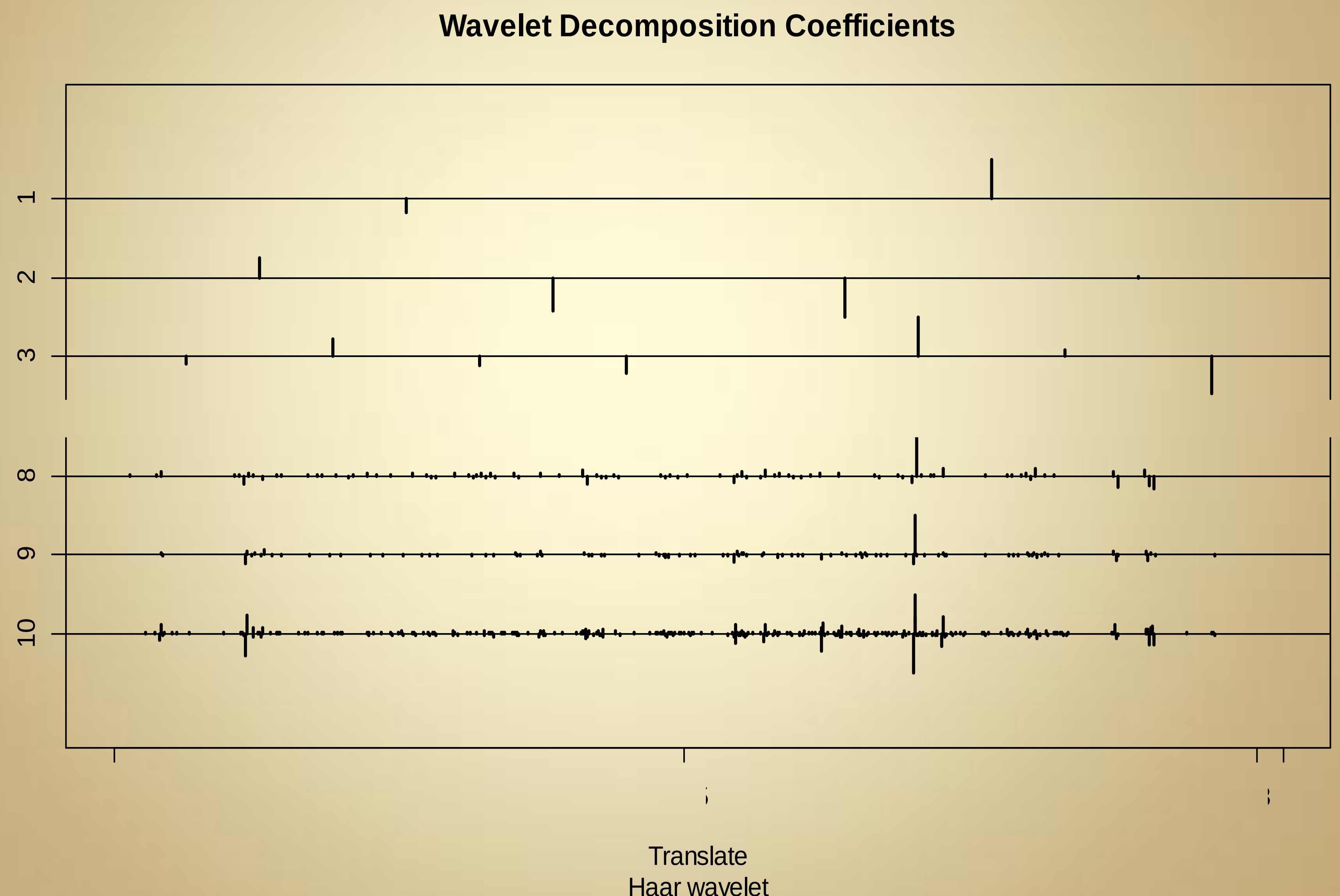


Possible reason could be considerably small ARF in that region? (~6%)

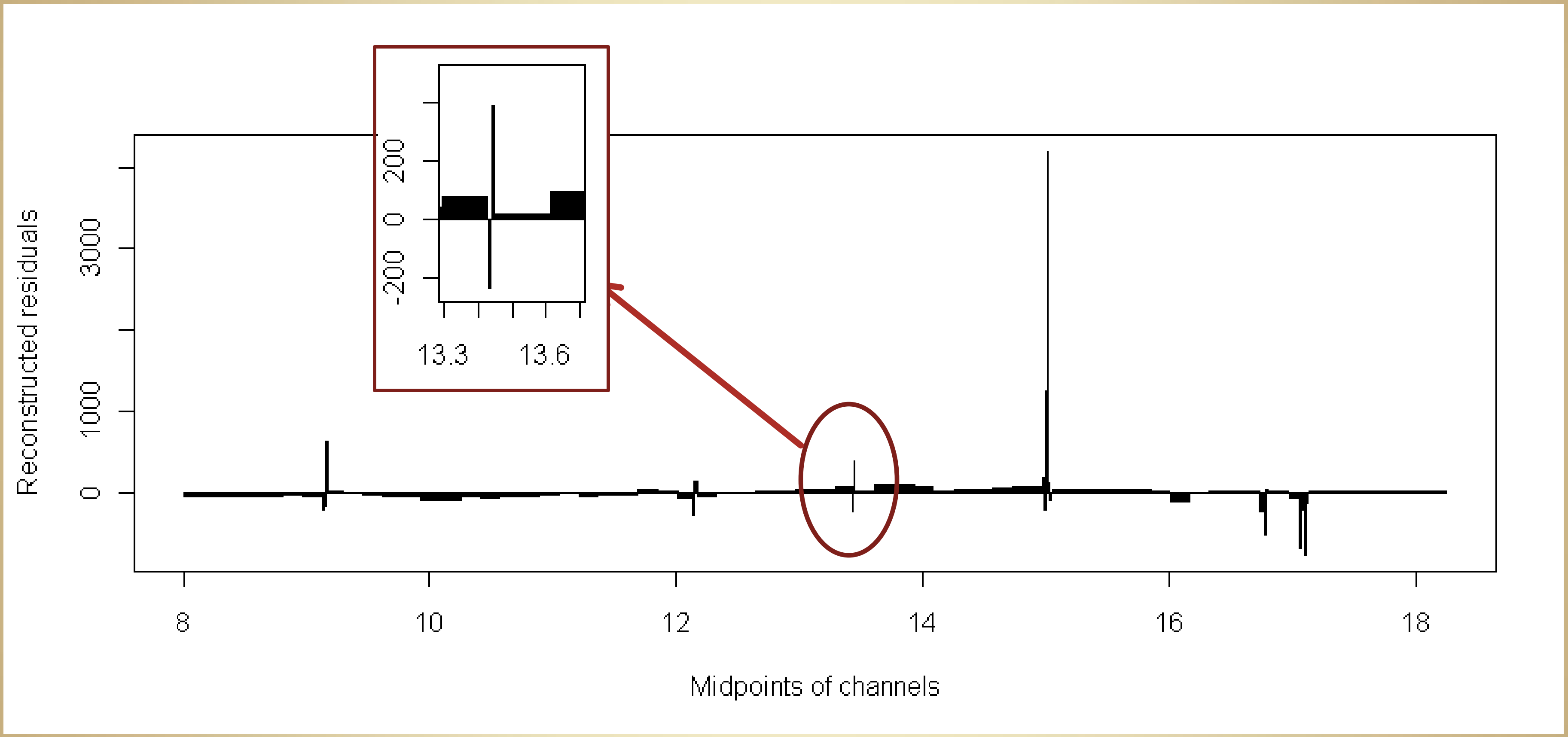
Observed counts prepared for wavelength decomposition



(Haar) Wavelet decomposition of estimated expected counts



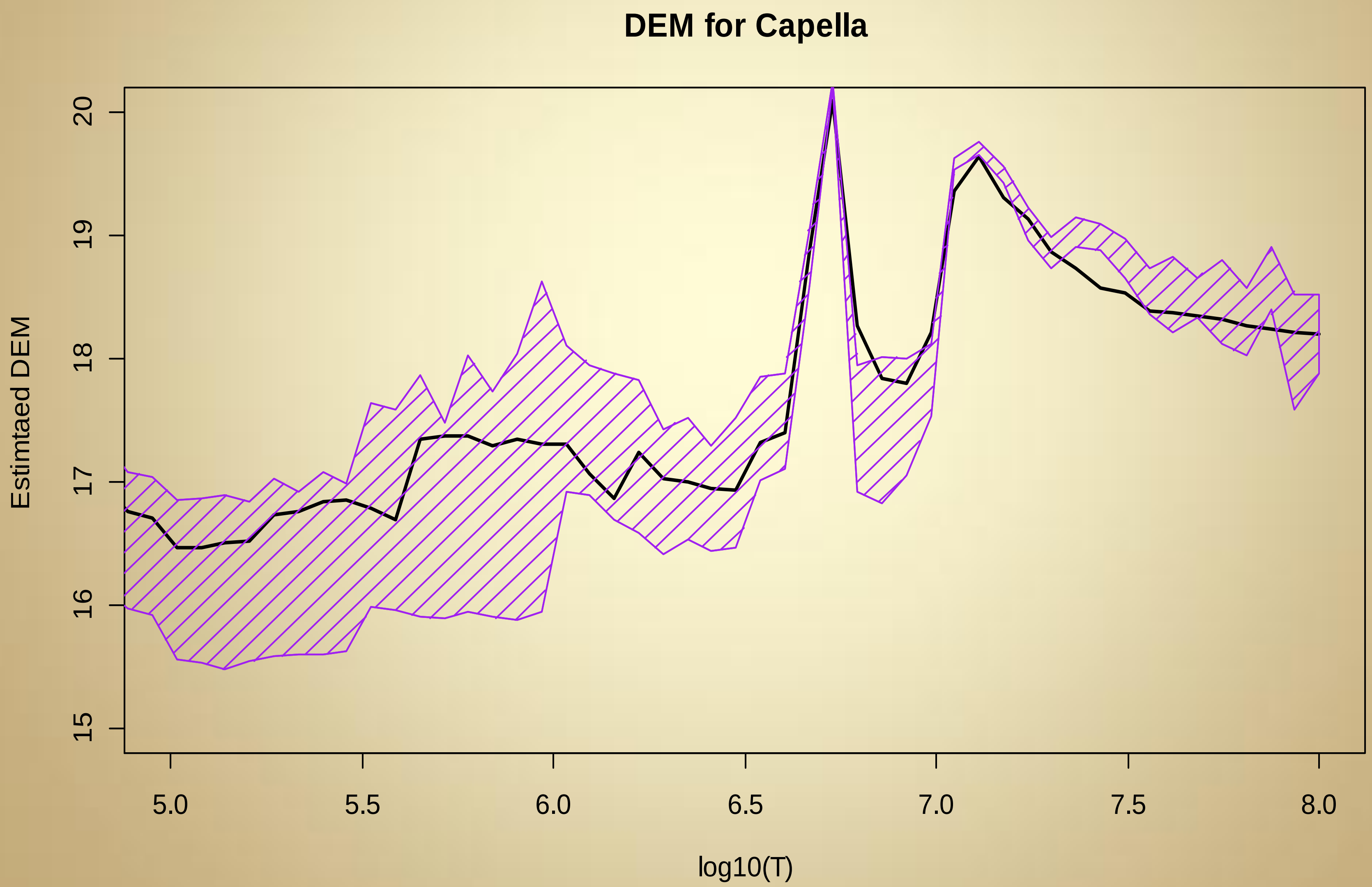
Reconstructed residuals using universal thresholds



Another region of interest is around ~13.5 A (this region has many unidentified Ne IX lines)

Negative residuals are in [13.43, 13.44], positive in [13.44, 13.45]

Reconstructed DEM with MAP taken from EM and 95 % posterior interval produced by MCMC



Future work

- Future work, long-term goals:
  - New implementation
  - Improved algorithm for DEM and abundance estimation (recalculate emissivity matrix for new abundance, convergence check)
  - Wavelet decomposition interpretation
  - A way of combining data from HETGS and LETGS and analyzing it
  - Include time component