Extraction of signal over background from spectral data using a Markov Chain Monte Carlo with a Gibbs sampler

Advanced Statistics for Physics Analysis

Prof. Alberto Garfagnini

Academic year
2019/2020

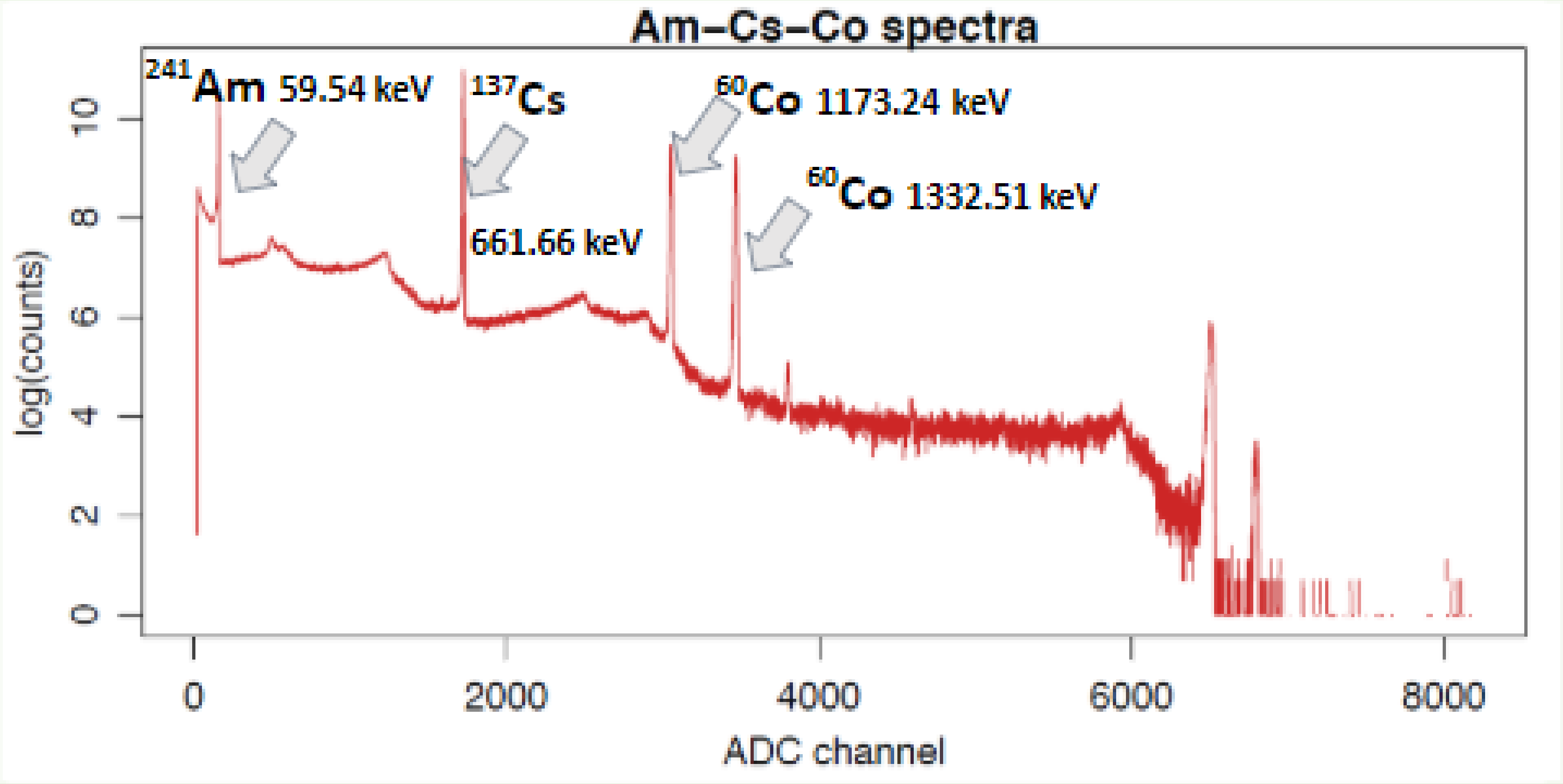
Università degli Studi di Padova

Content

1. **Peak by peak** analysis using JAGS
2. Finding the **best estimates** of parameters A, B, w, x0 and **95% credible interval**.
3. **Correlation among parameters** both Qualitative (2D posterior distributions) and Quantitative (correlation matrix)
4. Inference → **Histograms plot**(visual understanding)
5. Check for convergence by re-running with **different initial parameters**
6. **ACF**: Effect of **thinning** and **burn-in**
7. **Calibration**: Using best estimates of peak center and energy emission (given)
8. Estimating **number of events** under each peak.

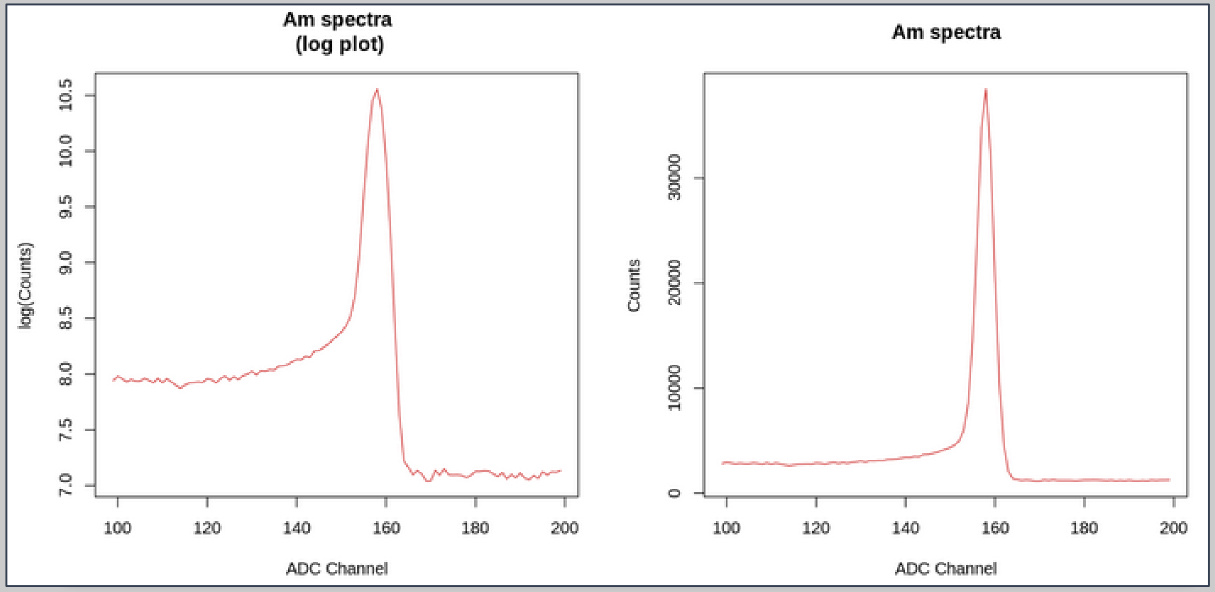
Our goal

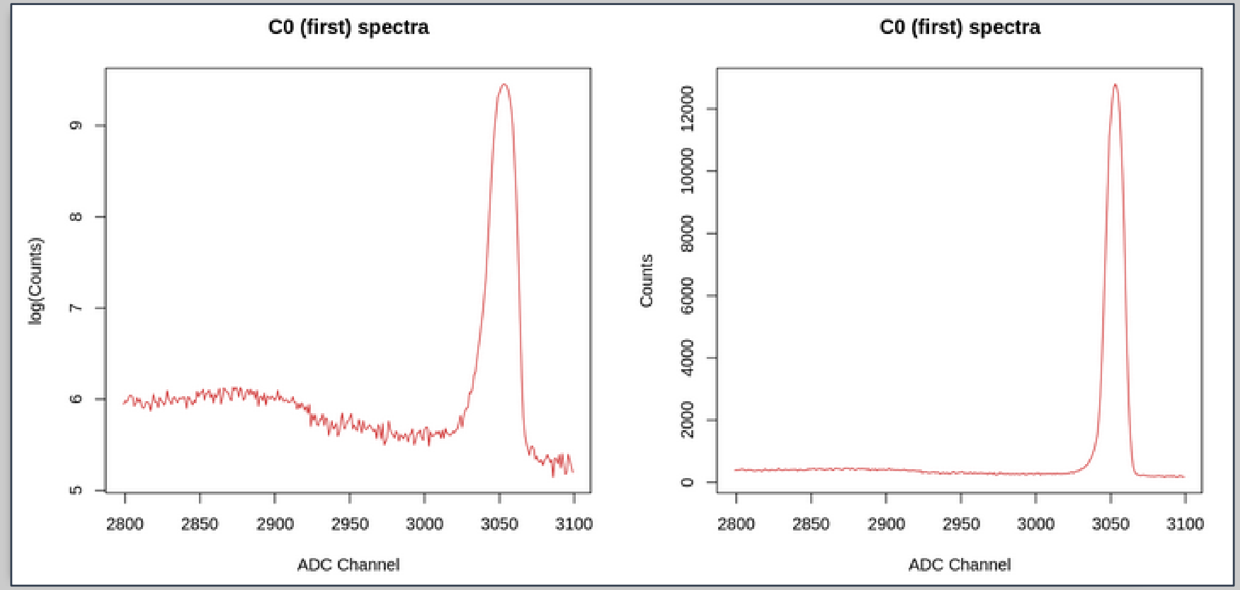
Determine the **number of events** under the source γ source taking into account the underlaying background

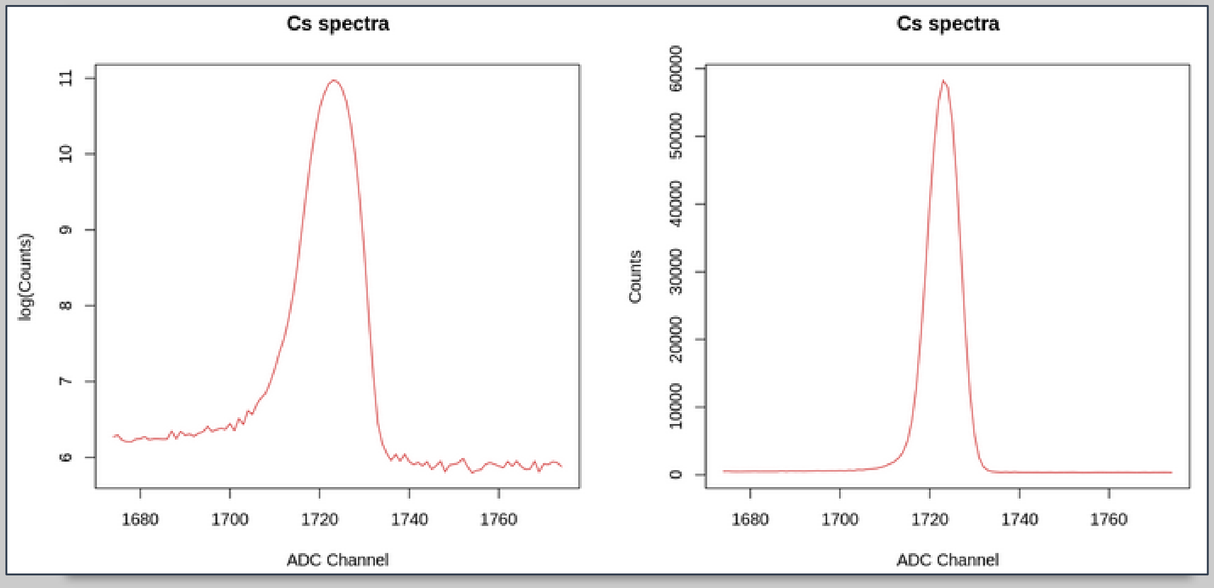


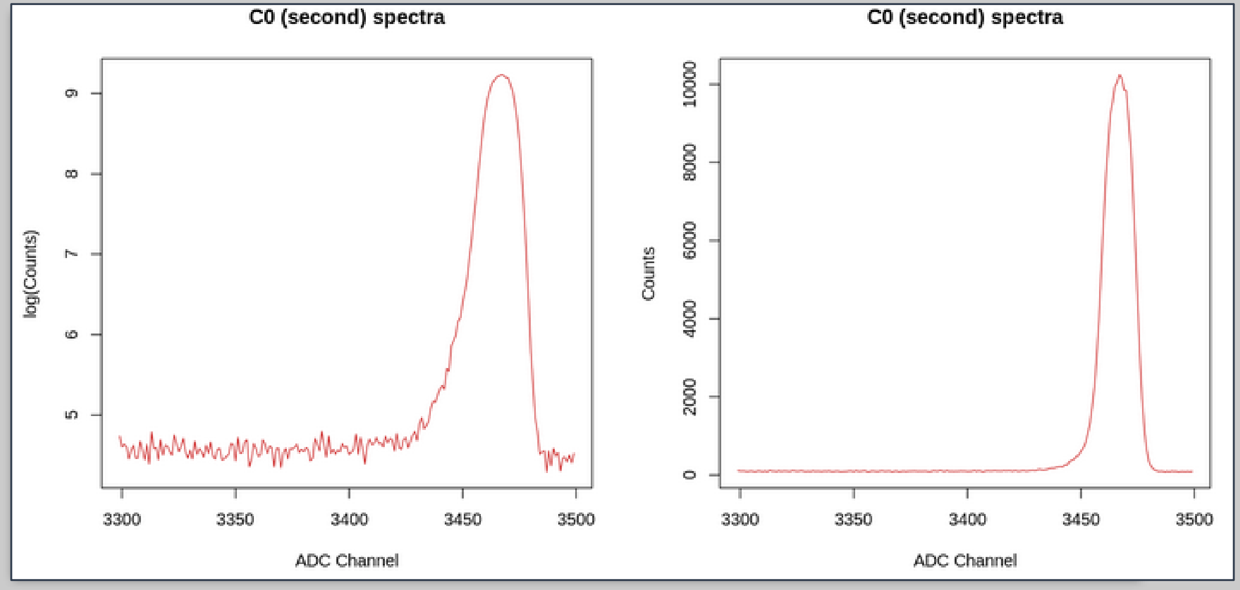
Data

- We begin by analysing peak by peak.
- The first 4 peaks are to be analysed independently.


Am spectra
(log plot)
log(Counts)
7.0 7.5 8.0 8.5 9.0 9.5 10.0 10.5
100 120 140 160 180 200
ADC Channel
Am spectra
Counts
0 10000 20000 30000
100 120 140 160 180 200
ADC Channel


C0 (first) spectra
log(Counts)
5 6 7 8 9
2800 2850 2900 2950 3000 3050 3100
ADC Channel
C0 (first) spectra
Counts
0 2000 4000 6000 8000 10000 12000
2800 2850 2900 2950 3000 3050 3100
ADC Channel

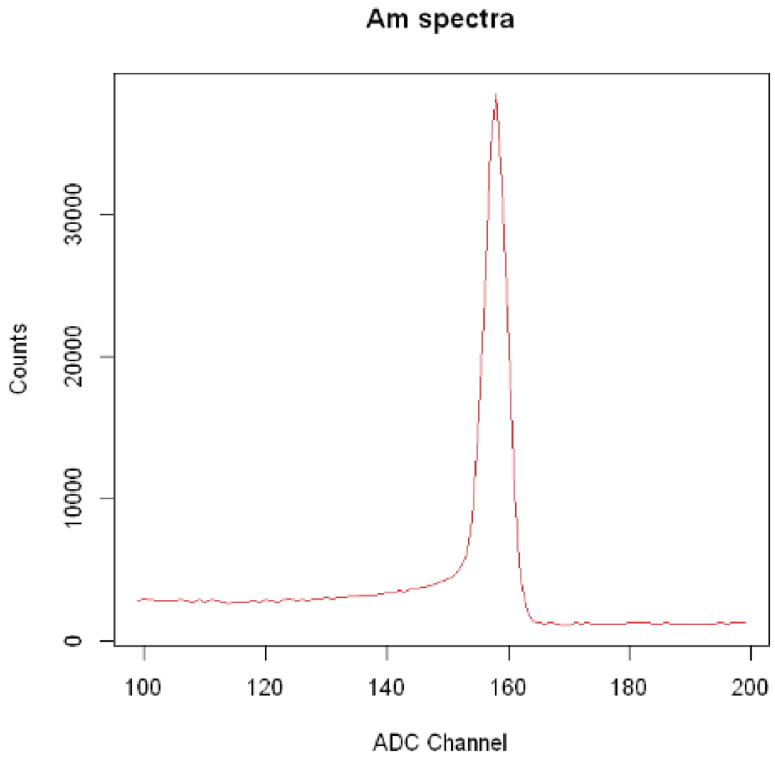

Cs spectra
log(Counts)
6 7 8 9 10 11
1680 1700 1720 1740 1760
ADC Channel
Cs spectra
Counts
0 10000 20000 30000 40000 50000 60000
1680 1700 1720 1740 1760
ADC Channel


C0 (second) spectra
log(Counts)
5 6 7 8 9
3300 3350 3400 3450 3500
ADC Channel
C0 (second) spectra
Counts
0 2000 4000 6000 8000 10000
3300 3350 3400 3450 3500
ADC Channel

Jags

Bugs model for Am-241

```
model{

    for (i in 1:length(x)){

        S[i] <- (A*exp((-(x[i]-x0)^2)/(2*w^2))+B)

        y[i] ~ dpois(S[i])     # likelihood of the
    data}

    A ~ dunif(0,60000)
    B ~ dunif(0,5000)
    w ~ dunif(0,35)
    x0 ~ dunif(140,180)
     }
```



- We assume **uniform prior** for parameters and arguments are considered by observing the graphs.
- The signal was modelled as a **Gaussian** whereas the number of observed photons follows **Poisson distribution**.

Unkowns

A	Signal Amplitude as counts
B	Background Amplitude as counts
w	Width of the signal peak
x0	Center of the signal peak

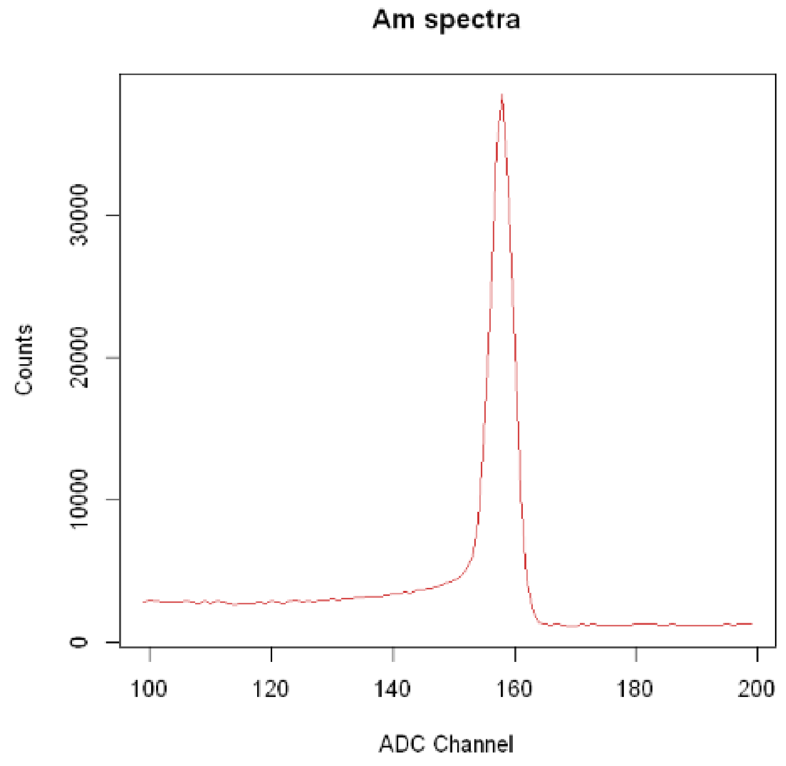
Jags

```r
# Initial parameter values
init <- NULL
init$A <- 30000
init$B <- 5000
init$w <- 8
init$x0 <- 150
```

```r
#running jags

# set the seed
set.seed(12345)

# Create the model and pass the parameters
jm <- jags.model("./Am241.bug", Am.data , init)

# Update the Markov chain (Burn -in)
update (jm , 4000)

chain <- coda.samples(jm , variable.names=c("A","B","x0","w"), n.iter = 2e5, thin = 100)
```

```
Compiling model graph
   Resolving undeclared variables
   Allocating nodes
Graph information:
   Observed stochastic nodes: 101
   Unobserved stochastic nodes: 4
   Total graph size: 922
```

```r
# plotting the chain, posterior distributions of the parameters
options(repr.plot.width=12, repr.plot.height=15)
plot(chain, col='navy')
```



Results from the MCMC

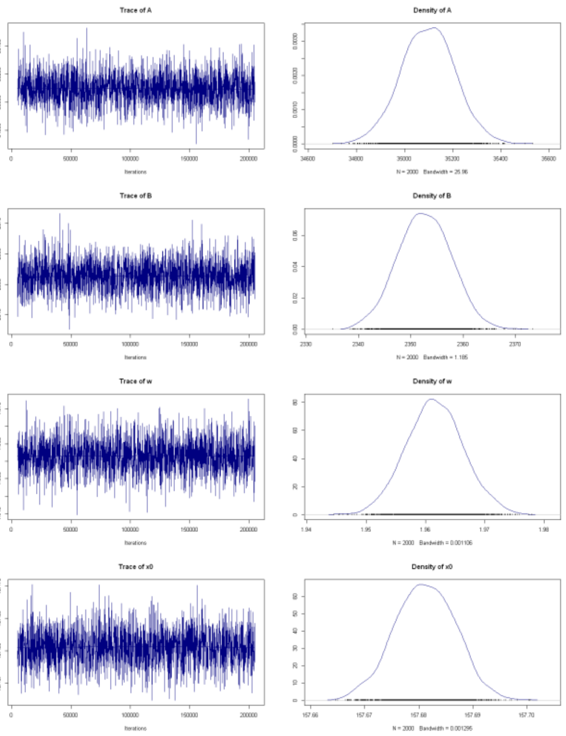
A

B

W

x0

Am^{241}

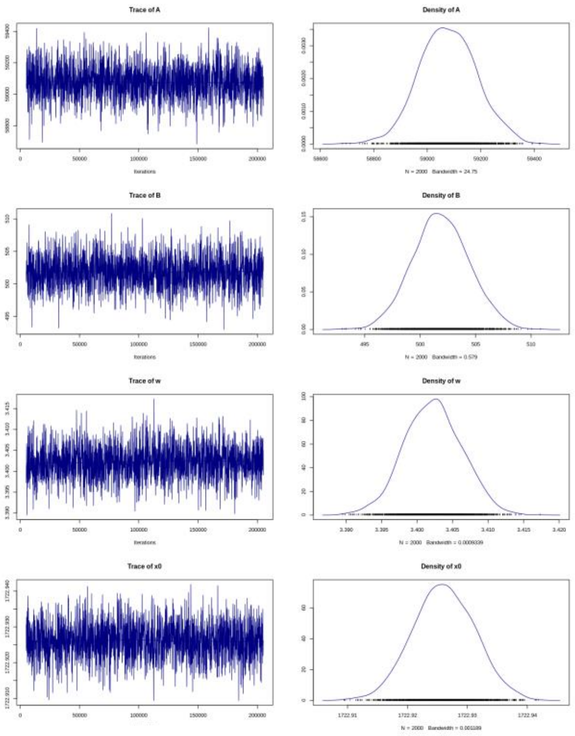


	Mean	SD
A	35097.234	1.120e+02
B	2352.666	5.111e+00
w	1.962	4.847e-03
x0	157.681	5.588e-03

95% CI

```
param: lower_limit - upper_limit
A : 34864.94 - 35326.59
B : 2342.511 - 2362.925
w : 1.951939 - 1.971753
x0 : 157.6698 - 157.6925
```

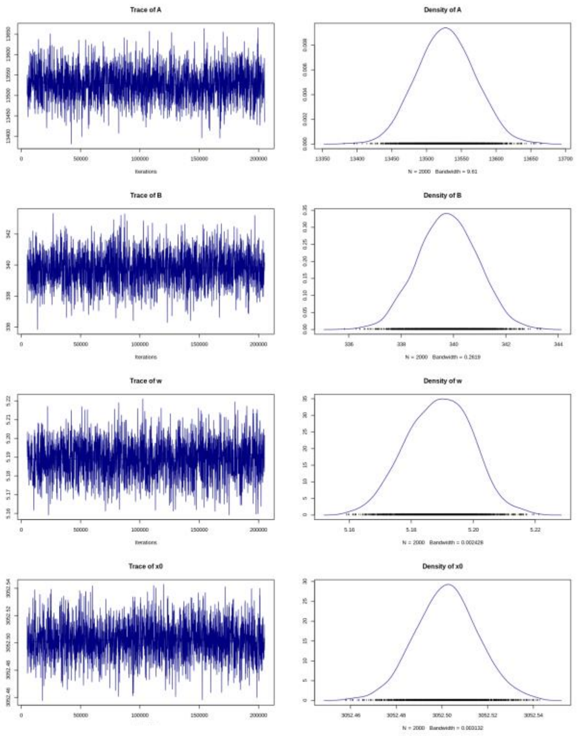
Cs^{137}



Mean	SD
59077.628	1.068e+02
501.870	2.534e+00
3.402	4.029e-03
1722.926	5.130e-03

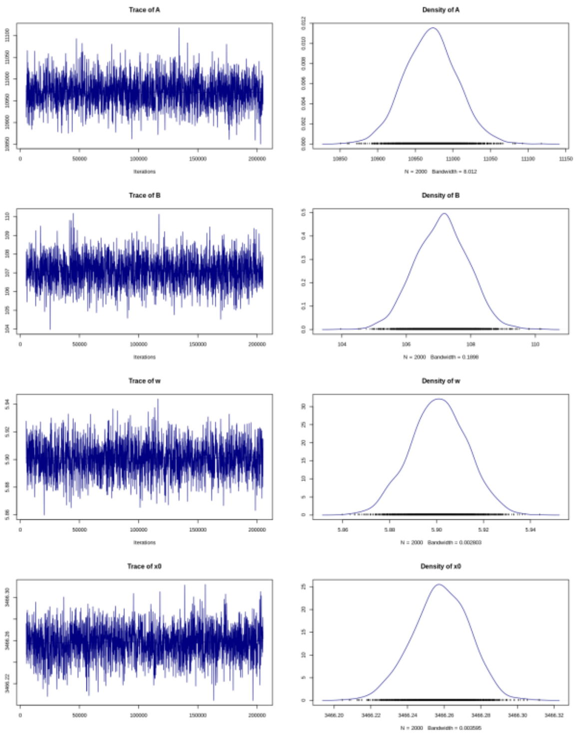
```
param: lower_limit - upper_limit
A : 58873.25 - 59292.86
B : 496.947 - 506.9042
w : 3.394245 - 3.410095
x0 : 1722.916 - 1722.936
```

Co^{60}



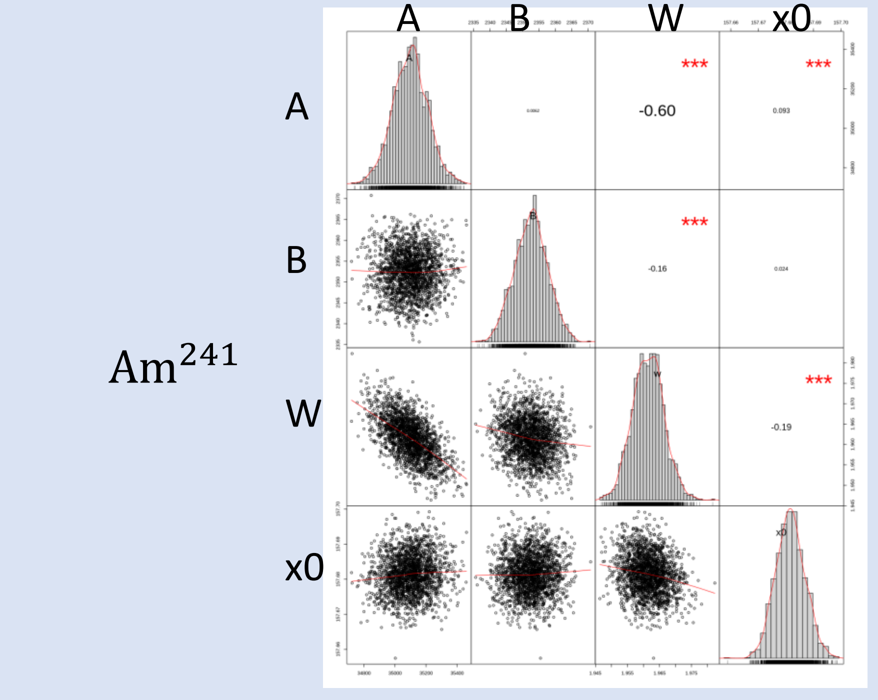
Mean	SD
13527.983	41.45953
339.770	1.13003
5.189	0.01048
3052.502	0.01379

```
param: lower_limit - upper_limit
A : 13448.28 - 13608.46
B : 337.6332 - 341.9345
w : 5.168487 - 5.208965
x0 : 3052.475 - 3052.53
```



Mean	SD
10970.682	34.95848
107.099	0.81889
5.901	0.01209
3466.258	0.01551

```
param: lower_limit - upper_limit
A : 10903.32 - 11040.81
B : 105.5348 - 108.6569
w : 5.877163 - 5.92465
x0 : 3466.227 - 3466.287
```

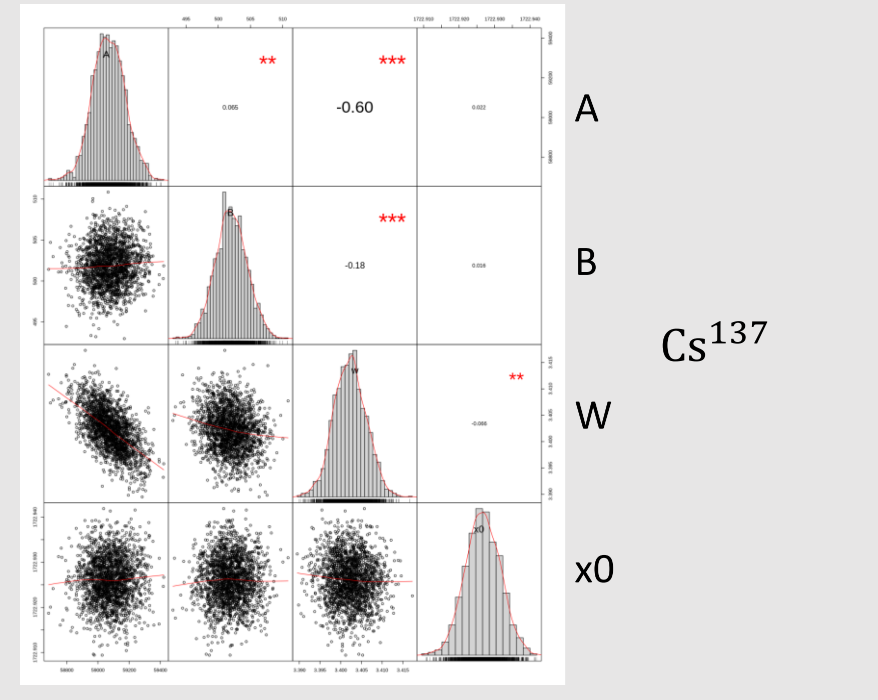


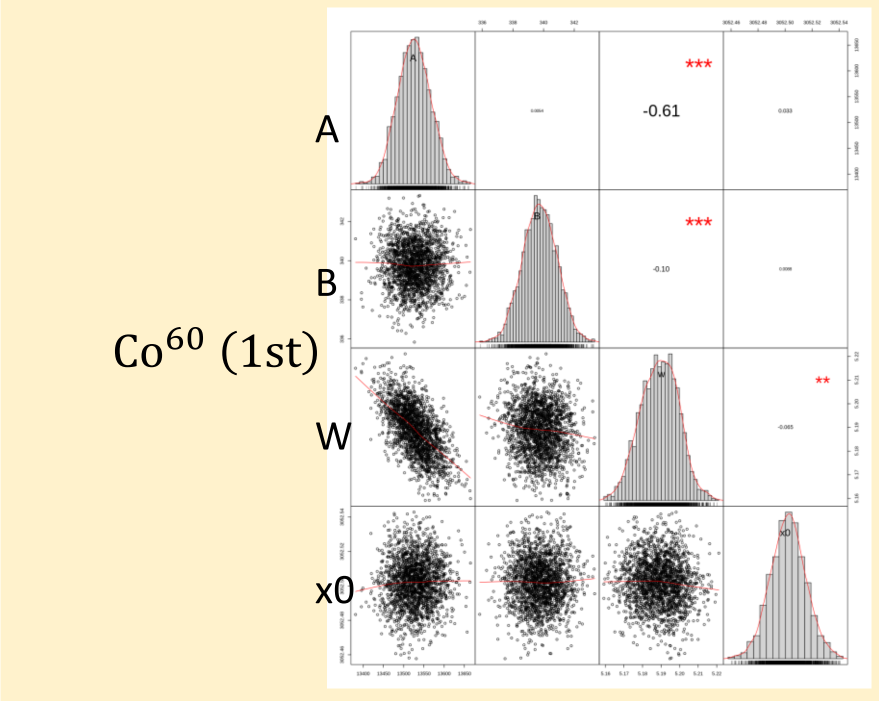
Correlation among parameters

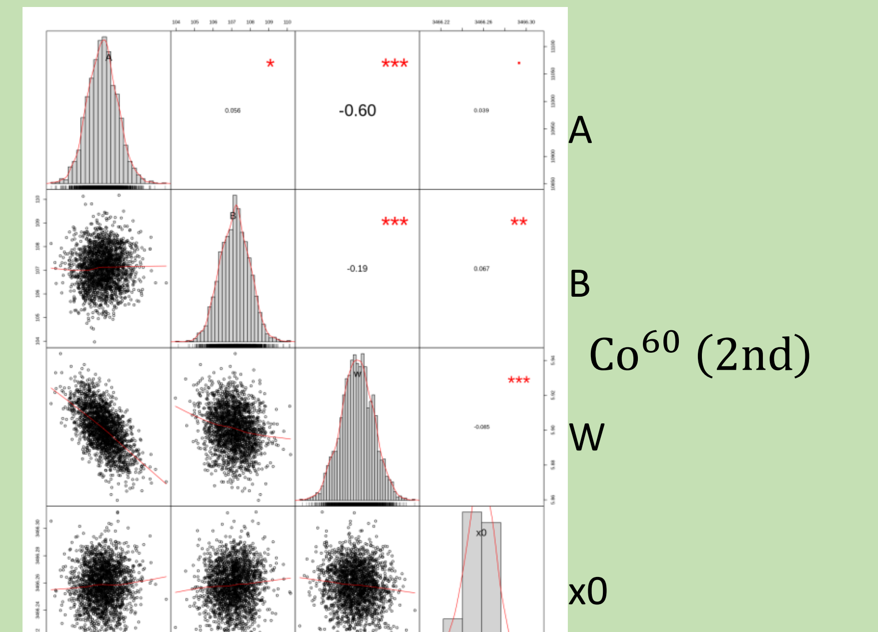
- Quantitatively, correlation matrix was found using

  cor(chain.df)

- Signal Amplitude A and width of signal peak w are negatively correlated.

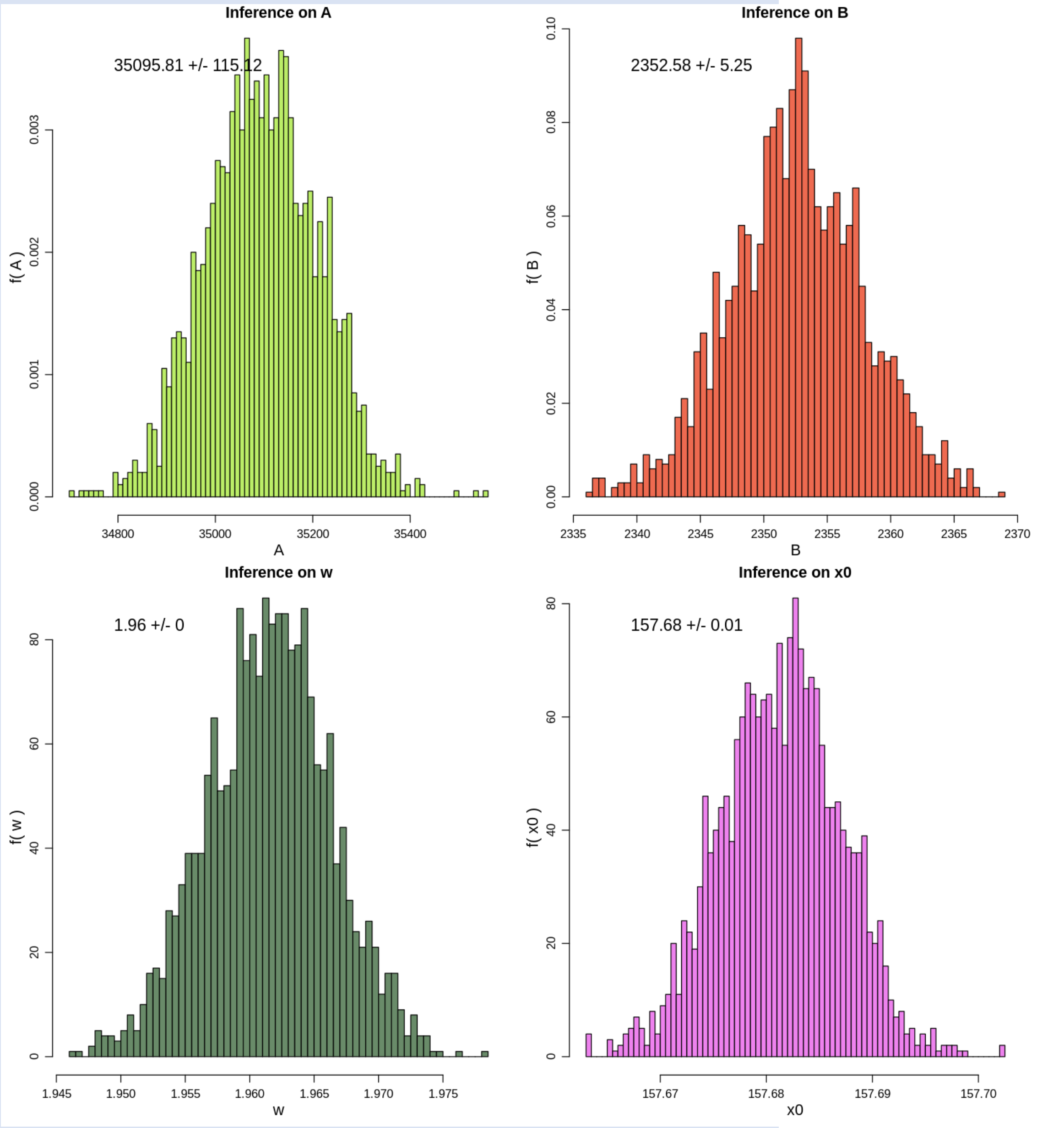




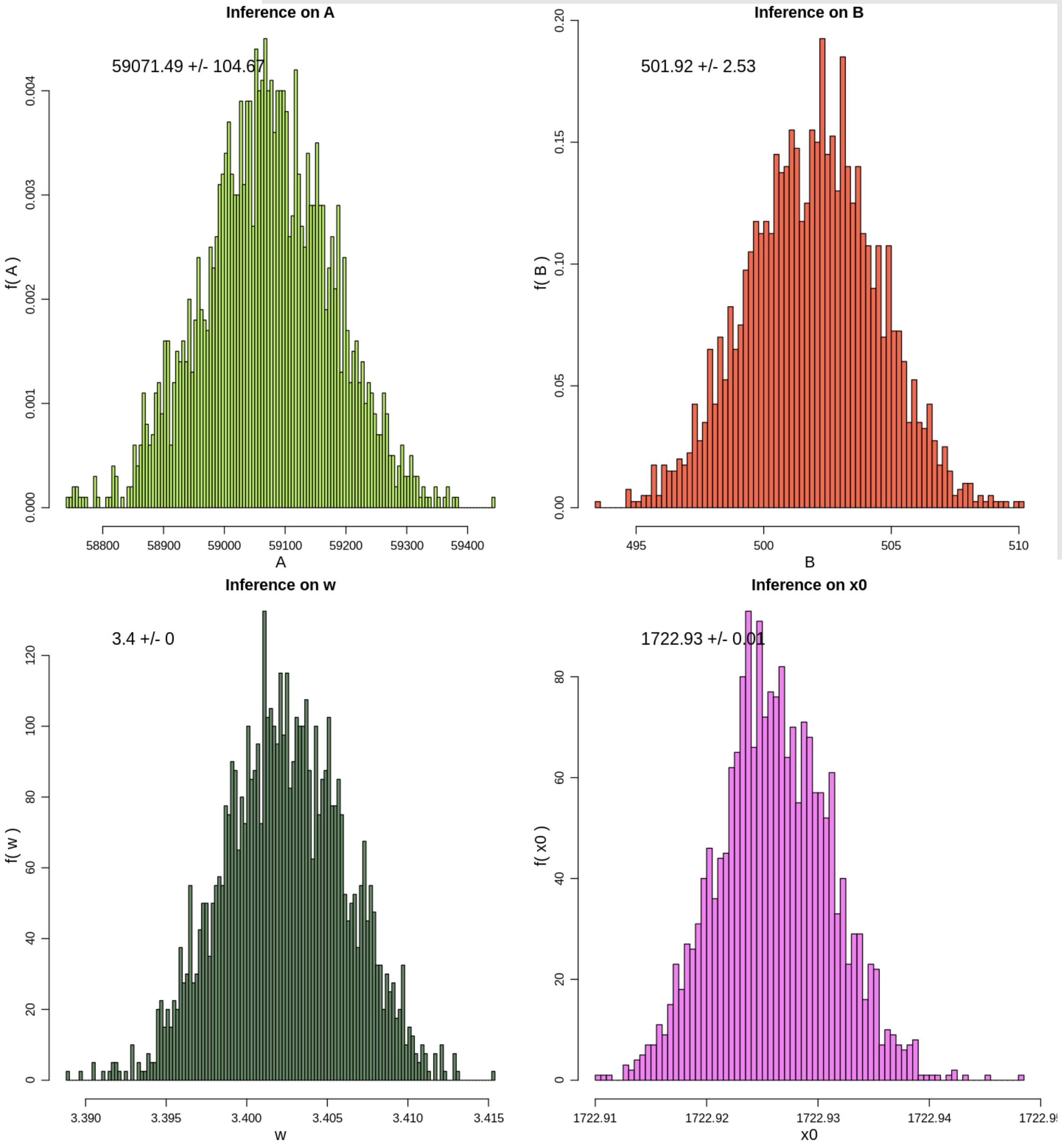


Histograms

Am^{241}

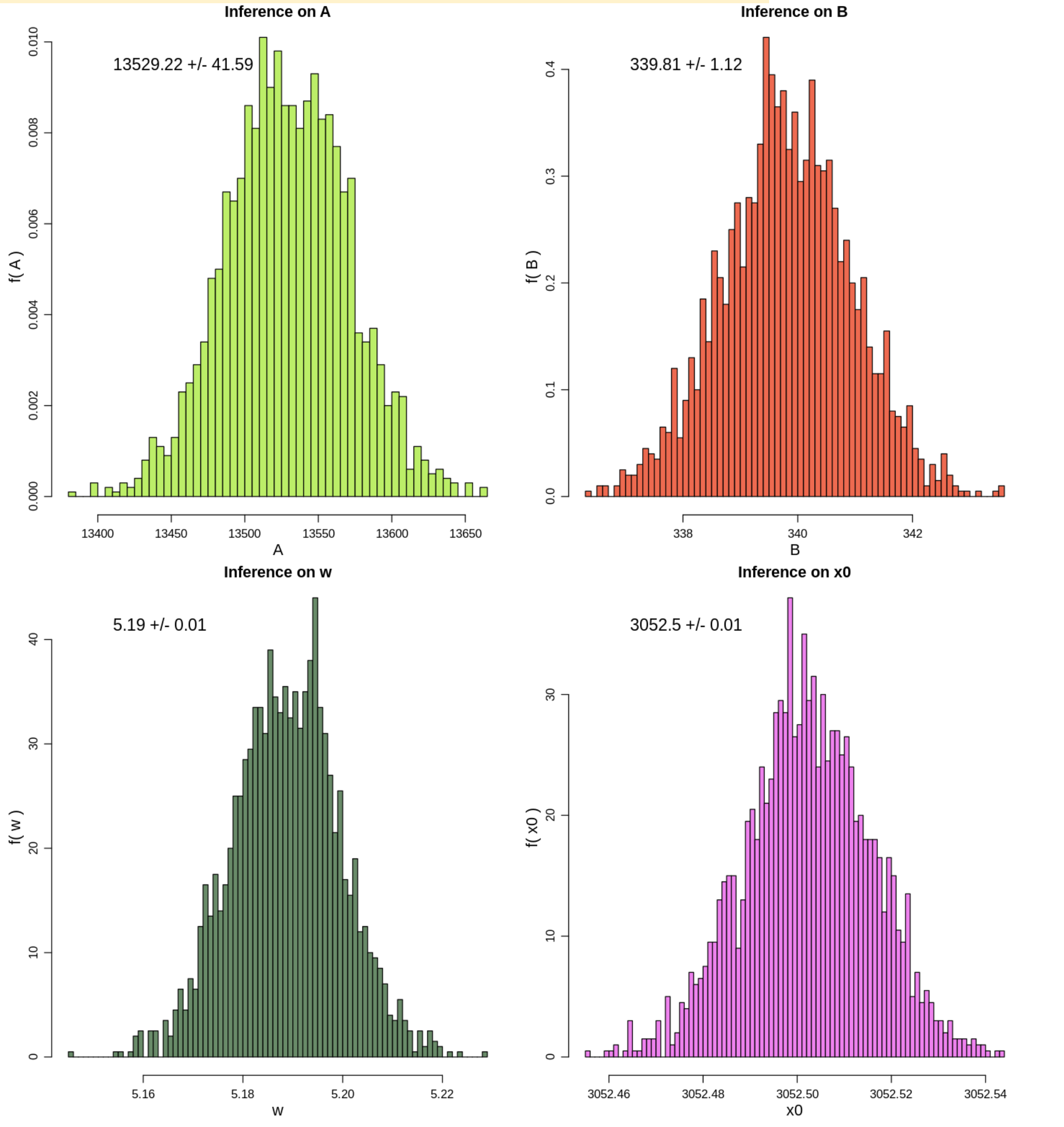


Cs^{137}

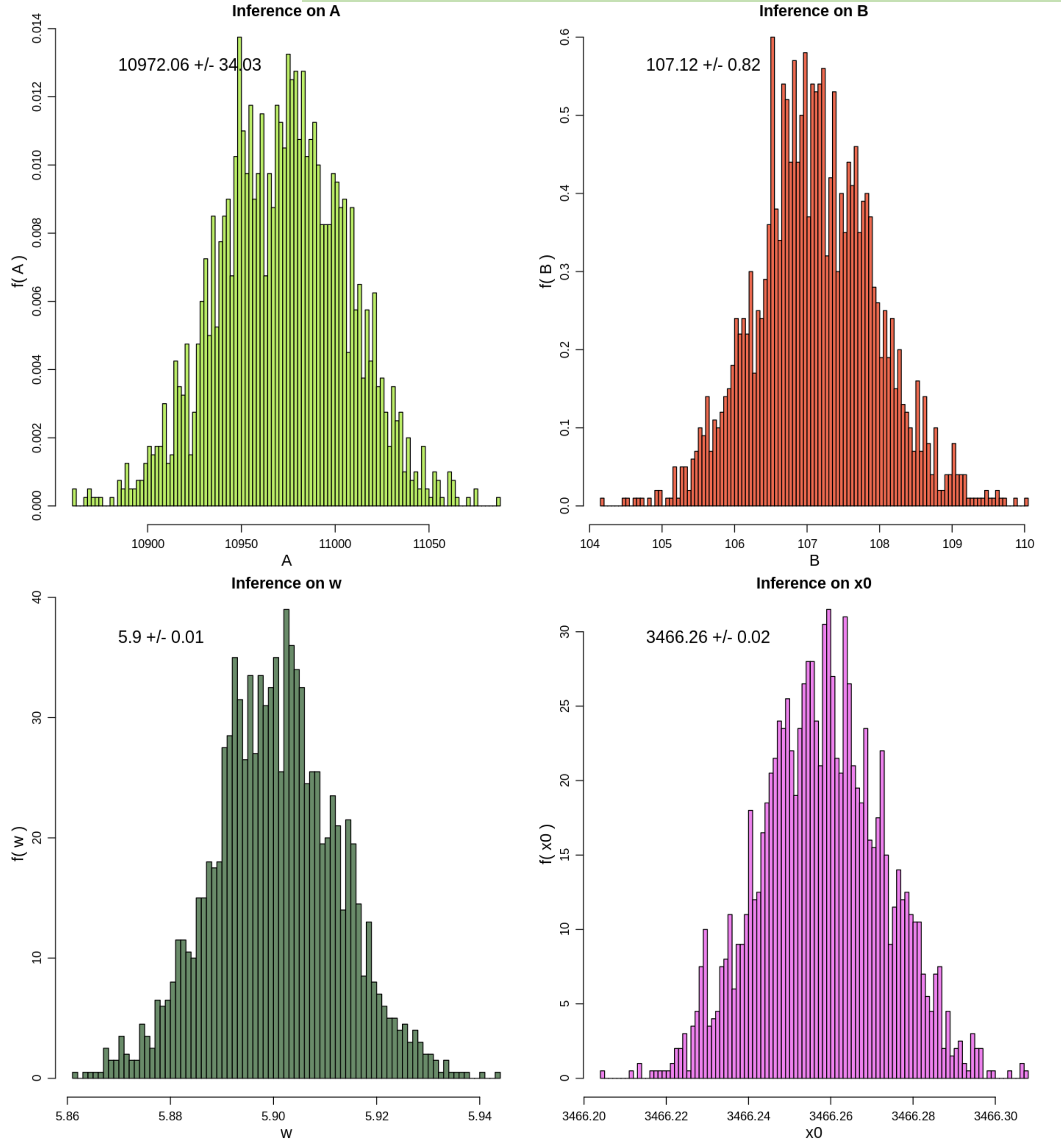


Histograms

Co^{60} (1st)



Co^{60} (2nd)



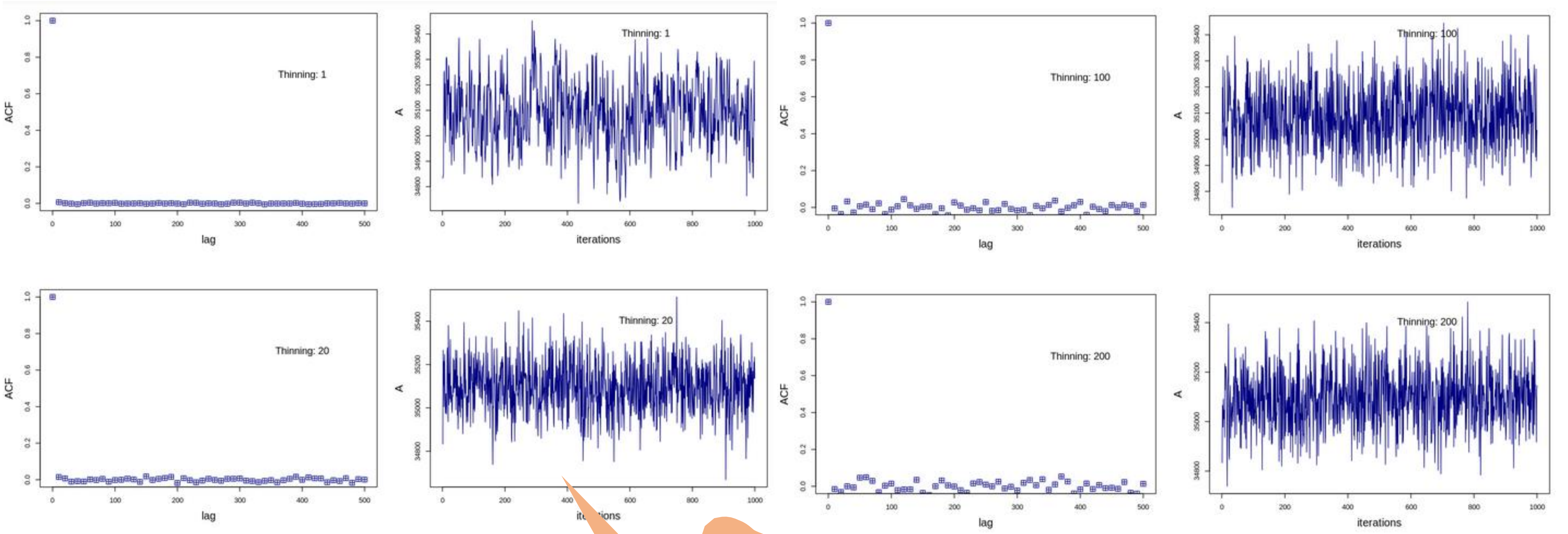
Convergence of MCMC

- Check whether the chain has reached **a steady state** by rerunning the sampling several times with different starting points
- All **chains should converge to the same** region of parameter space

Initial values	First experiment	Second experiment	Third experiment	Fourth experiment
A	30000	50000	25000	20000
B	5000	3500	2000	4000
W	8	20	15	5
X0	150	180	140	170
Estimates for the posterior distribution	A : 35096.5 +/- 113.582 B : 2352.595 +/- 5.1048 W : 1.961533 +/- 0.004806947 x0 : 157.6811 +/- 0.005723299	A : 35098.58 +/- 114.0157 B : 2352.465 +/- 5.256492 W : 1.961464 +/- 0.004956742 x0 : 157.6812 +/- 0.005701202	A : 35092.41 +/- 116.0715 B : 2352.489 +/- 5.174249 W : 1.961666 +/- 0.004920398 x0 : 157.6813 +/- 0.005606925	A : 35100.8 +/- 113.8174 B : 2352.439 +/- 4.973497 W : 1.961508 +/- 0.005030203 x0 : 157.6814 +/- 0.005631277
Estimates for the posterior distribution 95% CI	A : 34884.54 - 35325.63 B : 2342.175 - 2362.943 W : 1.951842 - 1.97039 x0 : 157.6699 - 157.6918	A : 34871.04 - 35317.44 B : 2342.518 - 2362.644 W : 1.951621 - 1.9714 x0 : 157.6702 - 157.6926	A : 34866.18 - 35322.96 B : 2342.323 - 2363.038 W : 1.952452 - 1.971218 x0 : 157.6709 - 157.6922	A : 34886.11 - 35328.36 B : 2342.896 - 2361.772 W : 1.95189 - 1.971365 x0 : 157.6705 - 157.6923

Convergence of MCMC

- Effect of **thinning** on Auto-correlation



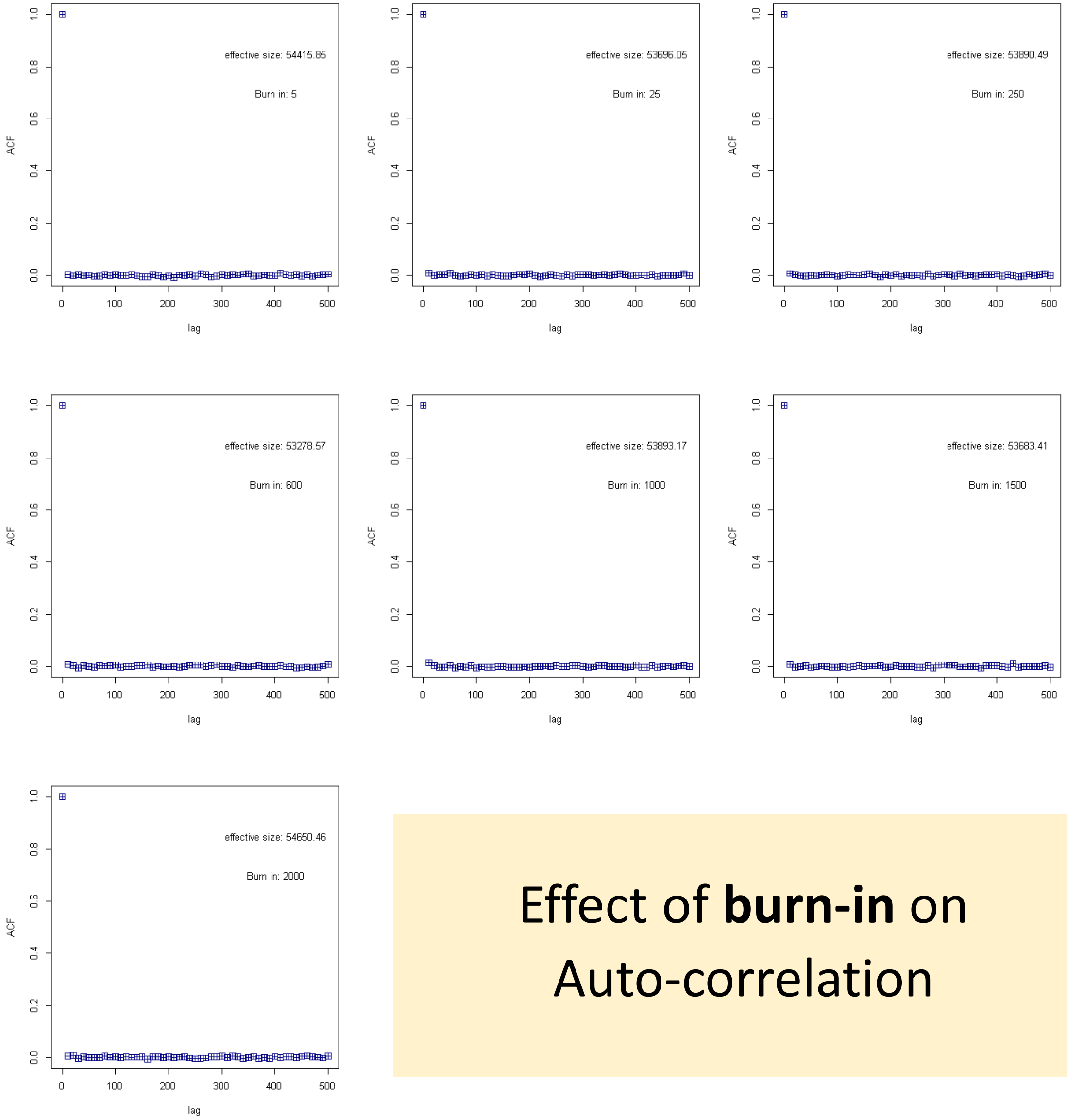
Auto-correlation or **ACF(h)** measures how closely the chain is correlated with itself h steps later

Burn in = 2000

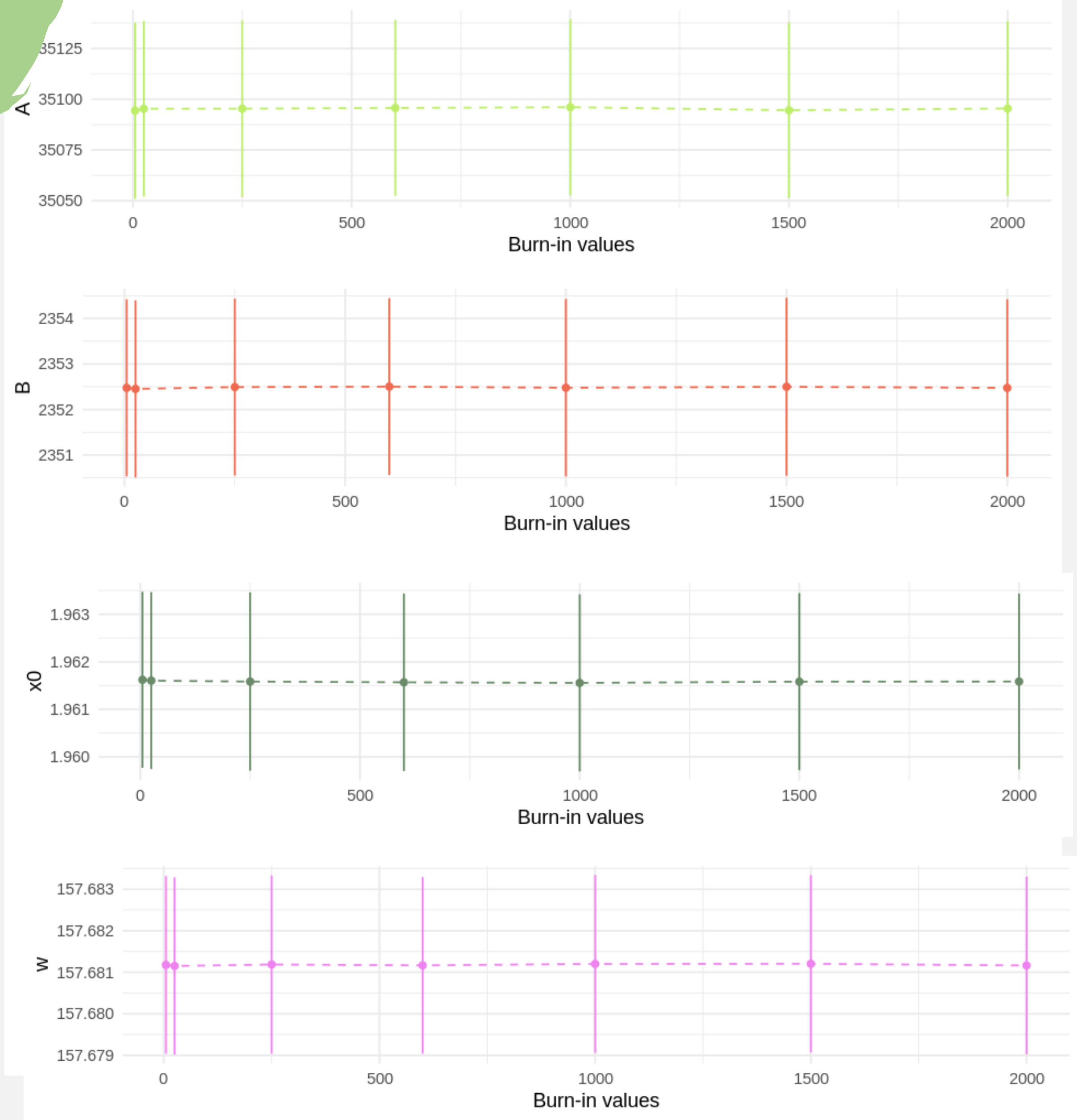
Even with **small values of burn-in** and **thin = 10**, for a given parameter (for eg, A), there ***is not much correlation*** h steps later for any h>1.

Convergence of MCMC

Thin = 1



Effective of **burn-in** on Auto-correlation



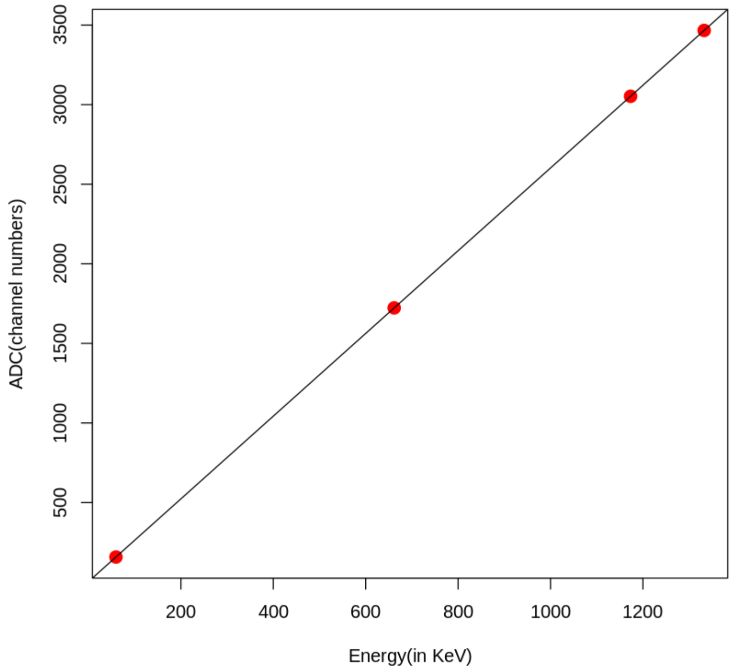
Calibration

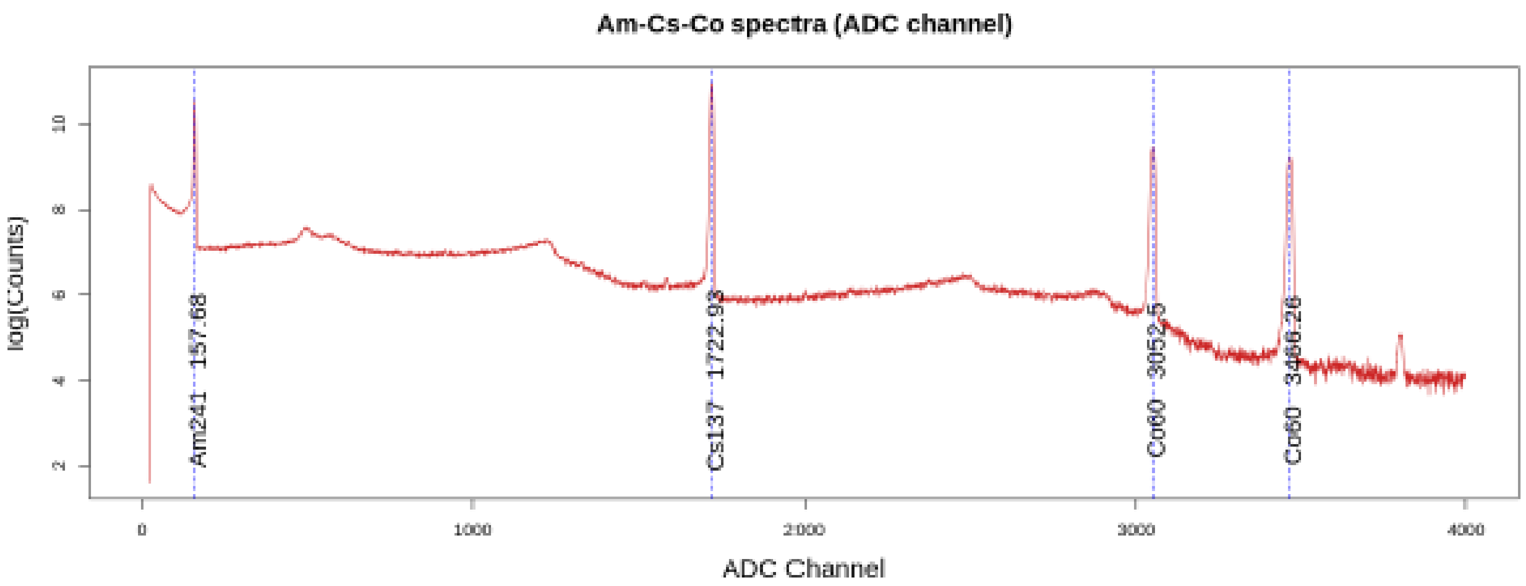
	Energy	x0
	<dbl>	<dbl>
Am-241	59.54	157.6812
Cs-137	661.66	1722.9260
C0-60.1	1173.24	3052.5027
C0-60.2	1332.51	3466.2582

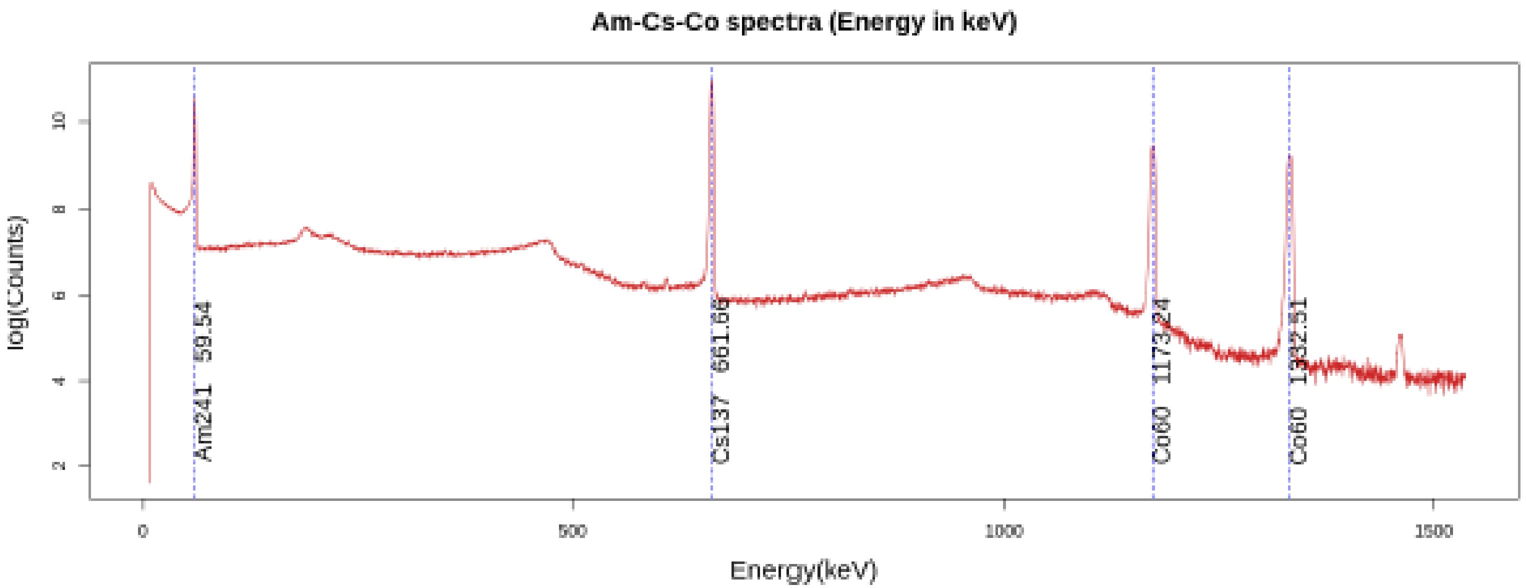
Energy = ADC channel slope + intercept

Intercept = -1.1627
Slope = 0.3847

```
# lm() in R is a good tool to model them.
regression <- lm(formula =   df$Energy~ df$x0 )
```





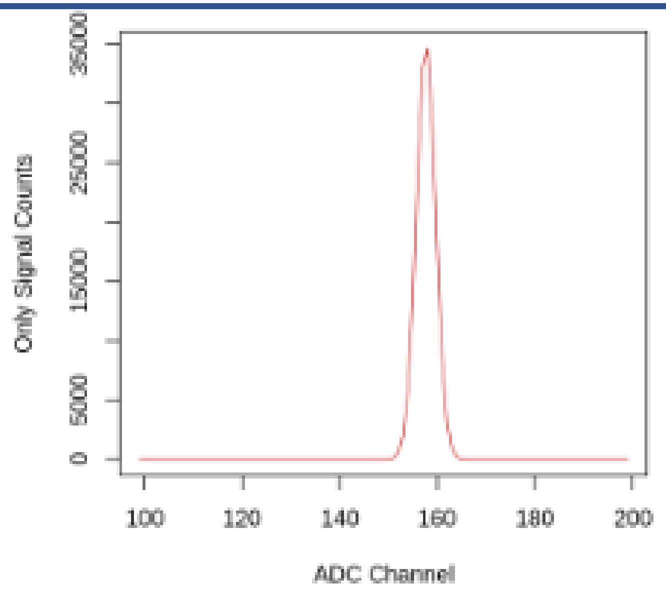


Estimating number of events

```
N.events <- function(nuclide.name,nuclide.data, nuclide.estimates){

    #empty vector to capture signal data
    S <- vector(mode="numeric", length=length(nuclide.data$x))
    #if we plot the gaussian noise without the constant background we obtained this graph:
    A <- nuclide.estimates$A
    w <- nuclide.estimates$w
    x0 <- nuclide.estimates$x0
    }
    for (i in 1:length(nuclide.data$x)){
    S[i]<- (A*exp((-(nuclide.data$x[i]-x0)^2)/(2*w^2)))}


    # plot the signal where the params were estimated using JAGS
    options(repr.plot.width=5, repr.plot.height=5)
    plot(nuclide.data$x, S, col='firebrick3', type='l',xlab="ADC Channel",ylab="Only Signal Counts",
     main=paste(nuclide.name," spectra"))

    Signal <- function(x) { (A*exp((-(x-x0)^2)/(2*w^2)))}
    # just integrate in the range of nuclide.data$x , so that the integral value will be zero in those places
    # where the gaussian function doesn't have any values.
    val <- integrate(Signal,lower = min(nuclide.data$x), upper = max(nuclide.data$x))$value
    cat("Number of events under the",nuclide.name,"peak:",val)
}
```

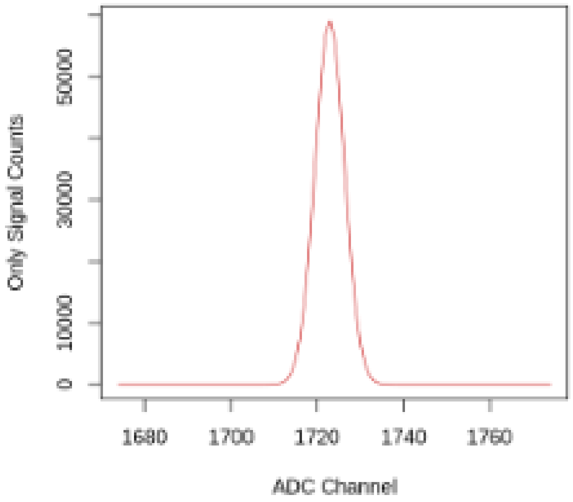
Estimating number of events

Results



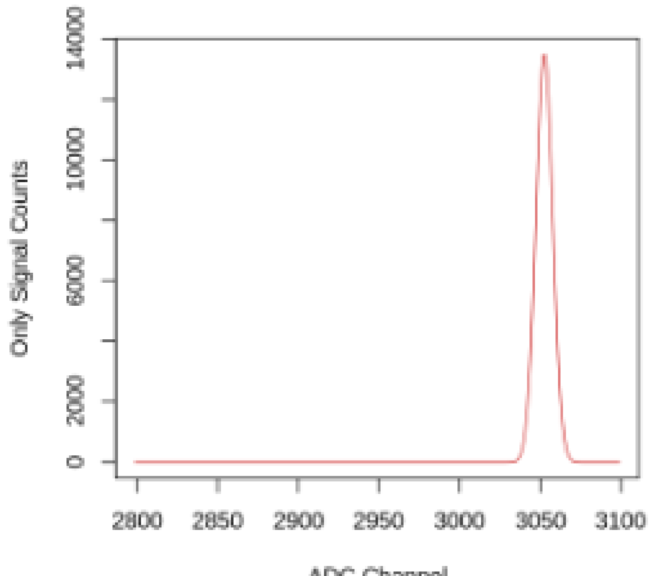
Number of events under the peak of $\mathbf{Am^{241}}$:

172585.2



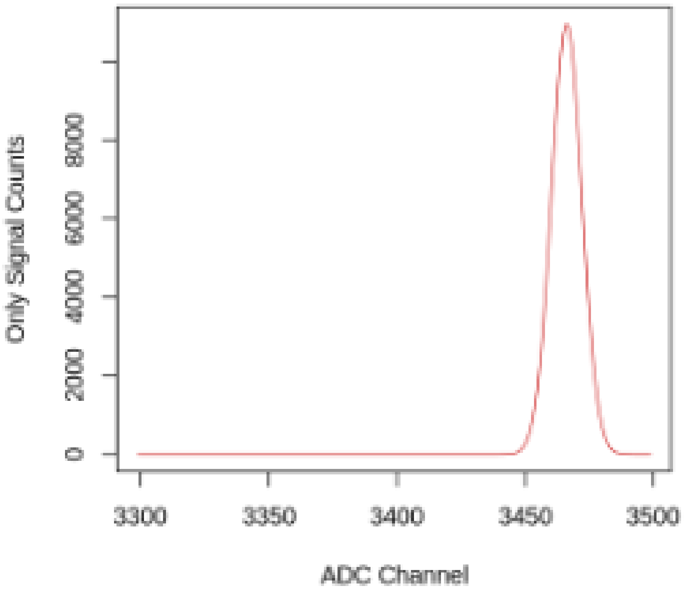
Number of events under the peak of $\mathbf{Cs^{137}(1st)}$:

503785



Number of events under the peak of $\mathbf{Co^{60}\ (1st)}$:

175961.2



Number of events under the peak of $\mathbf{Co^{60}\ (2nd)}$:

162277.2

Thank you for your attention