Nested sampling for Fully Bayesian Unfolding (PENDING TITLE)

by

Jens Bratten Due

THESIS

for the degree of

MASTER OF SCIENCE

Faculty of Mathematics and Natural Sciences
University of Oslo

May 2021

Abstract

This here is the abstract

Acknowledgements

Contents

Introduction

1   Introduction

Part I

Theory

Bayesian statistics

> Probability theory is nothing
> but common sense reduced to
> calculation.
>
> ———————————————
> Pierre-Simon Laplace [1]

Sivias book *Data Analysis - a Bayesian tutorial* [1] is a great read, and provides the theoretical foundation for the majority of topics discussed in this SECTION/CHAPTER.

1 Bayes' theorem

First, we consider probability theory and its basic algebra which includes the sum rule

$$P(X|I) + P(\overline{X}|I) = 1 \tag{2.1}$$

and the product rule

$$P(X, Y|I) = P(X|Y, I) \times P(Y|I). \tag{2.2}$$

Here P stands for probability, the bar "|" means "given" and $\overline{X}$ means "not X". Lastly, we have the symbol I, meaning all relevant background information. The sum rule can then be stated as "the probability of X being true plus the probability of X not being true, both given all relevant background, equals 1".

Using the product rule, and the fact that $P(X, Y|I) = P(Y, X|I)$ we get the following.

$$P(X|Y, I) \times P(Y|I) = P(Y|X, I) \times P(X|I)$$

Rearranging this leads to *Bayes' theorem*

$$P(X|Y, I) = \frac{P(Y|X, I) \times P(X|I)}{P(Y|I)} \tag{2.3}$$

To get a clearer picture of the significance of Bayes' theorem, we can replace X and Y with *hypothesis* and *data*. $P(hypothesis|data, I)$ is then given the formal name *posterior probability*,

$P(data|hypothesis, I)$ is called the *likelihood function* and $P(hypothesis|I)$ is called the *prior probability*, representing our knowledge about the truth of the hypothesis before any data has been analysed. The term in the denominator, $P(data|I)$, often called the *evidence*, is in many cases not shown, due to it often being absorbed by a normalization constant. We can then replace the equality sign with a proportionality.

$$P(hypothesis|data, I) \propto P(data|hypothesis, I) \times P(hypothesis|I) \tag{2.4}$$

In summary, Bayes' theorem describes a learning process, showing how a probability is augmented by the introduction of data.

Another useful result from using the sum and product rule is the *marginalization* equation

$$P(X|I) = \int_{-\infty}^{\infty} P(X, Y|I)dY \tag{2.5}$$

with a normalization condition

$$\int_{-\infty}^{\infty} P(Y|X, I)dY = 1. \tag{2.6}$$

The marginalization equation gives us the ability to integrate out so-called nuisance parameters, values of no interest to a specific problem, such as background signals and measurement byproducts. These rules of probability are widely applicable and provide a strong foundation for tackling data analysis problems. [1]

2   Parameter estimation I

We will now look at the act of estimating a single parameter using Bayes' theorem, such as the mass of a planet, or the charge of the electron. We will firstly go through the example of deducting the fairness of a coin. This can be represented by the *bias-weighting H*. $H = 1/2$ will mean the coin is fair, while $H = 1$ and $H = 0$ means the coin is showing only heads or tails every flip. This value is continuous on the range $[0, 1]$, and $P(H|\{data\}, I)$ describes how much we believe H to be true. For a range of H-values, $P(H|\{data\}, I)$ is a *probability density function* (pdf). To find this, we use Bayes' theorem.

$$P(H|\{data\}, I) \propto P(\{data\}|H, I) \times P(H|I) \tag{2.7}$$

We can, if needed, find the normalization constant using equation (2.6). To express ultimate ignorance, we can assign a flat pdf for the prior.

$$P(H|I) = \begin{cases} 1 & 0 \leq H \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

meaning we assume every value of H to be equally probable. Assuming each flip is an independent event, the likelihood function takes the form of the binomial distribution.

$$P(\{data\}|H, I) \propto H^R (1 - H)^{N-R} \tag{2.9}$$

where R is the number of heads and N is the number of flips.

Plugging (8) and (9) into Bayes' theorem results in the posterior probability, the shape of which varies significantly for the first few data points. When the number of data increases however, the pdf becomes sharper and converges to the most likely value. The choice of prior becomes mostly irrelevant when we have a large of number of data, as the majority of propositions will converge to the same solution, but the speed of convergence may vary. A very confident, but wrong prior will often approach the correct solution more slowly than an ignorant one.

2.1 Reliabilities: best estimates, error-bars and confidence intervals

One way to summarize the posterior pdf is with two quantities: the best estimate and its reliability. The best estimate is given by the maximum value of the pdf

$$\left.\frac{dP}{dX}\right|_{X_O} = 0 \tag{2.10}$$

where X_O denotes the best estimate. To make sure we have a maximum, we also need to check the second derivative

$$\left.\frac{d^2P}{dX^2}\right|_{X_O} < 0. \tag{2.11}$$

Using derivatives like this assumes X is continuous. If this is not the case, the best estimate will still be the value corresponding to the max of the pdf.

The reliability of the best estimate is found by considering the width of the pdf about X_O. We take the logarithm of the pdf as this varies more slowly with X, making it easier to work with.

$$L = ln[P(X|\{data\}, I)]. \tag{2.12}$$

Doing a Taylor expansion about X_O and using the condition

$$\left.\frac{dL}{dX}\right|_{X_O} = 0, \tag{2.13}$$

which is equivalent to (10), leads to

$$P(X|\{data\}, I) \approx A \exp\left[\frac{1}{2}\left.\frac{d^2L}{dX^2}\right|_{X_O}(X - X_O)^2\right]. \tag{2.14}$$

Here, we only show the dominating quadratic term of the expansion, with A being a normalization constant. We have now approximated our pdf by the *normal distribution*, typically taking the form

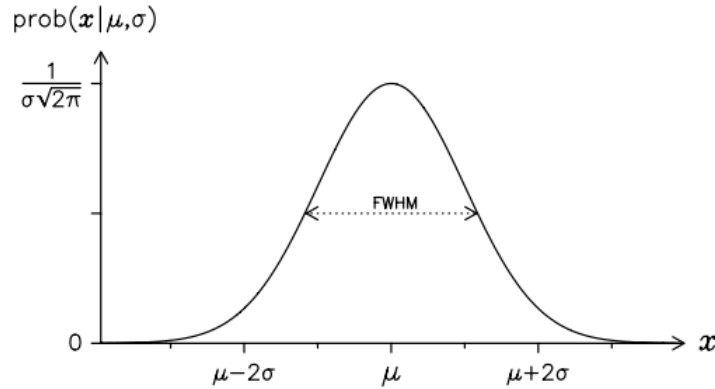$$P(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \tag{2.15}$$

Figure 2.1: The normal distribution with a maximum at $x = \mu$ and a full width at half maximum (FWHM) of 2.35σ. (Sivia, 2006, p. 22) [1]

The parameter σ is called the *error-bar* and is defined as

$$\sigma = \left(- \frac{d^2 L}{dX^2}\bigg|_{X_O} \right)^{-1/2}. \tag{2.16}$$

We then infer the quantity of interest by the following

$$X = X_O \pm \sigma. \tag{2.17}$$

By calculating the integral of the normal distribution in this range, we get a 67% chance that X lies within $X_O \pm \sigma$ and a 95% chance that it lies within $X_O \pm 2\sigma$.

2.1.1 Asymmetric pdfs

The error-bar needs a symmetric pdf to be valid, something that is often not the case. This is solved by replacing the error-bar with a *confidence interval* as a measure of reliability. It is defined as the shortest interval that encloses 95% of the area of the pdf. In short, we find X_1 and X_2 such that

$$P(X_1 \leq X \leq X_2|\{data\}, I) = \int_{X_1}^{X_2} P(X|\{data\}, I)dX \approx 0.95, \tag{2.18}$$

assuming the pdf is normalized. The 95% confidence level is conventionally seen as a sensible value, being a rather conservative estimate.

In the case of an asymmetric pdf, we may consider using the *mean* or *expectation* as the best estimate. This quantity takes the skewness of the pdf into account, and is given by

$$\langle X \rangle = \int XP(X|\{data\}, I)dX. \tag{2.19}$$

If the pdf is not normalized, we also need to divide the right-hand side by $\int P(X|\{data\}, I)dX$.

If the pdf is *multimodal*, meaning it has multiple maxima, it becomes more difficult to calculate a best estimate and its reliability. If one maximum is much greater than the
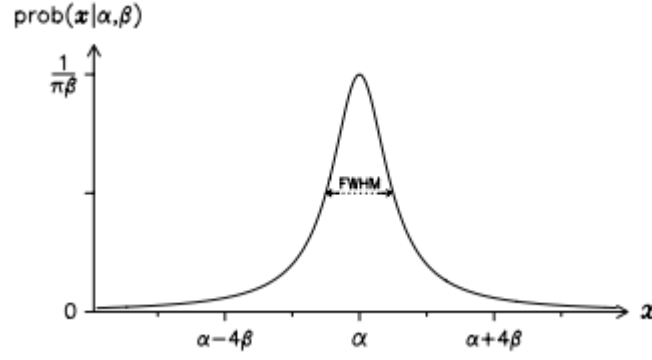
Figure 2.2: The Cauchy distribution, symmetric about $x = \alpha$ and has a FWHM of 2β. (Sivia, 2006, p. 31) [1]

others, we can ignore those other contributions and focus on the largest. However, if multiple peaks are of similar size, we would be better off displaying the pdf itself.

Another common pdf which is often used is the Cauchy distribution, shown below. This distribution has very wide wings and is in this case given by

$$P(x|\alpha, \beta, I) = \frac{\beta}{\pi[\beta^2 + (x - \alpha)^2]} \tag{2.20}$$

Part II

Methods

Unfolding Methods

1 What is unfolding (maybe in Theory-chapter?)

- The inverse problem $f = \boldsymbol{R}u$

- The physics behind?

- The detector/experimental setup

- SiRi, OSCAr

- The raw data matrix, how it is produced?

1.1 The Response Matrix

The response matrix contains what we know about the circumstances of the experiment, a large part of which are the characteristics of the detector. The response matrix has elements given by:

$$R_{ij} = P(\text{measured in bin i} \mid \text{true in bin j}) \tag{3.1}$$

[2]. This can be read as the probability of observing an event in energy bin i, given the true event happening in bin j. In a nutshell, the response matrix describes how a signal is smeared over the other bins in the spectrum.

- Multiplying with $\boldsymbol{R}^{-1}$ leads to fluctuations because we cannot assume the observed data equals the *expectation values* for the data. Statistical fluctuations in the data is assumed to come from a real structure in the true spectrum and will be magnified. http://www-library.desy.de/preparch/desy/proc/proc14-02/P52.pdf

2 The Folding Iteration Method

The following section describes the methods developed by Guttormsen et al. [3].

The folded spectrum f is on the form

$$f = \boldsymbol{R}u, \tag{3.2}$$

where $\boldsymbol{R}$ is the response matrix and u is the expectation values for the true spectrum. The iterative method can then be described in 4 parts.

- First we use the measured spectrum r as an initial guess,
  $u_0 = r$

- We then fold this with the response matrix,
  $f_0 = \boldsymbol{R}u_0$

- The difference between the folded and the raw spectrum is calculated and added to the initial guess, and we end up with the next trial spectrum,
  $u_1 = u_0 + (r - f_0)$

- This is then repeated according to the following iteration scheme,
  $u_{i+1} = u_i + (r - f_i)$

This method is performed until $f_i \approx r$ within the experimental uncertainties [3]. It is important to note that for each new iteration, the oscillations between channels increase, as the solution approaches the inverted matrix solution $u = \boldsymbol{R}^{-1}r$, which exhibits large oscillations. [2][3]

2.1 The Compton Subtraction Method

As the resulting spectrum from the folding iteration method often contains some degree of fluctuations, the Compton subtraction method is performed to obtain a significantly more stable spectrum.

The first step is to define a new spectrum $v(i)$ as the observed data excluding the Compton contribution:

$$v(i) = p_f(i)u(i) + w(i), \tag{3.3}$$

where $u(i)$ is the spectrum obtained from the folding iteration method, which multiplied with p_f gives the full energy contribution. The remaining contributions are contained in $w(i) = u_s + u_d + u_a = p_s(i)u(i) + p_d(i)u(i) + \sum p_{511}(i)u(i)$, representing single escape, double escape and annihilation (note the missing Compton contribution $"u_c"$). To match the observed energy resolution, each contribution is then smoothed with a Gaussian function. Next, we subtract this from the raw spectrum to obtain the Compton background spectrum:

$$c(i) = r(i) - v(i). \tag{3.4}$$

This spectrum may exhibit significant oscillations, and is thus further smoothed. This smoothing carries a low risk of loss of important information due to the nature of the

spectrum not containing any sharp, narrow peaks. After this smoothing procedure on the individual contributions, we now "return" to the unfolded spectrum like so:

$$u(i) = \frac{r(i) - c(i) - w(i)}{p_f}.$$ (3.5)

Finally, to get closer to the true number of events, we correct for the total detector efficiency:

$$U(i) = \frac{u(i)}{\epsilon_{tot}(i)}.$$ (3.6)

This final spectrum shows higher stability compared to the result of the iteration method, while keeping similar statistical fluctuations to the raw spectrum. [3]

3 Fully Bayesian Unfolding

Bayes' theorem succinctly describes what we are asking for in the problem of unfolding, showing the relation between the truth spectrum T, and the data we have obtained D.

$$P(T|D) \propto L(D|T) \cdot P(T)$$ (3.7)

The expected truth spectrum T and the raw spectrum D are binned with N_t and N_r bins, respectively. Each bin in T is assigned a prior probability distribution, describing our belief of the number of events expected to be present. We assume the data follows a Poisson distribution, meaning

$$L(D|T) = \prod_{r=1}^{N_r} \frac{f_r^{D_r}}{D_r!} e^{-f_r}$$ (3.8)

where

$$f_r = \sum_{t=1}^{N_t} T_t \cdot R_{rt}.$$ (3.9)

Here, R_{tr} is the element of the response matrix $R^{N_r \times N_t}$, corresponding to the probability that an event produced in the truth bin t is observed in the response bin r: $P(r|t)$. If we wish to include the background, all we have to do is add it to the sum:

$$f_r = B_r + \sum_{t=1}^{N_t} T_t \cdot R_{rt}.$$ (3.10)

The next step is to employ a sampling scheme of the parameter space, usually a MCMC algorithm, to calculate $L(D|T) \cdot P(T)$ and arrive at a posterior distribution per bin in the expected truth spectrum.

3.1 Priors

- Uniform

There is practically an infinite amount of choices one can make for assigning a prior, depending on what knowledge one has beforehand. If one wishes to make the least amount of assumptions about the truth, a *uniform* prior is suitable. The pdf of the uniform distribution is given as:

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

Ref? This flat distribution assigns equal probability to every outcome in the space of possibilities. The only assumption to be made here is determining the boundaries on this space. Complete ignorance would strictly be represented with a uniform prior without any boundaries. This would mean we believe all numbers on the interval $[-\infty, \infty]$ to be equally likely in a one-dimensional space. Such a space is of course not possible to explore completely, and otherwise extremely large limits will be computationally unfeasible. This is especially true considering the fact that many problems are complex and demand multidimensional parameter spaces. In addition to this, unfolding in physics is often related to physical experiments pertaining to the counting of a number of events measured by a detector. In these cases, the existence of negative counts is unphysical, meaning a lower prior limit can safely be set to 0 (Discuss possibility for negative counts?).

Choosing the upper limit is not as straightforward. The ideal choice would be the largest possible limit that still allows for reasonable computational performance. Of course, if we have some knowledge about the size and location of the domain of the possible truth-values, there is no need to pick a limit located significantly beyond this domain. Computational resources are wasted if spent on exploring a region we strongly believe (know?) will not improve our estimate.

Make sure the folded of prior contains the raw spectrum?

In this thesis, an upper prior limit of 10 times the raw data will be used for the uniform prior. In other words, we believe that the true value must be contained within an area relatively close to the observed value.

- Image of uniform distribution

- Logarithmic (using 'interpolate' in pymc3)

Another prior distribution we will use is the *log-uniform* distribution, also called the reciprocal distribution. This distribution has the characteristic that its logarithm is uniformly distributed. What this means for our prior belief is that each order of magnitude is given equal probability. In the case of a logarithm with base 10, we say that it is equally probable for our value of interest to lie between 10^0 and 10^1, as between 10^6 and 10^7, even though the second range is much larger. The pdf of the log-uniform distribution is defined as:

$$f(x; a, b) = \frac{1}{x \ln(b/a)} \quad \text{for } a \leq x \leq b \text{ and } a > 0. \tag{3.12}$$

Ref?

- Image of logarithmic distribution

Using the log-uniform distribution allows us to
Say we know there exists one or multiple peaks in our truth-spectrum consisting of a very large amount of counts, say $\sim 10^{10}$. If we also know that other bins in the spectrum should have close to 0 counts, how do we make sure both of these conditions are met? If we were to use a uniform prior between 0 and 10^{10}, we would firstly have an incredibly large space to explore, with 10^{10} possible values for each bin in the spectrum. Secondly, the probability of sampling a value close to 0 would be very small. Let's say that any value between 0 and 10000 is considered 'close' to 0, something that can be seen as very imprecise. According to the uniform prior, the probability of the true value being 'close' to 0 is thus:

$$P(0 \leq T_t \leq 10^4) = \int_0^{10^4} \frac{1}{10^{10} - 0} dT_t = \frac{10^4}{10^{10}} = 10^{-6} = 0.0001\%, \tag{3.13}$$

and the same result holds for values equally 'close' to the maximum of 10^{10}. This low probability means we might run out of computational resources long before our algorithm gets to explore those areas.

Now, if we instead use the log-uniform distribution as our prior, the probability of a value close to 0 is much higher, while still allowing for those tall peaks.

$$P(1 \leq T_t \leq 10^4) = \int_1^{10^4} \frac{1}{T_t \ln\left(10^{10}/1\right)} dT_t = 0.4 = 40\%. \tag{3.14}$$

Note here that we integrate from a count value of 1 instead of 0, as the distribution is undefined at 0. Since the logarithm of this function is uniformly distributed, looking at the exponents will give a good indication of the probability value. In the simple case above, the exponents are 4 and 10, $4/10 = 0.4 = 40\%$. Similarly, we can estimate the probability of the area between 10^6 and 10^{10} to be $(10 - 6)/10 = 0.4 = 40\%$, even though this range is much larger.

$$P(10^6 \leq T_t \leq 10^{10}) = \int_{10^6}^{10^{10}} \frac{1}{T_t \ln\left(10^{10}/1\right)} dT_t = 0.4 = 40\% \tag{3.15}$$

In summary, when we believe the truth spectrum to exhibit a very large difference between minima and maxima, the log-uniform distribution is a good candidate for the prior.
This will be implemented using the 'interpolated' class in PyMC3.

- Explain prior dist for each bin

Every bin in the data spectrum is assigned a prior distribution as well as a likelihood for which we apply Bayes' theorem.

3.2   Multiplying response with (total) efficiencies

3.3   Likelihood

- Poisson (already mentioned above)

- Modified Poisson to take into account the total amount of counts

3.4   Sampling

- NUTS

There are several sampling methods possible for the problem of unfolding, a common example being Markov Chain Monte Carlo (MCMC) algorithms such as the Metropolis-Hastings algorithm. In the PyFBU-package, a variant of a Hamiltonian Monte Carlo (HMC) Markov Chain Monte Carlo algorithm is the default sampler. HMC aims to be much more efficient than regular MCMC algorithms by avoiding both sensitivity to correlated parameters and random walk tendencies [4]. A drawback to this is a significant sensitivity to step size as well as the number of steps, requiring manual tuning of these parameters. To circumvent this, Hoffman and Gelman created the No U-turn Sampler (NUTS), a variant of HMC which removes having to specify the number of steps. They also implemented an adaptive step size, meaning no manual tuning is necessary for running NUTS. Furthermore, they observed similar to better performance than other fine-tuned HMC algorithms [4]. The NUTS algorithm is implemented in the PyMC3 package [5] and is the default sampling algorithm in PyFBU.

- ADVI?

Another sampling method which will be used in this thesis is called Automatic Differentiation Variational Inference (ADVI) [6].

It is included in PyMC3 as a possible choice for the **initialization** of NUTS. In some cases, the use of this initialization will help when FBU would otherwise crash. Maybe due to the 'jitter' part of 'auto'?

3.5   Posterior inference

Now that the unfolding has been performed, how do we interpret the resulting posterior distribution? While other methods may only return a point value, not necessarily accompanied by the uncertainties, FBU allows us to directly look at the final distribution per bin, and thus observe the result and its corresponding degree of belief. Of course, we are able to quantify these concepts in multiple ways. Here, we take a look at some of the methods of posterior inference.

3.5.1   Point estimates

We will consider three different point estimates, the posterior mean, median and mode:

- Posterior mean: The mean of the posterior distribution which minimizes the mean squared error (MSE) [7].

- Posterior median: The median of the posterior distribution which minimizes the expected absolute error [7].

- Posterior mode: The mode of the posterior distribution, also called the Maximum A Posteriori (MAP), which represents the most likely value for the parameter in question. This does not take into account any skewness of the posterior nor the existence of multiple modes of similar magnitudes.

3.5.2 Credible intervals

The credible interval is the Bayesian version of the frequentist confidence interval. It depends on the posterior and is defined as any interval that encompasses a certain percent of the posterior density. The difference between confidence and credible intervals is subtle, but not negligible. In the case of frequentist inference, the parameter in question, lets say θ, is treated as an unknown, but fixed value. The limits of the confidence interval are treated as random variables. Therefore, a confidence level of 95% means that for 100 repeated experiments, 95 of the confidence intervals will contain θ. Note that this does not mean there is a 0.95 probability of finding θ in every confidence interval. [8][9]

For bayesian inference however, the random-trait is switched, with credible interval limits being fixed, and θ treated as the random variable. The credible interval takes our prior belief into account, while the confidence interval relies only upon the data. A 95% credible interval covers 95% of the posterior and can then be said to contain θ with a probability of 0.95. [8][9]

There are many types of credible intervals, the only requirement being that it covers a certain amount of area of the posterior. Some examples of ways of constructing credible intervals are:

- Using the posterior mean as the interval center.

- Making sure the probability of being outside the interval is equal on all sides (equal tailed).

- Making the interval as narrow as possible, the Highest Posterior Density interval (HPD). This will include the most likely values, as well as the mode of the posterior if it is unimodal.

We will be using the HPD interval which, together with the point estimates mentioned above, will give a solid basis of comparison against the true energies of the γ-ray spectrum.

4   The FBU-package?

5   PyMC3?

- Built-in distributions possible to be used for both likelihoods and priors, such as Uniform, Normal, Poisson etc, as well as Truncated versions of some.

- Custom distribution! 'Interpolated' using x-values and pdf-values.

5.1   Theano?

6   Error statistics? MAE (MSE), R2-score, others?

- Compare unfolded with truth (if known)

- Compare unfolded*response with raw

- Residual plots for both

7   1-dimensional test spectrum

8   2-dimensional test spectrum

- Compare posterior with likelihood*prior

9   28 Si spectrum

- Compare result with Valas, using logscale prior, modified likelihood?

- Background?

Valsdóttir found that, when the background was known, including it in the unfolding significantly improved the results.

10   146 Nd spectrum

- 250 channels

- Background, nanoseconds

- 500 channels

- Background? nanoseconds?

- 453 keV

  Discrepancy at lower energies, unavoidable with this response, we have information to correct this. Maybe on the other states too? Results should match OMPy and thats what we want?

- 1-1.4 MeV

- 6-6.2 MeV

11 Model testing

- Cannot construct 1D likelihood Poisson distribution around data in bin X due to cross-bin dependencies.

- 2D example with 2D likelihood plots.

- Prior must be large enough such as the folded prior includes the observed counts.

- Can compare 1D prior and posterior.

- Can compare 2D likelihood and posterior if we only have 2 bins.

11.1 Synthetic data

For easier and more predictable testing, a synthetic data set is used for the raw spectrum. It consists of 3 excited states, a simplified representation of a physical case.
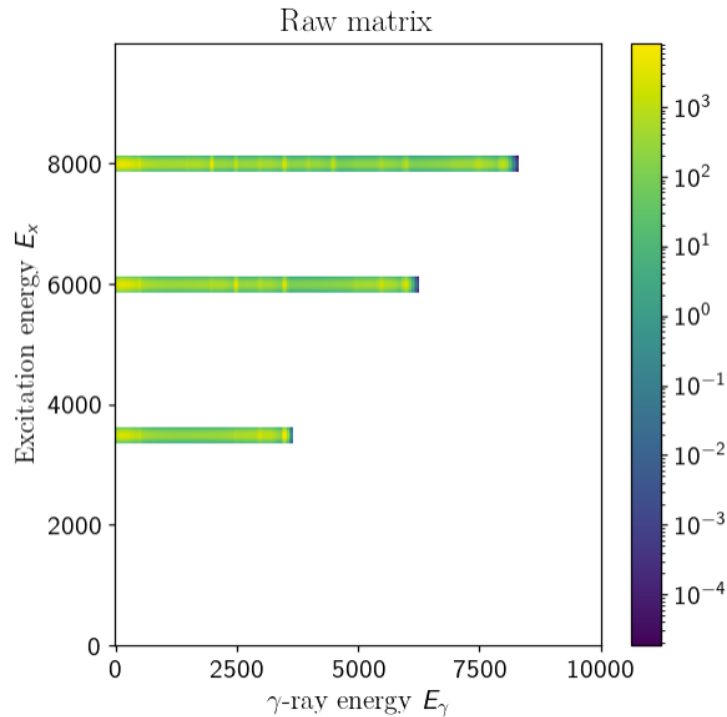
Figure 3.1: Synthetic raw $E_\gamma - E_x$ matrix.

11.2   Bayesian terms

An interesting aspect to look at is the terms in Bayes' theorem after unfolding has been performed. We should be able to reproduce the shape of the resulting posterior samples by multiplying the prior and likelihood. As the likelihood depends on f_r, rather than the truth-values t, direct comparison between this and the other terms is not easy at first glance. To arrive at comparative foundation, we perform the following modifications to the prior and the posterior to achieve f_r-dependence:

$$P(\boldsymbol{T}|\boldsymbol{D})_{f_r} = \boldsymbol{TR}, \tag{3.16}$$

$$P(\boldsymbol{T})_{f_r} = \boldsymbol{SR}, \tag{3.17}$$

where $\boldsymbol{S}$ is a matrix containing random samples between the upper and lower prior limits, with shape corresponding to the output $\boldsymbol{T}$. There is no easy way of extracting the likelihood from the PyFBU package, nor its corresponding f_r-values. However, as we know that it is a Poisson distribution, we are able to define a range of f_r-values determined by the prior and posterior, and use these as input to equation (3.8).

11.3   Response matrices

Here, different trial response matrices are tested for examining the impact on the final result, as well as the convergence of the implemented FBU method. As the response matrix is a vital part of the procedure, significant differences are expected when changes are made. In the end, the experimentally determined response should provide the best unfolded spectrum. It is however interesting to see the effect on the resulting Bayesian terms.

11.3.1   Normalized response from OCL

The following figures are results from running FBU using the response matrix from OCL, with normalization performed on each row.
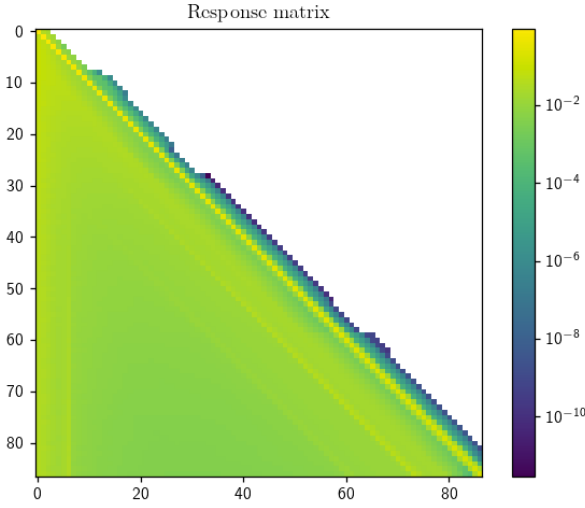
(a) Response matrix
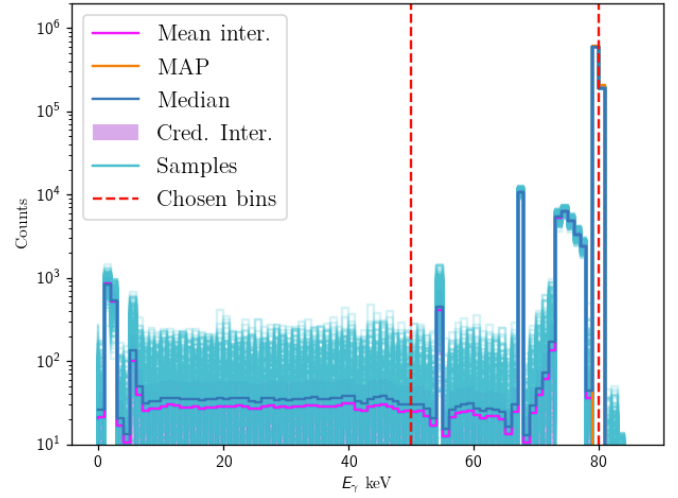
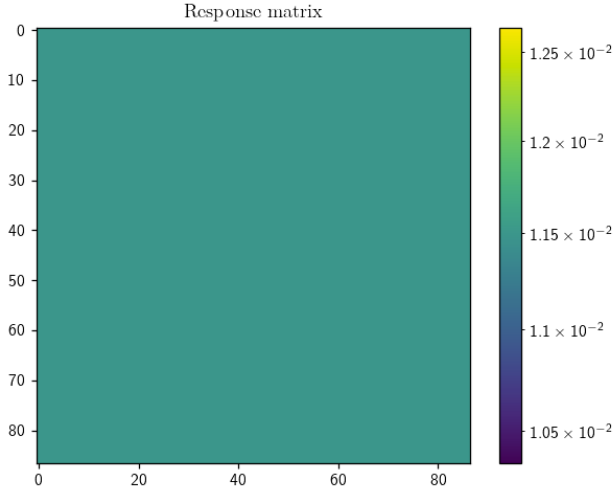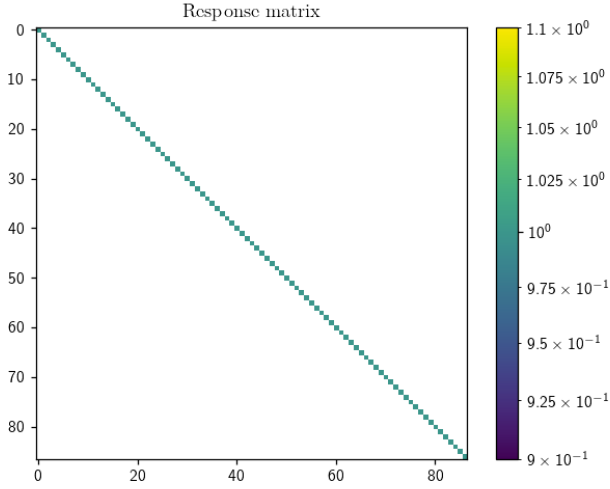(b) Result of unfolding. Dashed lines show the bins chosen for the analysis below.
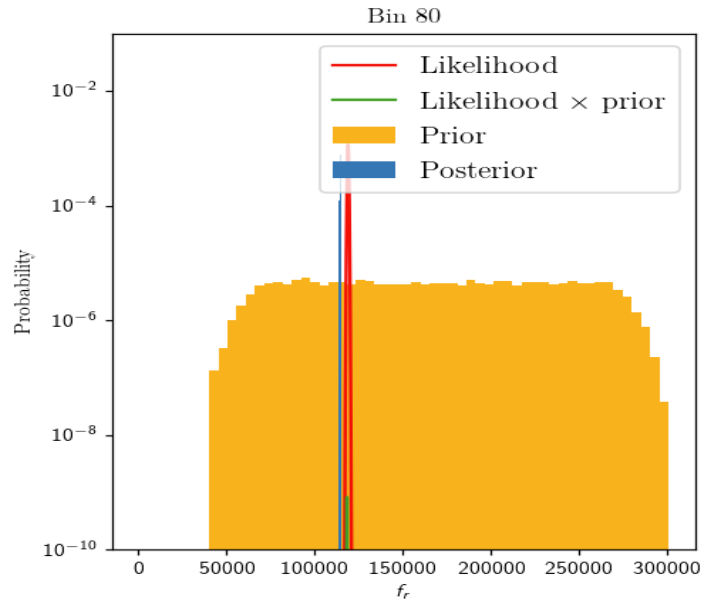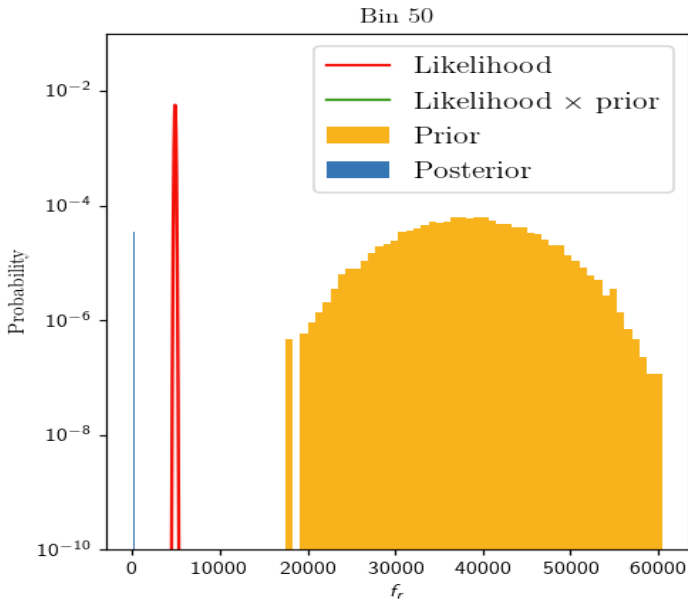
(c) The components of Bayes' theorem after unfolding, for the bins chosen above. Included is also a $L(\boldsymbol{D}|\boldsymbol{T}) \times P(\boldsymbol{T})$-function, which should have the same shape and position as the posterior, only differing by a normalisation constant.

11.3.2 Response from OCL with normalized columns

The following figures are results from running FBU using the response matrix from OCL, with normalization performed on each column instead.
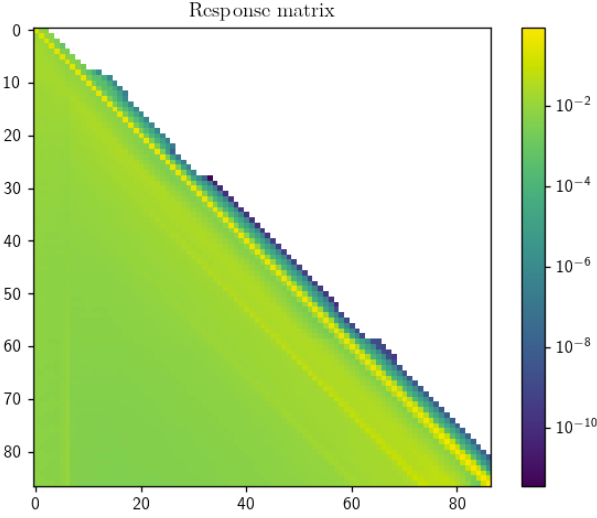
(a) Response matrix
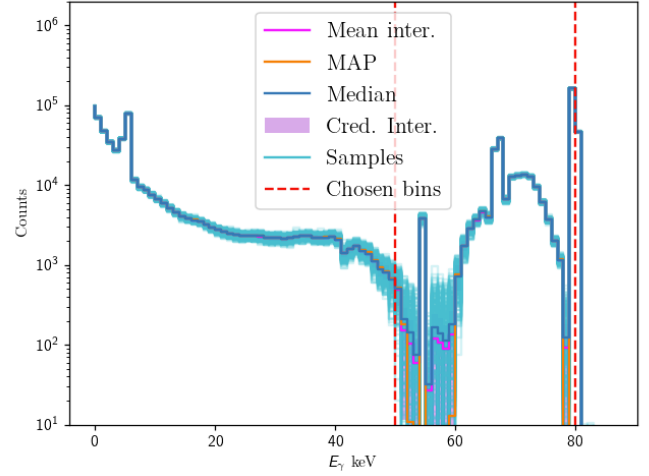
(b) Result of unfolding. Dashed lines show the bins chosen for the analysis below.
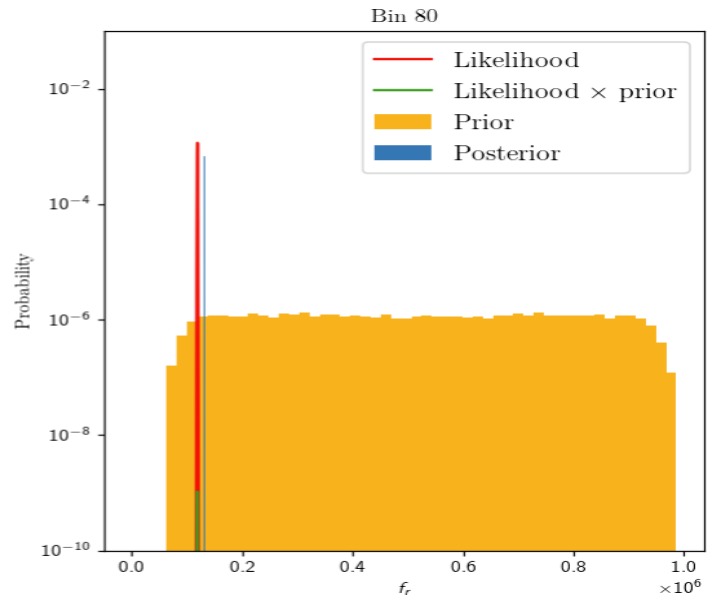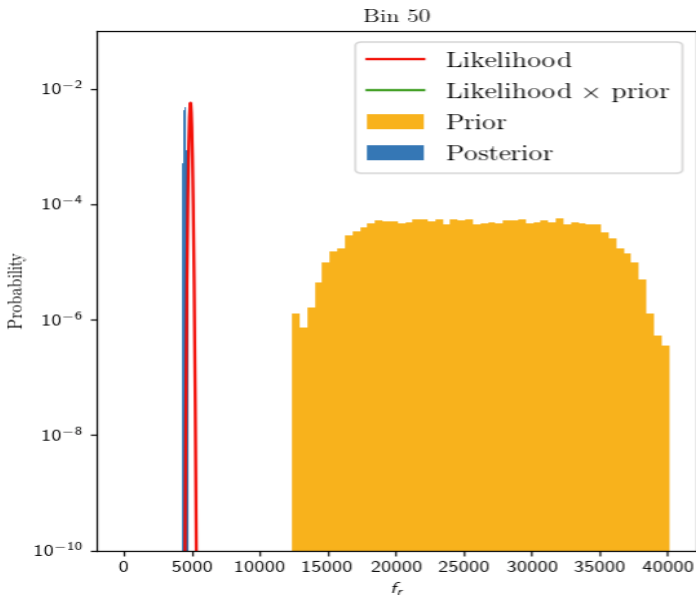
(c) The components of Bayes' theorem after unfolding, for the bins chosen above. Included is also a $L(\boldsymbol{D}|\boldsymbol{T}) \times P(\boldsymbol{T})$-function, which should have the same shape and position as the posterior, only differing by a normalisation constant.

11.3.3 Symmetrized response from OCL

The following is produced using a symmetrized version of the OCL response matrix, that is

$$R = \frac{R_{OCL} + R_{OCL}^T}{2} \tag{3.18}$$
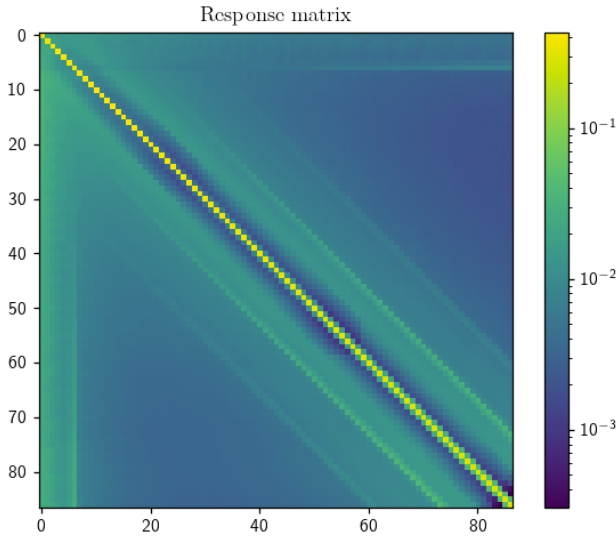
(a) Response matrix
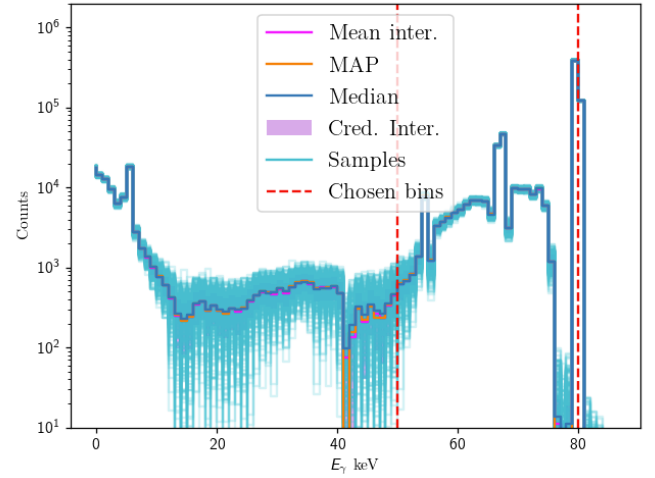
(b) Result of unfolding. Dashed lines show the bins chosen for the analysis below.

(c) The components of Bayes' theorem after unfolding, for the bins chosen above. Included is also a $L(\boldsymbol{D}|\boldsymbol{T}) \times P(\boldsymbol{T})$-function, which should have the same shape and position as the posterior, only differing by a normalisation constant.

27
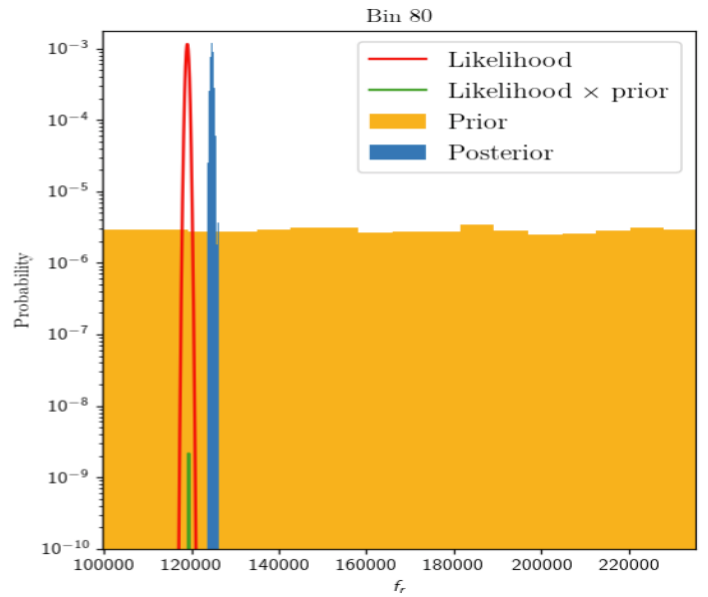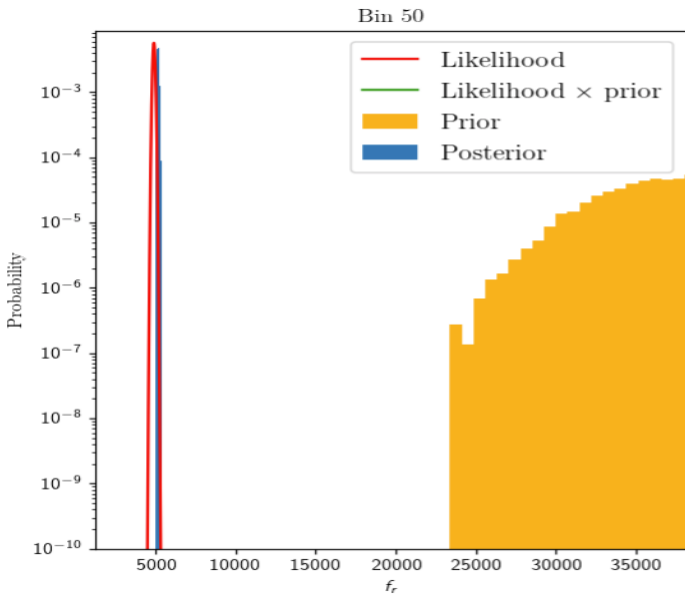
11.3.4   Response as a normalized matrix of ones

Now, the response is a normalized matrix of ones:

$$R^{N \times N} = \begin{bmatrix} 1/N & 1/N & \cdots \\ 1/N & \ddots & \\ \vdots & & \ddots \end{bmatrix} \tag{3.19}$$

(a) Response matrix
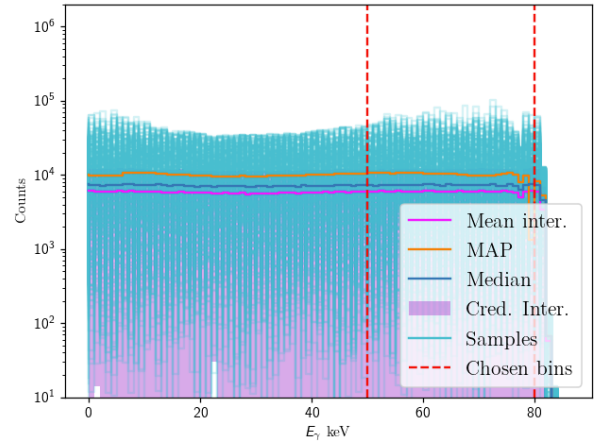
(b) Result of unfolding. Dashed lines show the bins chosen for the analysis below.
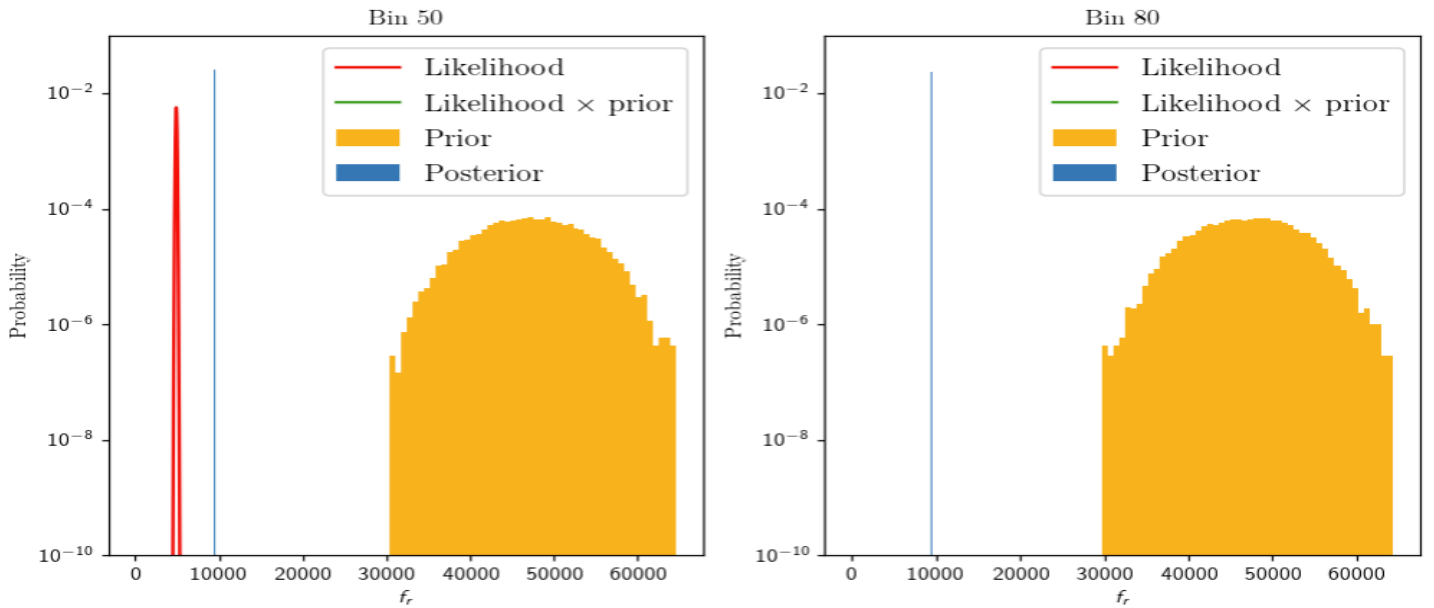
(c) The components of Bayes' theorem after unfolding, for the bins chosen above. Included is also a $L(\mathbf{D}|\mathbf{T}) \times P(\mathbf{T})$-function, which should have the same shape and position as the posterior, only differing by a normalisation constant.

11.3.5 Response as the identity matrix

The results below are produced using an identity response matrix.

$$R = \mathbb{I} \tag{3.20}$$

(a) Response matrix
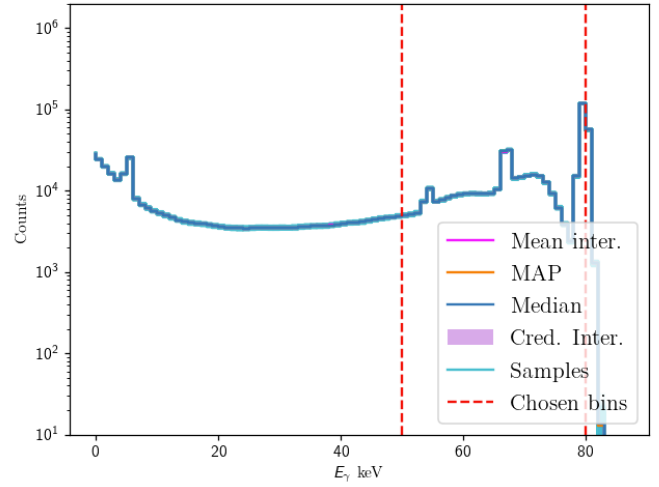
(b) Result of unfolding. Dashed lines show the bins chosen for the analysis below.
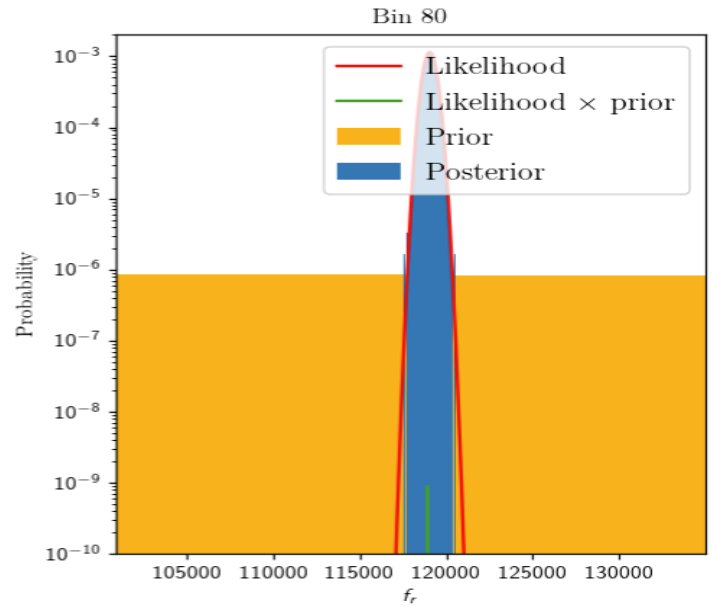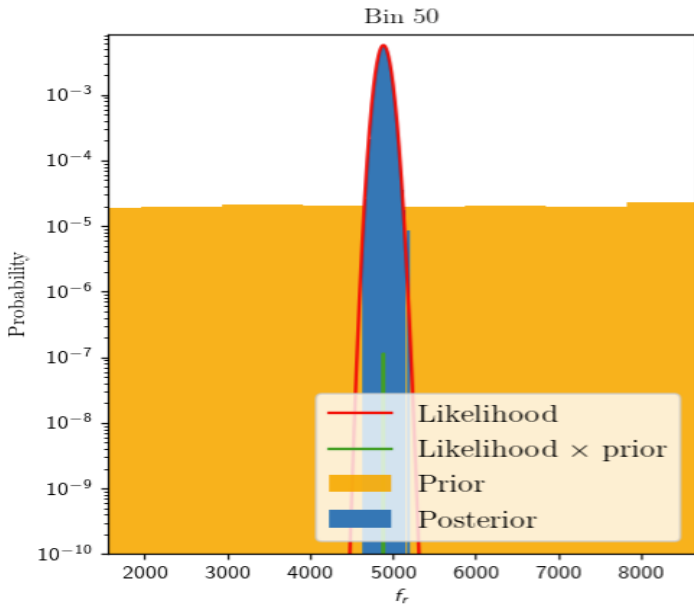
(c) The components of Bayes' theorem after unfolding, for the bins chosen above. Included is also a $L(\boldsymbol{D}|\boldsymbol{T}) \times P(\boldsymbol{T})$-function, which should have the same shape and position as the posterior, only differing by a normalisation constant.

Reproduction

1 Reproduction of results

Under follows a reproduction of the results achieved by Valsdóttir [10]. This procedure allows us to validate the results, as well as providing a benchmark for the new and hopefully improved evolution of these algorithms.

The figures presented as reproductions are outputs of the publicly available Jupyter notebooks on Valsdóttirs GitHub repository [11]. Some discrepancies are seen, pointing to the possibility that the results in Valsdóttirs thesis may stem from newer, locally stored versions of the files that have been made accessible on the repository. This may also be the reason to why the code for some results in the thesis is not found on the repository at all. Therefore, the below sections contain only the results for which corresponding output was found to be produced in the mentioned Jupyter notebooks.

(a) Valsdóttir

(b) Reproduction

Figure 4.1

(a) Valsdóttir

(b) Reproduction
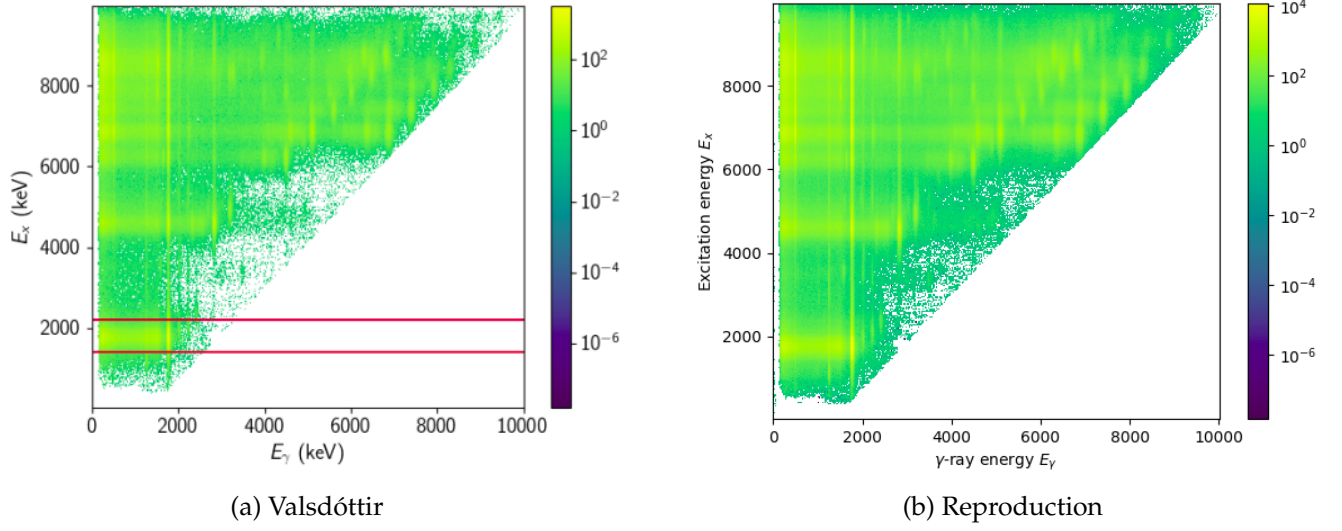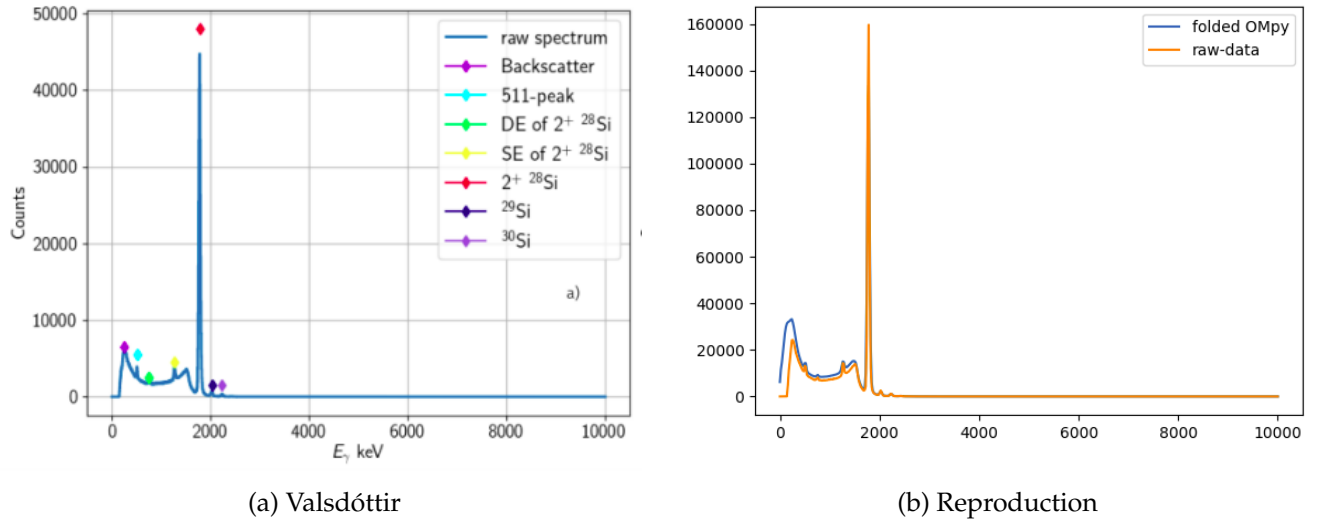
Figure 4.2

1.1 Fully Bayesian Unfolding Spectrum From the First Excited state of ^{28}Si

0

(a) Valsdóttir   (b) Reproduction

Figure 4.3

(a) Valsdóttir   (b) Reproduction

Figure 4.5

(a) Valsdóttir

(b) Reproduction

Figure 4.6

(a) Valsdóttir

(b) Reproduction
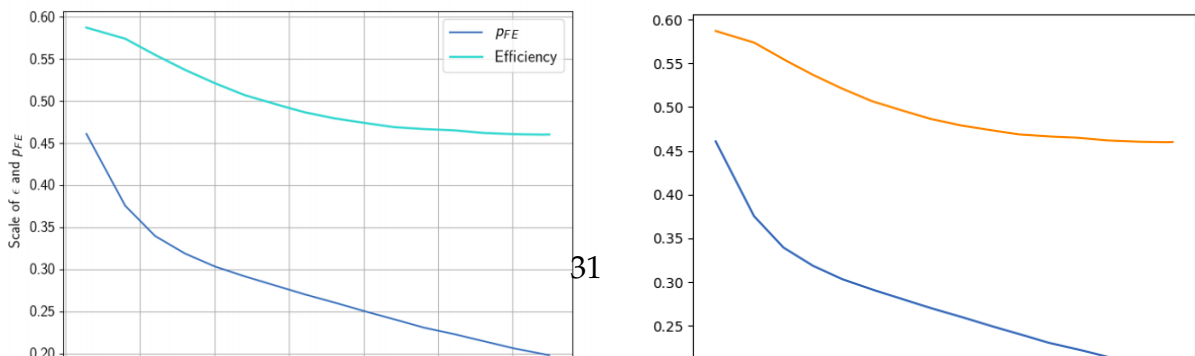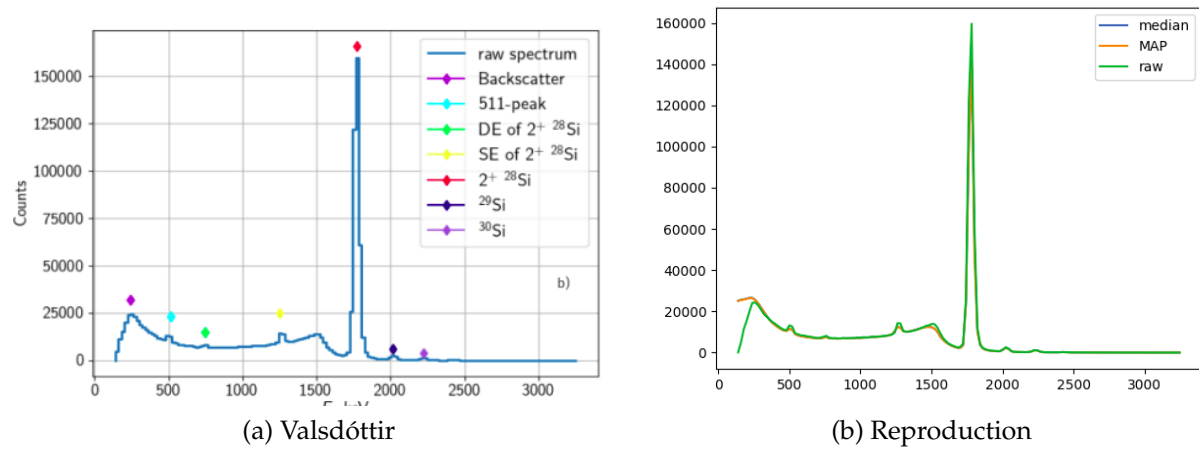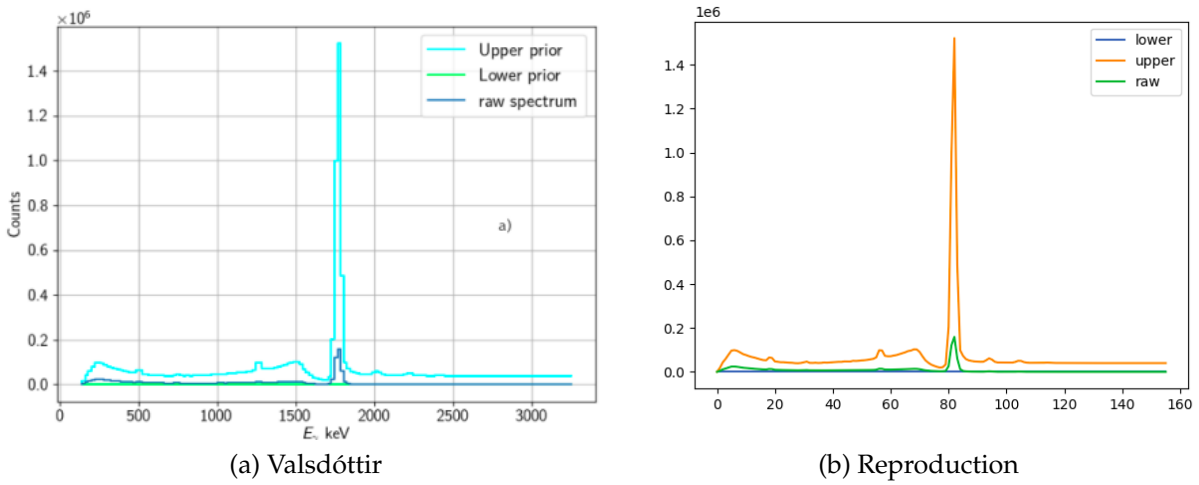
Figure 4.7

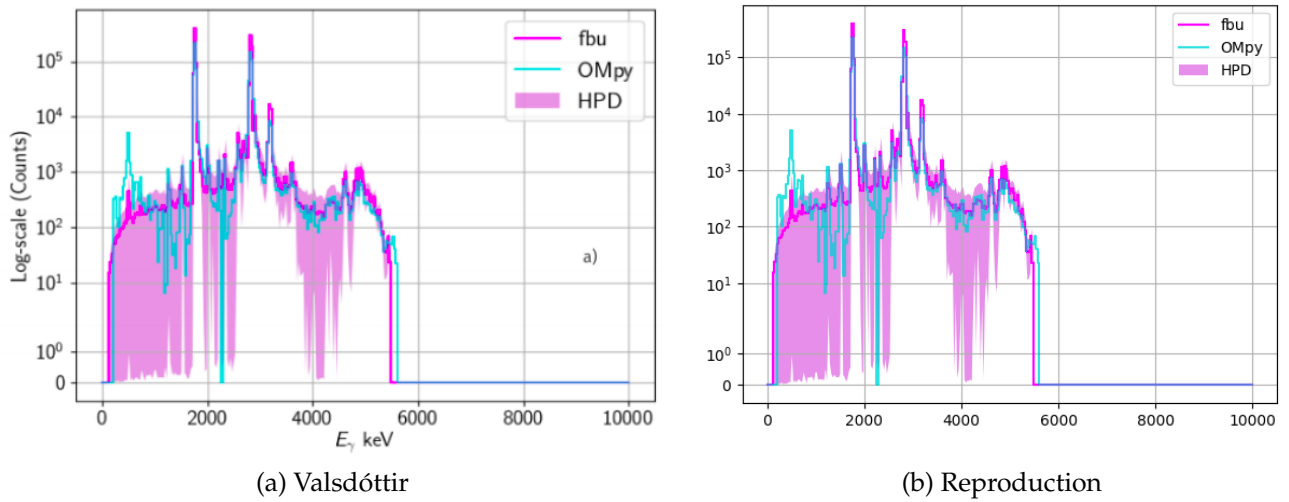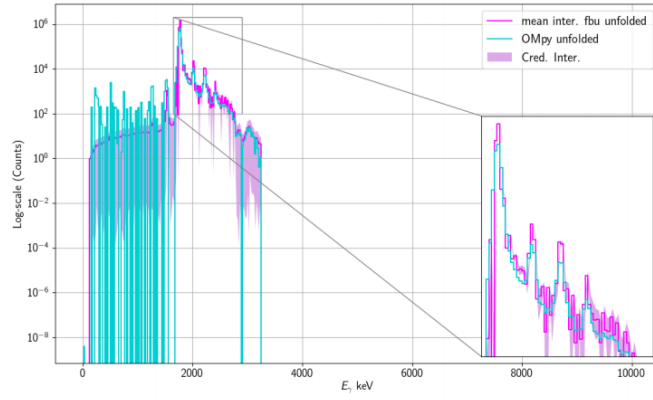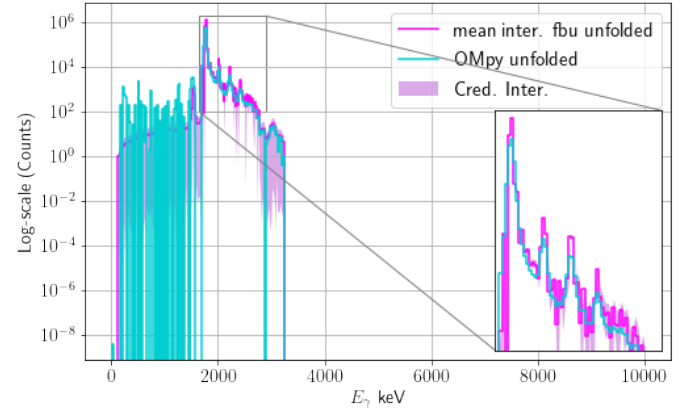## 1.2   Fully Bayesian Unfolding on first excited state of ^{28}Si including background

(a) Valsdóttir   (b) Reproduction

Figure 4.8

(a) Valsdóttir   (b) Reproduction
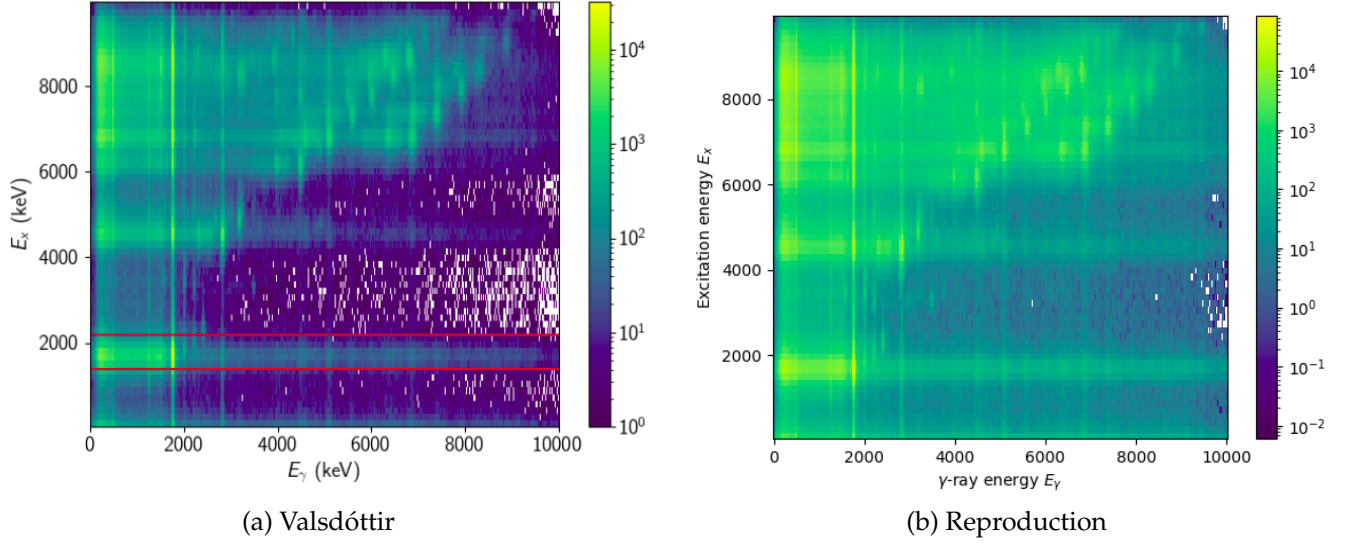
Figure 4.9
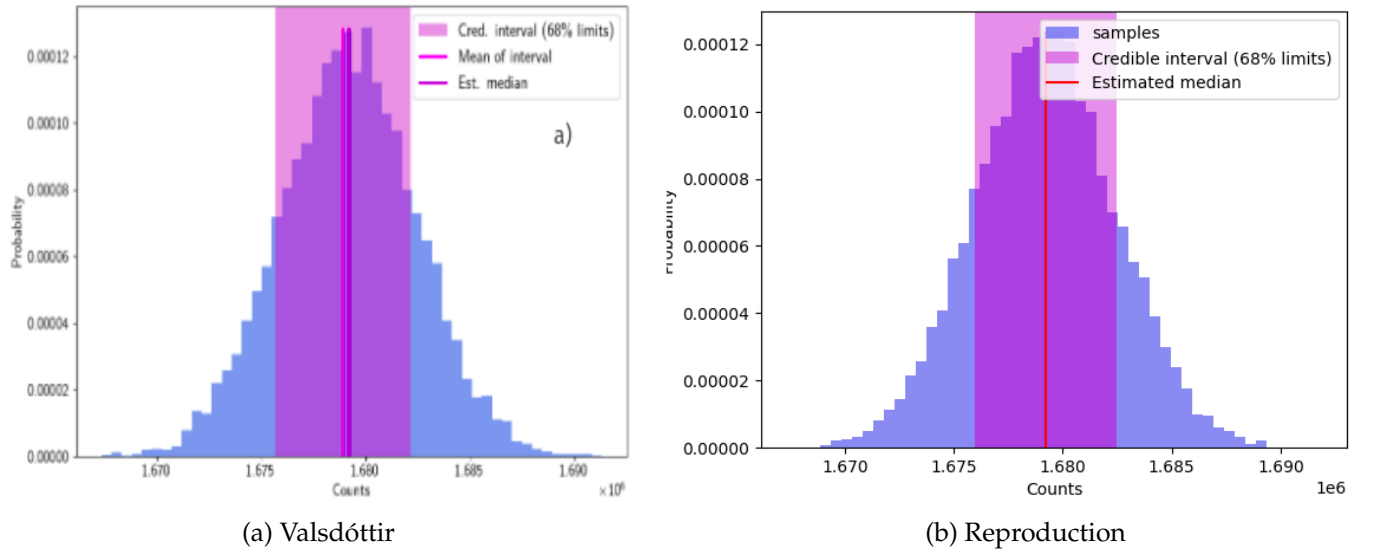
(a) Valsdóttir
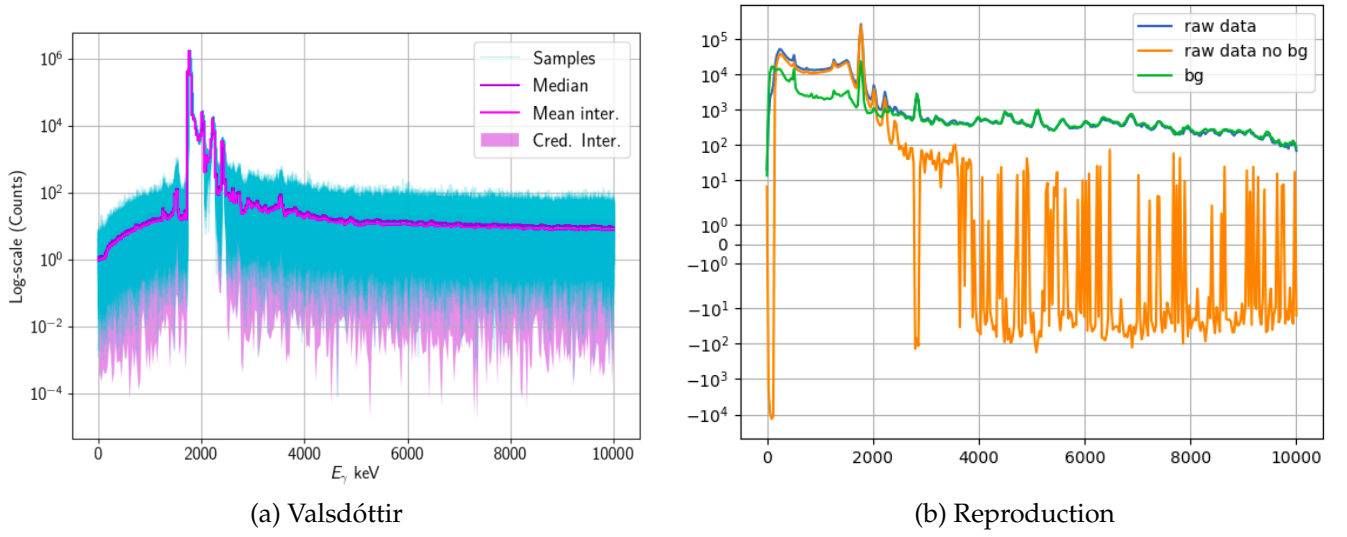
(b) Reproduction

Figure 4.10

1.3   Fully Bayesian Unfolding for all Excited States

(a) Valsdóttir

(b) Reproduction

Figure 4.11

(a) Valsdóttir

(b) Reproduction
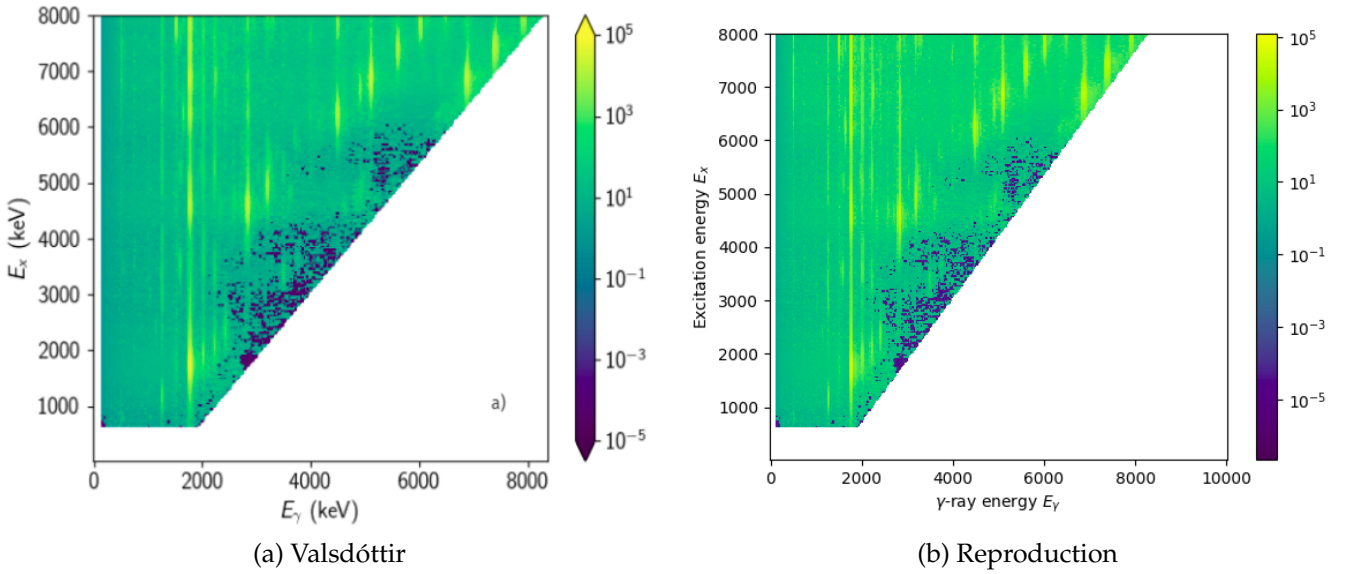
Figure 4.12
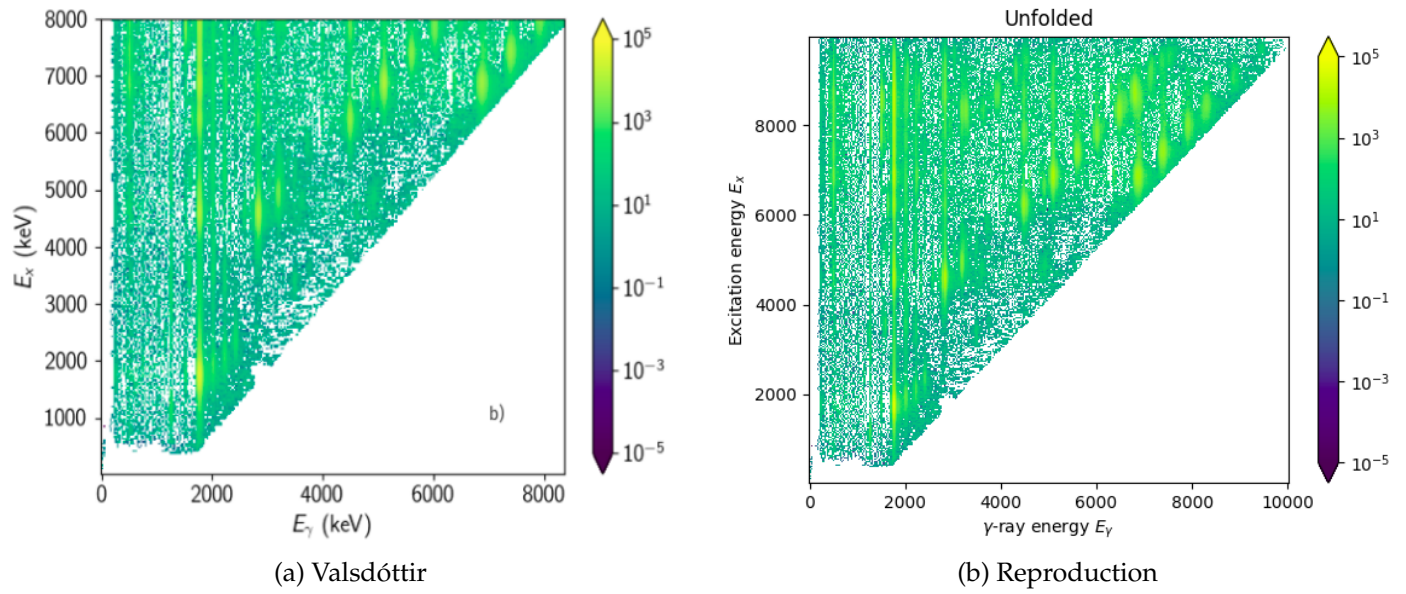
(a) Valsdóttir

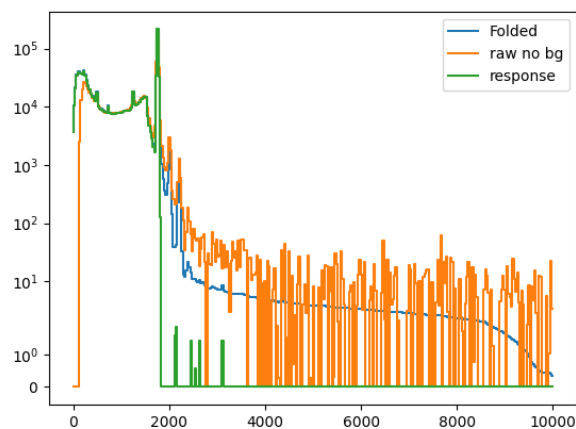(b) Reproduction

Figure 4.13

(a) Valsdóttir   (b) Reproduction
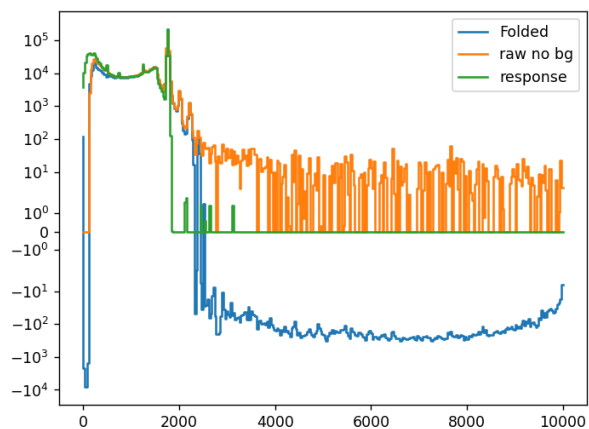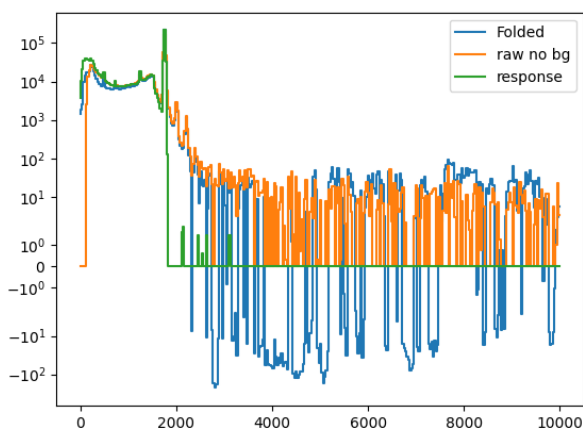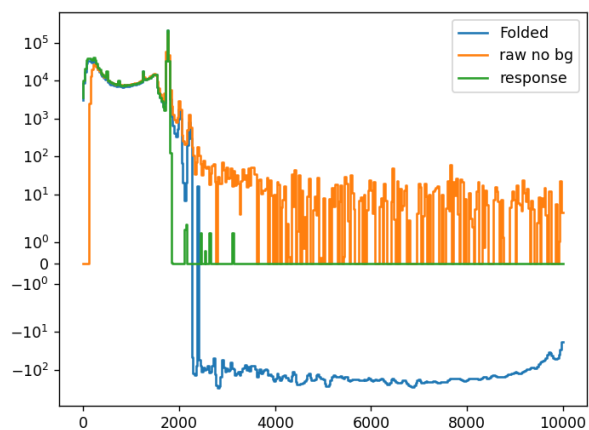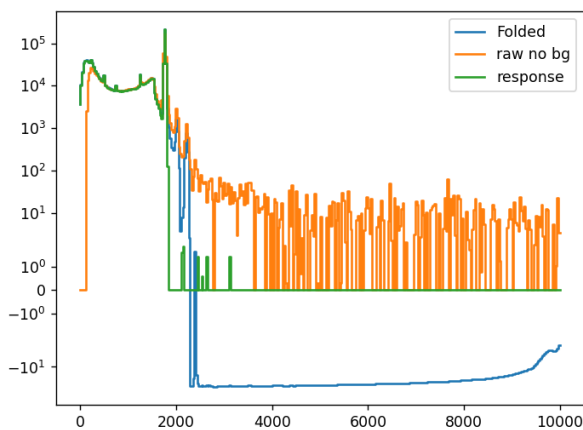
Figure 4.14

(a) Prior lower bound = raw/10

(b) Prior lower bound = raw/50
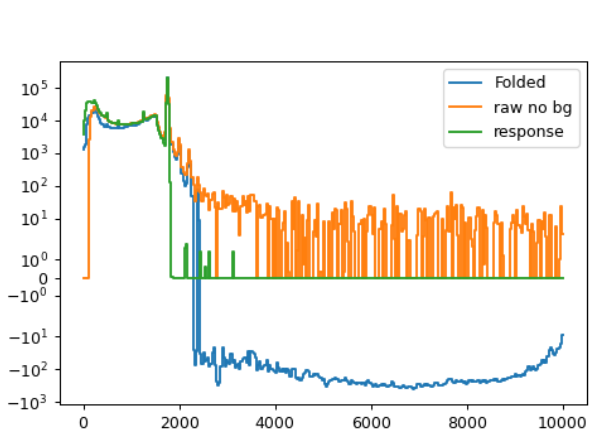

(c) Prior lower bound = raw/100

(d) Prior lower bound = raw/1000

(e) Prior lower bound = raw−100

(f) Prior lower bound = raw−20000

Figure 4.15

Part III

Implementation

Modification of PyFBU and PyMC3

- Different priors

- Modificated likelihood

Part IV

Results & Discussion

CHAPTER 6

Synthetic spectra

- 1dim test
- 2dim test (likelihood*prior)

Experimental spectra

1   146Nd

- Raw? 2 dim and projected

- Response?

- FBU unfolded

- FBU refolded

- Residuals

- MSE, R2

- OMPY unfolded

- OMPY refolded

- residuals?

- MSE, R2

Part V

Conclusion

Bibliography

1. D. S. Sivia, J. S. *Data analysis: A Bayesian Tutorial* 2nd ed. ISBN: 0–19–856831–2 (Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 2006).

2. Cowan, G. *A survey of unfolding methods for particle physics* in *Prepared for Conference on Advanced Statistical Techniques in Particle Physics, Durham, England* (2002), 18–22.

3. Guttormsen, M., Tveter, T., Bergholt, L., Ingebretsen, F. & Rekstad, J. The unfolding of continuum γ-ray spectra. *Nuclear instruments & methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment* **374,** 371–376. ISSN: 0168-9002 (1996).

4. Hoffman, M. D. & Gelman, A. *The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo* 2011. arXiv: 1111.4246 [stat.CO].

5. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science* **2.** Publisher: PeerJ Inc., e55. ISSN: 2376-5992. https://peerj.com/articles/cs-55 (2021) (Apr. 6, 2016).

6. Kucukelbir, A., Ranganath, R., Gelman, A. & Blei, D. M. *Automatic Variational Inference in Stan* 2015. arXiv: 1506.03431 [stat.ML].

7. Jaynes, E. T. *Probability Theory: The Logic of Science* (ed Bretthorst, G. L.) (Cambridge University Press, 2003).

8. Lee, P. M. *Bayesian statistics : an introduction* eng. Chichester, West Sussex ; 2012.

9. Edwards, W., Lindman, H. & Savage, L. J. Bayesian statistical inference for psychological research. *Psychological Review* **70.** Publisher: US: American Psychological Association, 193. ISSN: 1939-1471. http://psycnet.apa.org/fulltext/1964-00040-001.pdf (1964).

10. Valsdóttir, V. M. *Exploring Fully Bayesian Unfolding for γ-ray Spectra* MA thesis (University of Oslo, Oslo, 2020).

11. Valsdóttir, V. M. *Exploring Fully Bayesian Unfolding for γ-ray Spectra - GitHub repository* https://github.com/valamaria89/Exploring-Fully-Bayesian-Unfolding-for-gamma-ray-Spectra (2020).