# Frequentist and Bayesian Approaches in X-ray Spectral Fitting

Tejas Sewak[1]

[1] *Department of Physics & Astronomy, Texas Tech University*

## ABSTRACT

Spectral fitting of X-ray data is traditionally performed using the frequentist approach due to its computational efficiency. However, this method does not fully explore parameter degeneracies and may become trapped in local minima, making the fitted model dependent on initial parameters and minimization techniques. To draw more robust conclusions, a Bayesian approach can be applied, allowing for a broader exploration of the parameter space and providing posterior distributions to assess degeneracies among parameters. The main goal of this project is to employ both frequentist and Bayesian approaches to fit X-ray spectral data and compare their results and computational times. The frequentist analysis is done using XSPEC and for the Bayesian analysis I utilised the EMCEE method. I have fitted spectral data of the active galactic nucleus (AGN) NGC 4051 from NICER (0.2–15 keV) using the RELXILL (García et al. 2014) model. I compared two different versions of the model: fixed low density ($10^{15}$ cm$^{-3}$) and a free density ($< 10^{20}$ cm$^{-3}$) based on criteria such as reduced chi-squared, Bayesian Information Criterion (BIC), and Akaike Information Criterion (AIC). I further discussed importance of posterior distributions for robust scientific conclusions and how both frequentist and Bayesian approach go hand-in-hand for X-ray spectral fitting.

## 1. INTRODUCTION

Active Galactic Nuclei are supermassive black holes at the center of galaxies which are actively accreting matter. The accretion disk is typically modeled as a geometrically thin, optically thick disk which emits thermal radiation as a blackbody (Shakura & Sunyaev 1973). The temperature of the disk is radially dependent but the continuum emission peaks in UV/Optical. However, there is also strong X-ray radiation observed from such sources. This X-ray radiation is assumed due to inverse Compton scattering of seed photons from the disk due to presence of hot-relativistic electrons ($kT_e \sim 50 - 100$ keV) in "corona". The observed spectrum from such an inverse Compton scattering process is a power-law feature with a high energy cutoff related to the temperature of the electrons. This X-ray emission can also hit the accretion disk and reflect back due to ionized emission. Such reflection features have been observed in the spectrum with a detection of broad Fe K$\alpha$ line at a rest frame energy of 6.4 keV. However, certain Active Galactic Nuclei also show an additional feature at lower energies above the power law, called as the soft excess. NGC 4051 is one such AGN which shows very strong soft excess features at lower energies (see Figure 1). The soft excess feature has been interpreted in terms of relativistically broadened emission lines at lower energies (García et al. 2016). In this paper, we have used the model RELXILL (García et al. 2014) which was developed to model the relativistically broadened reflection features of an X-ray spectrum. Separating the X-ray spectrum into different spectral components can allow us to see their variability over time and help constrain the reflection models better. Recently, a new flavor of RELXILL model was developed (García et al. 2016) where the density of the accretion disk was allowed to vary up to $10^{20}$ cm$^{-3}$ instead of the older versions with a fixed density of $10^{15}$ cm$^{-3}$. It was shown that increasing the density of the disk enhances the reflection spectrum at lower energies and could possibly be the case for AGNs which show soft-excess features. To test the X-ray spectrum modeling of NGC 4051 using the variable density RELXILL model, we perform a Frequentist and Bayesian analysis of data taken from NICER ($0.22 - 15.0$ keV).

## 2. OBSERVATIONS AND DATA REDUCTION

The intention of spectral modeling is to separate model components from different radiative processes. Thus, to study AGNs with soft excess, we made use of data from NICER instrument which has a high signal to noise ratio at lower energies. We have used the observations of NGC 4051 from the NICER X-ray Timing Instrument (XTI). All NICER data is publicly available and can be accessed through NICER Master Catalog. We chose data for one such observations dated 2022-09-29. The data reduction was performed using the standard procedure outlined in the NICER data analysis thread [1]. Using the pipeline, the X-ray spectral file (.pha) and background files were generated. The exposure for one such observation was about $\sim 3000$ seconds, with an average count rate of 1.33 cts/s. The X-ray spectrum observed is shown in Figure 1 over the energy range $0.22 - 15.0$ keV. As evident from Figure 1, the background started to dominate for higher energies ($>\sim$ 8keV) which is accounted by the variable SCORPEON background model.

[1] https://heasarc.gsfc.nasa.gov/docs/nicer/analysis_threads/

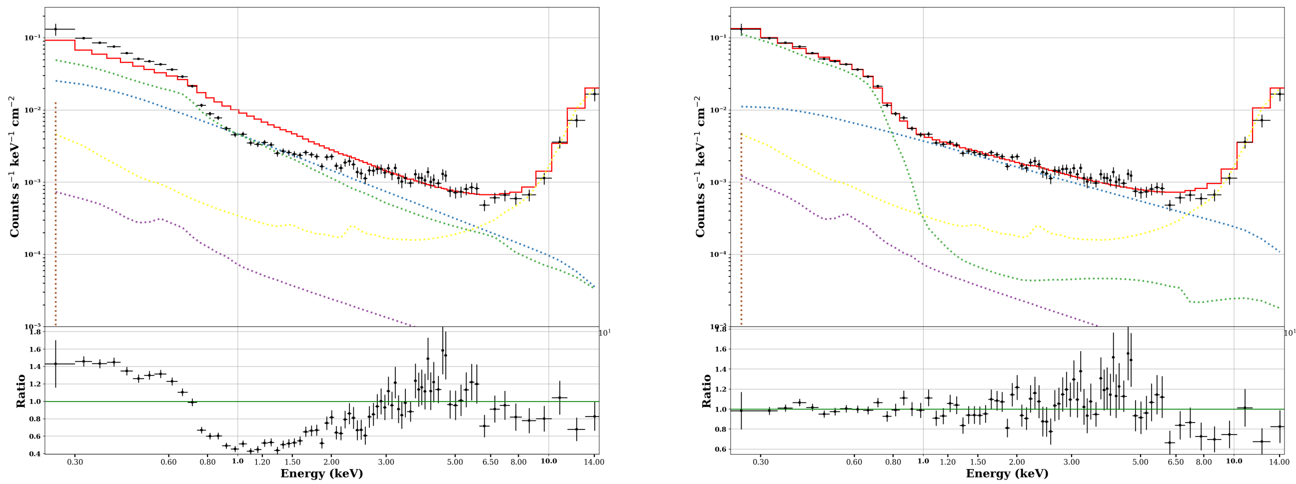


**Figure 1.** The total source+background spectrum is shown above. The two plots show the spectrum with the initial model and the best-fit model obtained using the Frequentist approach. In the figures, the red line indicates the complete model (source+background). The dashed blue and green line indicates the `nthcomp` and `RELXILLCP` model components while the yellow, purple and orange represents the non X-ray, X-ray and noise background components. Overall, we see a strong soft-excess above power law which is modeled with the `RELXILLCP` reflection component above the `nthcomp` component of the model. The best-fit model parameters are presented in Table 1.

## 3. ANALYSIS

We have performed two sets of analysis, one following the Frequentist approach and the other using the Bayesian approach. For the Frequentist method, we extensively used the software XSPEC (Arnaud et al. 1999) for $\chi^2$ minimization to find the best fit parameters. The Bayesian analysis was performed using the EMCEE sampler (Foreman-Mackey et al. 2013) on PYTHON to obtain the posterior distributions of the parameters.

For any kind of spectral modeling, we need to assume a possible model which can replicate the true distribution of spectrum. For our analysis, we are using a combination of two different physical models: `nthcomp` and `RELXILLCP`. The component `nthcomp` (Zdziarski et al. 1996; Życki et al. 1999) models the thermally Comptonized continuum which has an asymptotic power-law feature with photon index as the power. This model is chosen over a simple power law because it has a high energy cutoff dependent on temperature of Comptonizing medium. It has two free parameters, the photon index ($\Gamma$) and flux of the component. The model `RELXILLCP` (García et al. 2016) is a component that models the reflection part of the spectrum. It is a combination of physical parameters of the accretion disk such as spin, inclination, truncation radius for the disk, corona size, density of the disk, iron abundance, emissivity parameter and ionization of the disk. A combination of these parameters generate a reflection component of the spectrum, for which either the best-fit parameters or the posterior distributions can be obtained. In total, our model has 8 free parameters, 2 for `nthcomp` and 6 for `RELXILLCP`.

### 3.1. *Frequentist Analysis*

Frequentist analysis gives us best-fit point estimates for our parameters. In the context of this paper, the Frequentist analysis was done to compare the results with the Bayesian method. To minimize the $\chi^2$ (maximize likelihood), we made use of the software `XSPEC` which is the traditionally used software in X-ray spectral fitting. Our motivation is to compare the best fit parameters obtained using `XSPEC` with the posterior distribution for each parameter. This is to say that we want to check if our best-fit parameters are not stuck in a local minima but indeed have achieved the best-fit. If the best-fit parameters are found to be in $1\sigma/2\sigma$ region of our posterior distribution, we can safely concluded that the best-fit was found.

We made use of the `PYTHON` version of `XSPEC` —`PYXSPEC` for the analysis. After loading the spectrum file (along with the background, redistribution matrix, and response files), we created a model which was a combination of `nthcomp` and `RELXILLCP`, as discussed above. Based on the count rate of our source, different statistics can be chosen for spectral fitting. If we are in the low-count regime, i.e., Poisson distributed data, the maximum likelihood-based statistic is the Cash statistic (Cash 1979). However, in a high-count regime, where the distribution can be approximated to be Gaussian, $\chi^2$ can be used. Although for NGC 4051, we are in a low count regime, the spectrum was grouped together to ensure a minimum of 25 cts/bin, where the distribution approaches Gaussian. This was done for a easier check of the best fit found using the scale $\chi^2$/dof $\sim 1$. Another reason for that is because NICER data is dominated by systematics rather than statistics, the choice

**Table 1.** Best-fit and Posterior Estimates of Spectral Parameters

| Parameter | Frequentist Best-fit | Bayesian Posterior Median |
|---|---|---|
| Flux $\log(\Phi_{\rm nth})$ [erg cm$^{-2}$ s$^{-1}$] | $-10.94 \pm 0.05$ | $-10.94^{+0.03}_{-0.03}$ |
| Flux $\log(\Phi_{\rm rel})$ [erg cm$^{-2}$ s$^{-1}$] | $-11.18 \pm 0.11$ | $-11.17^{+0.03}_{-0.03}$ |
| Inclination $i$ [deg] | $40.22 \pm 6.49$ | $40.07^{+3.34}_{-3.37}$ |
| Photon Index $\Gamma$ | $1.35 \pm 0.11$ | $1.3496^{+0.0381}_{-0.0383}$ |
| $\log \xi$ [erg cm s$^{-1}$] | $2.16 \pm 0.06$ | $2.1637^{+0.0531}_{-0.0535}$ |
| $\log N$ [cm$^{-3}$] | $20.00 \pm 3.07$ | $19.8474^{+0.1903}_{-0.1895}$ |
| $A_{\rm Fe}$ [solar] | $6.70 \pm 1.40$ | $6.8075^{+1.1432}_{-1.1832}$ |
| Fit Statistic | $\chi^2/\text{dof} = 259.05/133$ | $\ln \mathcal{L}_{\max} = -129.65$ |

Note—Frequentist errors are $1\sigma$ values estimated using the Fisher matrix and are symmetric. Bayesian values are posterior medians with 68% credible intervals derived from the marginal distributions and could be asymmetric. Parameters like $kT_e = 100$ keV, spin $a = 0.998$, corona size Rbreak $= 6.0$ r$_{\rm g}$, and reflection fraction $= 1$ were fixed during fitting.

of statistics may not make much difference. Although both can be done and compared, for this analysis we assumed that our data points (number of counts) are from a Gaussian distribution. Thus, we made use of $\chi^2 = \Sigma_{i=1}^{N} (y_i - m_i)^2/\sigma_i^2$ statistics for minimizing.

The fit was performed and a $\chi^2/\text{dof} = 120.57/78 = 1.54$ value was obtained for energy range up to 8.0 keV, as above that the spectrum is background dominated. This is an acceptable fit and the best-fit parameters obtained for our model are presented in Table 1. The errors reported are the $1\sigma$ uncertainties obtained using the Fisher Matrix. To see if this best-fit obtained with an acceptable $\chi^2/\text{dof}$ value is indeed the best-fit, we will compare results with the posterior distribution of the parameters using a Bayesian approach.

### 3.2. *Bayesian Analysis*

To perform Bayesian analysis, we used the EMCEE sampler available on PYTHON (Foreman-Mackey et al. 2013). EMCEE requires a log-likelihood function to determine the probability of some combination of parameters. The priors for all parameters were kept uniform but bounded by the possible range for each parameter. This was the only choice of prior as no conclusive knowledge about the system is known which can be incorporated in our models. For example, if we had known the precise value of inclination of the system through a different study (say radio or polarization technique), we might have used a Gaussian prior around the obtained value. However, in absence of any such prior knowledge, a uniform prior is the only choice. This implies that the prior will have no impact on our likelihood evaluation. To compute the log-likelihood value, we used the statistic value ($\chi^2$) value obtained for any combination of parameters. This was because the models are created to be loaded in XSPEC configuration and a model spectra can not be generated for certain parameters in PYTHON itself. The starting guess for the parameters were chosen from a normal distribution created for each parameters centered at their best-fit values found from the Frequentist approach with a scale of $1\delta$ where $\delta$ is the step size used in Frequentist method (usually 1% of the parameter). The choice of the initial guess was motivated for faster convergence fo the chain. Since, this is a complex model with 8 free parameters and possible correlations between parameters may exist, the time required for the sampling chain to converge will be more. For this reason, we used a total of 128 walkers. Each walker took 10000 steps with a burn in period of 500 steps. In total, this will give a chain length of $\sim 10^6$ steps for one walker which will be enough for convergence. However, obtaining such a long chain comes with computational cost. On a 4-core machine, the estimated time for a $\sim 10^6$ chain is $\sim 24$ hours. To increase the speed of computation, we parallelized the process on TTU HPCC cluster with 1 node using 128 cluster, i.e., by making use of 128 parallel walkers. The computational time was reduced to $\sim 1$ hour.

To check for convergence of the chain, we obtained the trace plot for each parameter and checked if the chains are mixed or not at each step visually. Figure 3 shows an example trace plot for the flux of `nthcomp` component of our model which hints towards the chain being converged. As a further check, we checked the average acceptance rate across all walkers which was found to be $0.383( > 0.2)$ which also concludes that the chain was converged. The average autocorrelation length was found to be $\sim 100$ which resulted in about 10k independent steps which also implies that the chain has sufficient number of independent samples and that the chain is converged. All chain diagnostics are presented in Table 2.

To visualize the posterior of the parameters, we plotted a corner plot (Foreman-Mackey 2016) for all free parameters as shown in Figure 2. All parameters have been constrained as visually seen from the individual posterior distributions. For each parameter, the median value of the posterior along with $1\sigma$ errors were mentioned and marked on the distributions. Thus, using

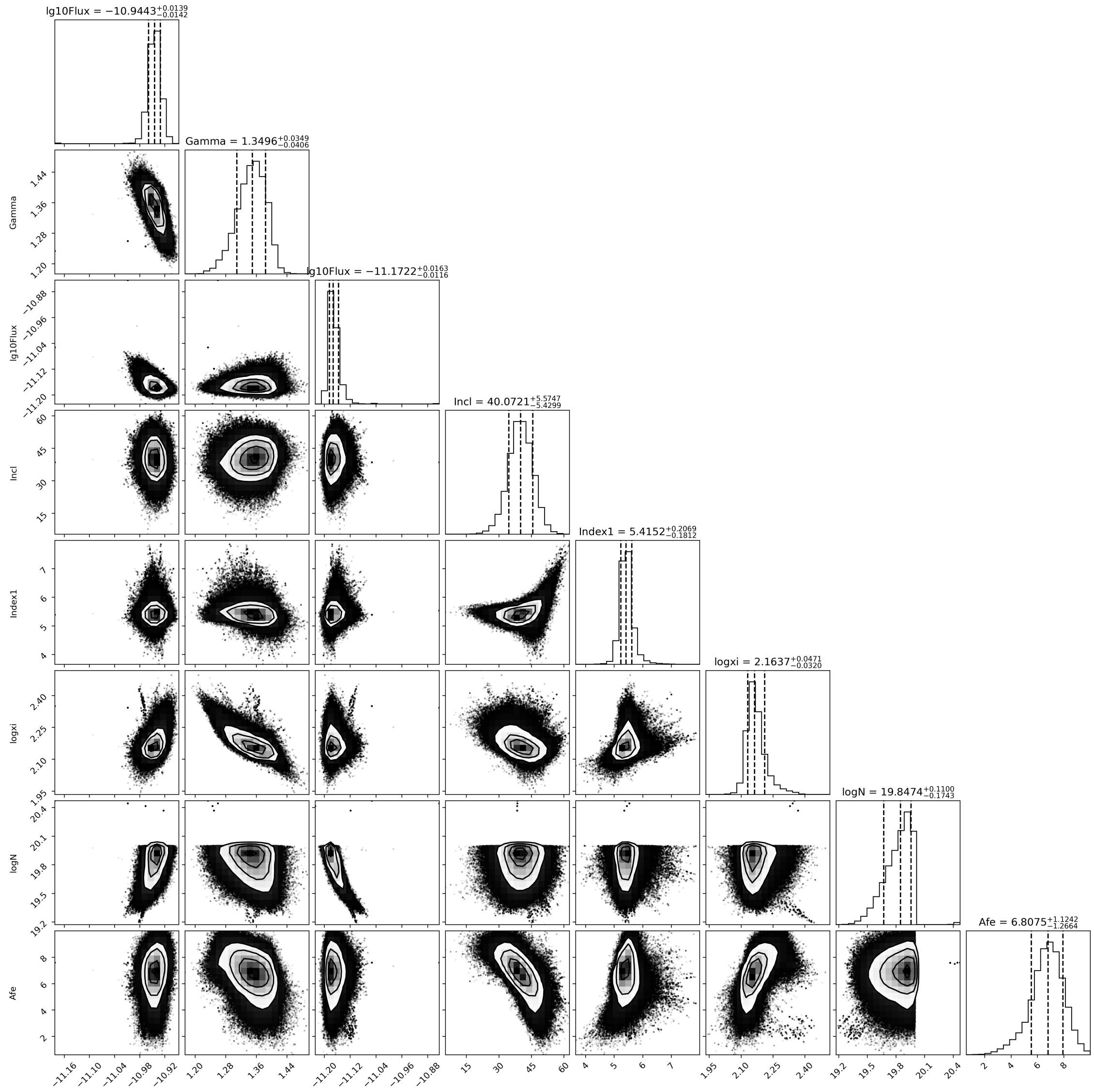


**Figure 2.** Corner plot showing posterior distributions for all free parameters in our model: `nthcomp`(2 free parameters)+`RELXILLCP`(6 free parameters). The 2-d histogram can be used to see for correlation between parameters. The median of the distribution along with the $1\sigma$ errors are marked on the posteriors.

the Bayesian method, we have obtained the posterior distributions for all our free parameters.

Table 1 shows the best-fit parameters obtained using Frequentist approach and the medians of our posterior distribution. It is evident that both the approach gave similar results. Thus, we can conclude that our Frequentist best-fit was indeed the best-fit and did not get stuck in a local minimum.

### 3.3. *Model Comparison*

The reflection spectra generated by RELXILLCP strongly depends on free physical parameters in our model. One such free physical parameters is the density of electrons in the accretion disk. This was recently shown by García et al. (2016) to significantly enhance the reflection component of the model at lower energies ($<$ 2 keV). Canonically, a low density of electron ($10^{15}$ cm$^3$) model was used. However, the model has been updated to allow for higher order of densities (upto $10^{20}$ keV). For our analysis, we used the latest RELXILLCP model which allows for a variable density of electrons in the accretion disk. This was motivated to produce the strong reflection component seen in the NGC 4051 spectrum evident from Figure 1. However, to test the significance of the free density parameter model (Model 2) over a fixed low density model (Model 1), we performed model comparison. Model comparison can be done two ways: physical interpretation or statistical evidence. However, for this analysis, we are only concerned with distinguishing two different models solely based on

**Table 2.** Bayesian Chain Diagnostics and Information Criterion Comparison for Models

| Quantity | Model 1 (Fixed Density) | Model 2 (Variable Density) |
|---|---|---|
| Mean Acceptance Fraction | 0.280 | 0.383 |
| Mean Autocorrelation Time | 720.88 | 134.90 |
| Number of Effective Samples | 3551 | 9489 |
| Bayesian Information Criterion (BIC) | $BIC_1 = 433.05$ | $BIC_2 = 371.8$ |
| Akaike Information Criterion (AIC) | $AIC_1 = 343.76$ | $AIC_2 = 275.3$ |
| ΔBIC (Model 1 - Model 2) | — | 61.25 |
| ΔAIC (Model 1 - Model 2) | — | 68.45 |
| Bayesian Evidence (Z) | $9.244 \times 10^{-74}$ | $2.715 \times 10^{-58}$ |
| Log Bayesian Evidence (log Z) | -168.2 | -132.6 |
| $\Delta \log Z$ (Model 2 - Model 1) | — | 35.6 |
| Bayesian Preference | — | Strong evidence for Model 2 |

Note— Chain diagnostics for Model 1 (fixed density) and Model 2 (variable density), from which it can be inferred that both the chains converged. The information criteria (BIC, AIC) suggest a statistically significant preference for Model 2 over Model 1. Interpretation of $\Delta \log Z$ follows the Jeffreys scale (Jeffreys 1961; Kass & Raftery 1995), where values $> 6$ indicate strong evidence in favor of the lower $\log Z$ model. The $\log Z$ values are on the lower end because for Likelihood $= 0.5 \times \chi^2$, the $\chi^2$ values used were for the entire energy range ($0.22 - 15.0$ keV). However, above 8.0keV the background model dominates and reduce the $\chi^2$ value significantly.

statistical evidence. This can again be done both in the Frequentist regime as well as Bayesian regime.

For the Frequentist approach, we can obtain the best fit parameters for both Model 1 and Model 2 and compare the $\chi^2$ statistics. For Model 1, the $\chi^2/\mathrm{dof} = 2.83$, and for Model 2, the $\chi^2/\mathrm{dof} = 1.54$. From the $\chi^2/\mathrm{dof}$ values it can be inferred that the Model 2 performs much better than Model 1 in producing the observed data. However, $\chi^2/\mathrm{dof}$ does not give a sense of how badly Model 1 performs, i.e., there is no scale on which the $\chi^2/\mathrm{dof}$ values can be compared. The only scale for $\chi^2/\mathrm{dof}$ is that it should be $\sim 1$. However, NICER data is dominated by systematics and any outlier there might increase the $\chi^2$ value. So although the Frequentist approach allows us to infer that Model 1 performs badly than Model 2, it does not give us a scale for us to compare the two different models. This can be done in the Bayesian approach where we can obtain Information Criterion for two different models. The difference between the two Information Criterion can be compared on a Bayesian scale which will allow us to infer that how significantly one model is preferred over the other. We obtained the posterior distribution for a fixed density model using the Bayesian approach described above. The chain was checked for convergence using the same criteria above and was concluded to be converged (chain statistics in Table 2). Figure 4 shows the corner plot for all free parameters for the fixed density model (Model 2), from which it can be inferred that all parameters are constrained.

From the two chains obtained for Model 1 and 2, we obtained the Bayesian Information Criterion and Akaike information criterion. $\mathrm{BIC} = -2 \times log(L) + k \times log(n)$ and $AIC = 2k - 2 \times log(L)$ where L is the maximum likelihood of the model, k is the number of free parameters in the model, and n is the length of the chain. A lower BIC/AIC value indicates a better model. We also obtained the Bayesian evidence (Z) for our models, where $Z \approx \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(\theta_i)$. Table 2 summarizes the BIC, AIC, $\log Z$, ΔBIC/AIC and $\Delta \log Z$ values. According to Jeffreys scale (Jeffreys 1961; Kass & Raftery 1995), we obtained a very strong evidence for a free density model over the fixed low density model.

## 4. SUMMARY AND CONCLUSIONS

We model the X-ray spectra from NGC 4051 taken using NICER satellite for observation dated 2022-09-29. To model the spectra, we used both Frequentist and Bayesian approach to obtain the best-fit and posterior distribution for parameters in our model. The model chosen was `nthcomp` + `RELXILLCP`. The model was a combination of two radiative processes: inverse Compton scattering of seed photons in corona and X-ray reflection from a high-density accretion disk. Using $\chi^2$ minimization, we obtained the best-fit parameters. The Bayesian setup was done using the `EMCEE` package in `PYTHON`. For our complex model with 8 free parameters and possible correlations, we sampled for a chain length of $\sim 10^6$ and was checked for convergence. The median values from the posterior distribution were compared with the best-fit parameters and were found to be same. This concluded that the Frequentist approach did not get stuck in a local minima and indeed found the best-fit model parameters. Next, we checked for evidence for employing a variable high-density ($10^{15} < N < 10^{20}$ cm$^{-3}$) disk model (Model

2) over a canonical fixed low-density $10^{15} = N$ cm$^{-3}$ disk model (Model 1). Using $\chi^2$ values we inferred that Model 2 gives a much better fit to the data when compared to Model 1. However, to robustly judge the significance of one model over the other, we performed Bayesian model comparison. Posterior distribution for Model 1 was found using the same method for Model 1. $\Delta$BIC/AIC and $\Delta \log Z$ values found indicated Very Strong evidence for a variable high-density disk model.

In conclusion, the statistical techniques developed here allows us to conclude physical interpretation over complex physical models in a simpler sense. For X-ray spectral fitting, the traditional Frequentist approach is extremely fast when compared to Bayesian analysis. However, Bayesian analysis gives much more robust conclusions and physical interpretation can easily be made by looking at the posterior distributions for parameters like density and emissivity of the disk.

## REFERENCES

Arnaud, K., Dorman, B., & Gordon, C. 1999, XSPEC: An X-ray spectral fitting package, Astrophysics Source Code Library, record ascl:9910.005

Cash, W. 1979, ApJ, 228, 939, doi: 10.1086/156922

Foreman-Mackey, D. 2016, The Journal of Open Source Software, 1, 24, doi: 10.21105/joss.00024

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306, doi: 10.1086/670067

García, J. A., Fabian, A. C., Kallman, T. R., et al. 2016, MNRAS, 462, 751, doi: 10.1093/mnras/stw1696

García, J. A., Dauser, T., Lohfink, A., et al. 2014, ApJ, 782, 76, doi: 10.1088/0004-637X/782/2/76

Jeffreys, H. 1961, Theory of Probability, 3rd edn. (London: Oxford University Press)

Kass, R. E., & Raftery, A. E. 1995, Journal of the American Statistical Association, 90, 773, doi: 10.2307/2291091

Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337

Zdziarski, A. A., Johnson, W. N., & Magdziarz, P. 1996, MNRAS, 283, 193, doi: 10.1093/mnras/283.1.193

Życki, P. T., Done, C., & Smith, D. A. 1999, MNRAS, 309, 561, doi: 10.1046/j.1365-8711.1999.02885.x

# APPENDIX

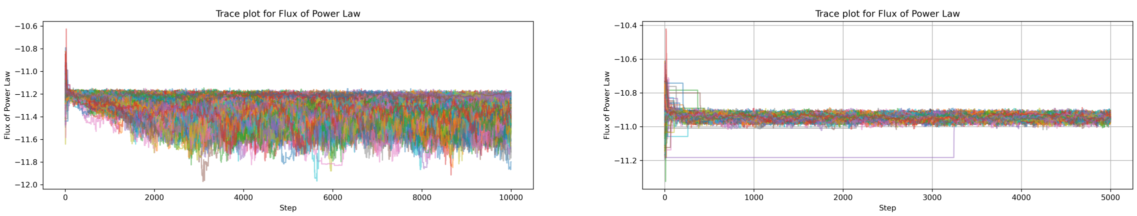


**Figure 3.** Trace plot for the parameter flux $\log(\Phi_{\rm nth})$ for Model 1 (fixed density) and Model 2 (variable density) indicating that both the chains are well mixed and converged.

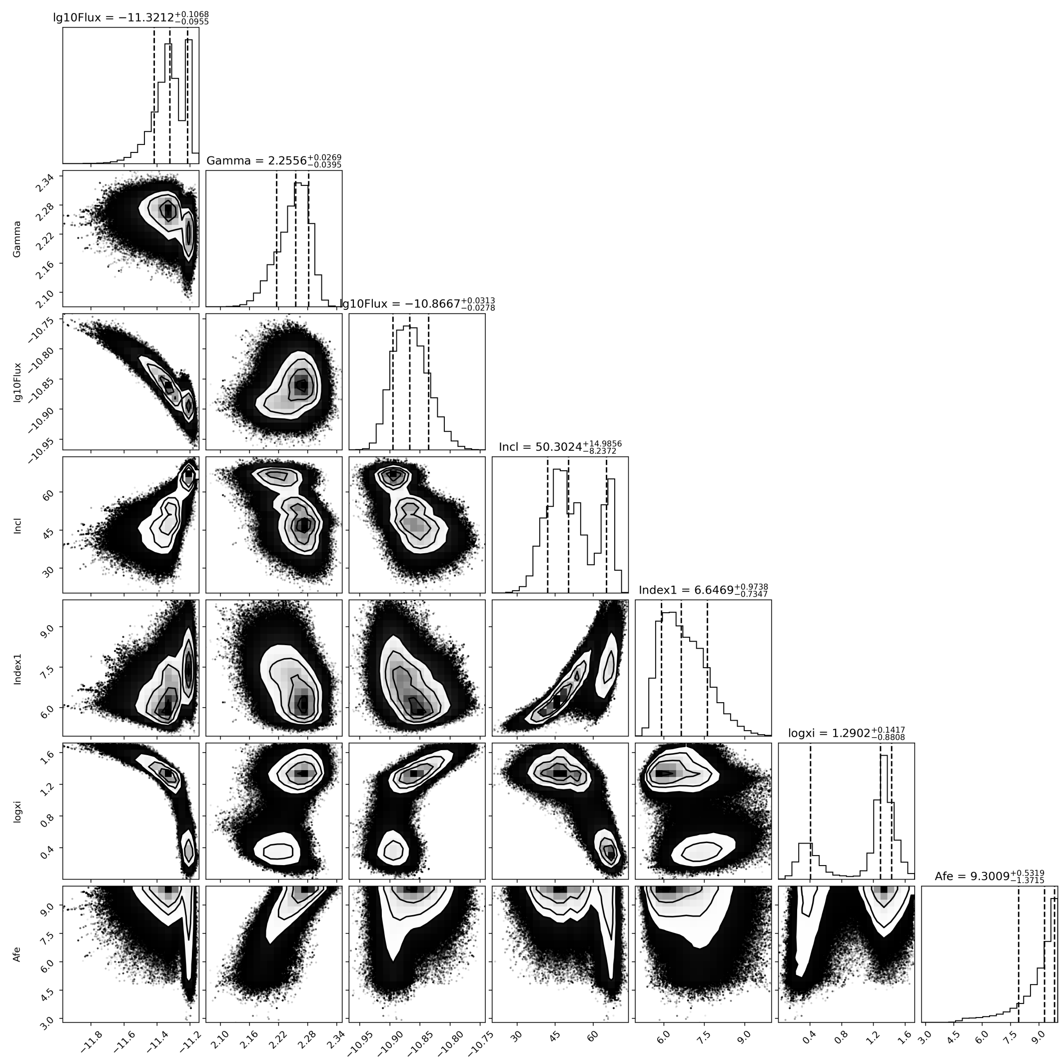


**Figure 4.** Corner plot showing posterior distributions for all free parameters in our model 1 (fixed density model): `nthcomp`(2 free parameters)+`RELXILLCP`(5 free parameters). The 2-d histogram can be used to see for correlation between parameters. The median of the distribution along with the $1\sigma$ errors are marked on the posteriors.

**Listing 1.** Python script used for the Bayesian spectral fitting analysis.

```
import time
import xspec
import numpy as np
import emcee
from multiprocessing import Pool
import matplotlib.pyplot as plt
import pickle

xspec.Xset.restore("5660010101/xti/event_cl/fitted_files/analysis0_rexillCp_manual_chisqgti1.xcm")
initial_model = xspec.AllModels(1)

par_val = [5,6,14,15,20,24,25,26] # thawed parameter number in the model. Example: 6th parameter is
    gamma of nthcomp.

def log_Likelihood(params):
    model = xspec.AllModels(1) # load the current model
    # update each model parameter
    for i in range(len(par_val)):
        model(par_val[i]).values = params[i]
    # calculate updated statistic and likelihood for those model parameters
    cstat = xspec.Fit.statistic
    log_likelihood = -0.5*cstat
    return log_likelihood

def log_Posterior(params):
    # Defined for a uniform prior between parameter range.
    # if any parameter is outside range, the prior value = 0, log(prior) = -inf, thus, log(Posterior
        ) = -inf
    for i in range(len(par_val)):
        pval, pdelta, pmin, pbottom, ptop, pmax = initial_model(par_val[i]).values # the range is
            same throughout
        # define a box-prior with the param range (in logarithm)
        if params[i] < pmin or params[i] > pmax:
            return -np.inf # log(0)
    # if all parameters are within the range, the prior value = 1, log(prior) = 0, log(posterior) =
        log(likelihood) + 0
    log_posterior = log_Likelihood(params)
    return log_posterior

# Run MCMC
ndim = len(par_val)  # number of parameters in the model
nwalkers = 256  # number of MCMC walkers
burn = 500  # "burn-in" period to let chains stabilize
nsteps = 10000  # number of MCMC steps to take **for each walker**

# initialize theta

best_fit = np.zeros(len(par_val))
delta_values = np.zeros(len(par_val))
for i in range(len(par_val)):
    best_fit[i] = initial_model(par_val[i]).values[0]
    delta_values[i] = initial_model(par_val[i]).values[1]

starting_guesses = np.random.normal(loc=best_fit, scale=5*delta_values, size=(nwalkers, ndim))

# the function call where all the work happens:
start = time.time()
with Pool() as pool:
    sampler = emcee.EnsembleSampler(nwalkers, ndim, log_Posterior,pool=pool)
    start = time.time()
    sampler.run_mcmc(starting_guesses, nsteps)
    end = time.time()
    print("Time to run 100 iterations =", end-start,"s")
    from multiprocessing import cpu_count
    ncpu = cpu_count()
    print("{0} CPUs".format(ncpu))

# save the sampler
```

```python
print("Saving data")
filename = f"sampler_w{nwalkers}_s{nsteps}.pkl"
with open(filename, "wb") as f:
    pickle.dump(sampler, f)


# sampler.chain is of shape (nwalkers, nsteps, ndim)
# throw-out the burn-in points and reshape:
emcee_trace  = sampler.chain[:, burn:, :].reshape(-1, ndim)
np.save("emcee_trace_w{nwalkers}_s{nsteps}.npy", emcee_trace)

print("done")

print(sampler.chain.shape) #original chain structure
print(emcee_trace.shape) #burned and flattened chain

import corner
import matplotlib.pyplot as plt
import xspec

# emcee_trace shape: (n_samples, len(par_val))
nparams = len(par_val)
assert emcee_trace.shape[1] == nparams, "Mismatch in number of parameters"

param_labels = [initial_model(par).name for par in par_val]

# Corner plot
figure = corner.corner(
    emcee_trace,
    labels=param_labels,
    quantiles=[0.16, 0.5, 0.84],
    show_titles=True,
    title_fmt=".4f",
    title_kwargs={"fontsize": 12}
)

plt.savefig('emcee_p_w{nwalkers}_s{nsteps}.png',dpi=300)
print("Congrats, job done.")
```