CrossMark

Testing the Performance and Accuracy of the RELXILL Model for the Relativistic X-Ray Reflection from Accretion Disks

Kishalay Choudhury[1], Javier A. García[2,3,4,7], James F. Steiner[5,8], and Cosimo Bambi[1,6,7]

[1] Center for Field Theory and Particle Physics and Department of Physics, Fudan University, 200433 Shanghai, China; bambi@fudan.edu.cn
[2] Cahill Center for Astronomy and Astrophysics, California Institute of Technology, Pasadena, CA 91125, USA
[3] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[4] Remeis Observatory & ECAP, Universität Erlangen-Nürnberg, Sternwartstr. 7, D-96049 Bamberg, Germany
[5] MIT Kavli Institute for Astrophysics and Space Research, MIT, 70 Vassar Street, Cambridge, MA 02139, USA
[6] Theoretical Astrophysics, Eberhard-Karls Universität Tübingen, D-72076 Tübingen, Germany

Abstract

The reflection spectroscopic model RELXILL is commonly implemented in studying relativistic X-ray reflection from accretion disks around black holes. We present a systematic study of the model's capability to constrain the dimensionless spin and ionization parameters from ~6000 *Nuclear Spectroscopic Telescope Array* (*NuSTAR*) simulations of a bright X-ray source employing the lamp-post geometry. We employ high-count spectra to show the limitations in the model without being confused with limitations in signal-to-noise. We find that both parameters are well-recovered at 90% confidence with improving constraints at higher reflection fraction, high spin, and low source height. We test spectra across a broad range—first at 10^6–10^7 and then ~10^5 total source counts across the effective 3–79 keV band of *NuSTAR*, and discover a strong dependence of the results on how fits are performed around the starting parameters, owing to the complexity of the model itself. A blind fit chosen over an approach that carries some estimates of the actual parameter values can lead to significantly worse recovery of model parameters. We further stress the importance to span the space of nonlinear-behaving parameters like $\log \xi$ carefully and thoroughly for the model to avoid misleading results. In light of selecting fitting procedures, we recall the necessity to pay attention to the choice of data binning and fit statistics used to test the goodness of fit by demonstrating the effect on the photon index Γ. We re-emphasize and implore the need to account for the detector resolution while binning X-ray data and using Poisson fit statistics instead while analyzing Poissonian data.

Key words: accretion, accretion disks – black hole physics – X-rays: general

1. Introduction

The modeling of the X-ray reflection spectrum is a very important method for understanding the physics of accreting compact objects. X-ray spectra from active galactic nuclei (AGNs) and X-ray binaries often show evidence of interaction between radiation emitted near the compact object and the nearby gas, which leads to signatures imprinted on the observed spectrum. Modeling the observed X-ray reflection features can lead to important constraints on the ionization state of the inner accretion disk (Ross et al. 1999; García & Kallman 2010; García et al. 2011, 2013). The reflecting region of the disk may be subject to relativistic effects that can blur and distort the emission features (Fabian et al. 1989; Laor 1991), leading to measurements of inner disk radii and black hole spin (Brenneman & Reynolds 2006; Reynolds & Fabian 2008). The most notable feature is the Fe-Kα emission complex (e.g., Guilbert & Rees 1988; Lightman & White 1988; Fabian et al. 1989; George & Fabian 1991). An astrophysical black hole in general relativity is completely specified by its angular momentum J and its mass M (Kerr 1963).[9] Black hole spin, defined by the dimensionless spin parameter $a_* = cJ/GM^2$ with theoretical values $|a_*| \leqslant 1.0$, is arguably the most important parameter whose estimate is

affected by strong field gravity near the black hole. One of the best-known AGN systems and the first to have observationally confirmed broad iron line detection is the Seyfert I galaxy MCG-6-30-15 (Tanaka et al. 1995; Iwasawa et al. 1999), for which the Fe emission appears to be broad and skewed well beyond the instrumental resolution. Such an X-ray reflection spectrum has been observed from accretion disks around numerous black holes (e.g., see Reynolds 2014).

Numerous reflection model computations have been published over the past two decades, with the most notable including PEXRAV (Magdziarz & Zdziarski 1995), REFLIONX (Ross & Fabian 2005), and XILLVER (García & Kallman 2010; García et al. 2013). These models were originally decoupled from the relativistic smearing associated with strong gravity and implemented in broadening kernels such as DISKLINE (Fabian et al. 1989), LAOR (Laor 1991, extended to KDBLUR later), KERRDISK (Brenneman & Reynolds 2006), KY (Dovčiak et al. 2004), and RELLINE (Dauser et al. 2010, 2013). These have been applied in a great many observational papers, (e.g., Miller et al. 2008; Steiner et al. 2011; Dauser et al. 2012; Fabian et al. 2012; Reynolds et al. 2012).

The X-ray blurring code RELXILL presented by García et al. (2014) is the current most advanced relativistic reflection model, which has addressed many of the deficiencies of previous models. It is the result of the angle-dependent reflection code XILLVER (García et al. 2013) convolved with the relativistic blurring code RELLINE (Dauser et al. 2013). XILLVER uses the atomic data of XSTAR (Bautista &

---

[7] Alexander von Humboldt Fellow.
[8] Einstein Fellow.
[9] Because of the difference between the masses of electrons and protons, astrophysical objects can have a nonvanishing equilibrium electric charge. However, for macroscopic objects, the value of the equilibrium electric charge is small, and completely negligible in the background metric.

Kallman 2001) to calculate the specific intensity of the radiation field as a function of energy, position in the accretion disk, and emission angle. Lamp-post geometry (Matt et al. 1991; Martocchia & Matt 1996; Frank et al. 2002; Dauser et al. 2013), which describes an isotropically irradiating primary X-ray source located on the rotation axis of the black hole has been implemented as relxilllp in the RELXILL model family. Given the progress made on reflection computations, and the complexity of the these calculations, it is important to test and understand how reliably, in an idealized case, the model performs in yielding estimates of spin and other quantities of interest.

The *Nuclear Spectroscopic Telescope Array* (*NuSTAR*; Harrison et al. 2013) mission is currently the best resource for the study of relativistic reflection. Its broad bandpass enables simultaneous measurements of both the iron emission and the Compton hump. Examples of recent, high-count reflection studies using *NuSTAR* data are spin measurements for the high-mass X-ray binary Cygnus X-1 (Walton et al. 2016) in the soft state, the low-mass X-ray binary (LMXB) GS 1354-645 (El-Batal et al. 2016) in the hard state, and for the LMXB GX 339-4 in the very high state along with *Swift*-XRT data (Parker et al. 2016). All works show that good constraints on spin can be achieved using *NuSTAR* for such bright sources, and their results agree well with previous observations. A more recent work with V404 Cygni has established the source's first ever spin constraint using *NuSTAR* (Walton et al. 2017b). Hard X-ray observations of accreting black holes provide a probe of the inner regions of the accretion disk where strong gravity prevails (Remillard & McClintock 2006; Bambi et al. 2016).

Using data with good signal and enough counts is a very common practice for simulations in order to test a model's efficiency. But one also needs to be cautious of systematic errors that may be introduced as a consequence of the ubiquitous customs. Humphrey et al. (2009) presented a thorough analysis and stated via Monte Carlo tests that performing fits employing χ^2-statistics at high counts can lead to heavy bias in contrast to the popular belief that the effect is minimal if a minimum-counts-per-bin sampling approach is taken up at high counts, which falls in line with the expectancy from Cash statistics (Cash 1979). The latter statistic should be employed for analyzing Poisson-distributed data and has been shown in their work to yield unbiased parameter estimates at high counts. As explained in the paper, the "approximated" χ^2_d test (followed as the *de facto* standard χ^2-test on XSPEC) can prove to underestimate values largely as counts in the spectrum increase, and unless the number of data bins in the data set are far less than $\sqrt{N}$, where N is the number of counts in the energy range considered in the fit, the bias with high-count data will not cease to exist irrespective of sampling with high counts per bin in a χ^2 test. In fact, the bias can be of the order of or even higher than the statistical error, and has been shown to increase as the number of counts increases. We will, however, show in the next section that the kind of binning we adopt avoids this problem for the total counts we work with.

In this paper, we aim to demonstrate how well RELXILL can constrain spin and related spectral parameters under optimal conditions. The work is presented as follows: Section 2 describes the model, software, and our methodology. Our results are given in Section 3. In Section 4, we also present a comparison of results from our fitting approach with that of a recent paper, and in Section 4.2 we further elucidate the role of

proper data analysis techniques, which are important when assessing subtle features in the reflection signal to infer physical constraints. We present our conclusions in Section 5.

2. Simulations as Proxy for Ideal Observational Data

Lamp-post geometry has been employed for explaining the observed X-ray spectra of numerous sources in the past (e.g., Duro et al. 2011; Wilkins & Fabian 2011; Dauser et al. 2012; Marin et al. 2012; Miller et al. 2015), and also in a few recent analyses (Beuchert et al. 2017; Walton et al. 2017a, 2017b). We use v0.4a[10] of the lamp-post model of RELXILL in this paper to carry out our analysis. Our main goal is to determine the robustness of the model, when employing ideally for the simplistic case of an axisymmetric, stationary source irradiating isotropically.

For this paper, we have simulated observational data using the fakeit routine in XSPEC (Arnaud 1996) v12.9.0i, implemented with the Python interface PyXspec (Arnaud 2016) v1.1.0. The simulations use the instrumental response of the FPMA of *NuSTAR* in the 3–79 keV energy band. Ancillary response and background files were selected assuming a circular extraction region with $60''$ radius centered $60''$ off-axis. All the instrumental files used here are available on the official website of *NuSTAR*.[11]

The resolution of the *NuSTAR* detectors is 0.4/0.9 keV at 6/60 keV. Based on this, we approximate the detector resolution as scaling as $E^{2/5}$ with a constant offset, and use the FTOOLS subroutine GRPPHA to bin our spectra so that the energy resolution is oversampled by a factor of 3. Approximately 90% of the grouped bins had >50 cts bin^{-1}, with the typical number of counts per bin even higher in harder input spectra compared to softer ones. Readers may refer to Kaastra & Bleeker (2016) for a more operationally crafty optimal binning method laid down for X-ray data analysis. This concerns both signal-to-noise and binning according to detector resolution. It adopts a variable binning scheme, thereby greatly reducing response matrix size and the number of model bins to save a lot of storage and computational time. This is, however, a much more sophisticated method than the one we have employed here. Nevertheless, we find that our approach serves well the main goals of the present paper.

To study the relativistic effects on incoming photons from the lamp-post source, we simulate our observations for a bright X-ray corona positioned at each of two lamp heights: $h = 3R_g$ and $h = 5R_g$. We simulated *NuSTAR* observations of a bright X-ray source (at $z = 0$) with a discrete sample space for all parameters of interest. Each observation generated was analogous to a 100 ks long exposure to ensure sufficiently high counts at all energies, and Poisson noise was included in the simulations. After grouping with our adopted method, we were left to fit 301 PHA bins in the desired energy range of 3–79 keV for each spectrum. This number is very small compared to the range of counts we work with here ($\sim 10^6$–10^7). Thus, we overcome the possibility of a fitting bias that using χ^2-statistics could have imposed as per Humphrey et al. (2009). We used nine representative values

---

[10] Newer versions of RELXILL were made available during the course of this project. But our results are still applicable because of our source constraints and methodology adopted.

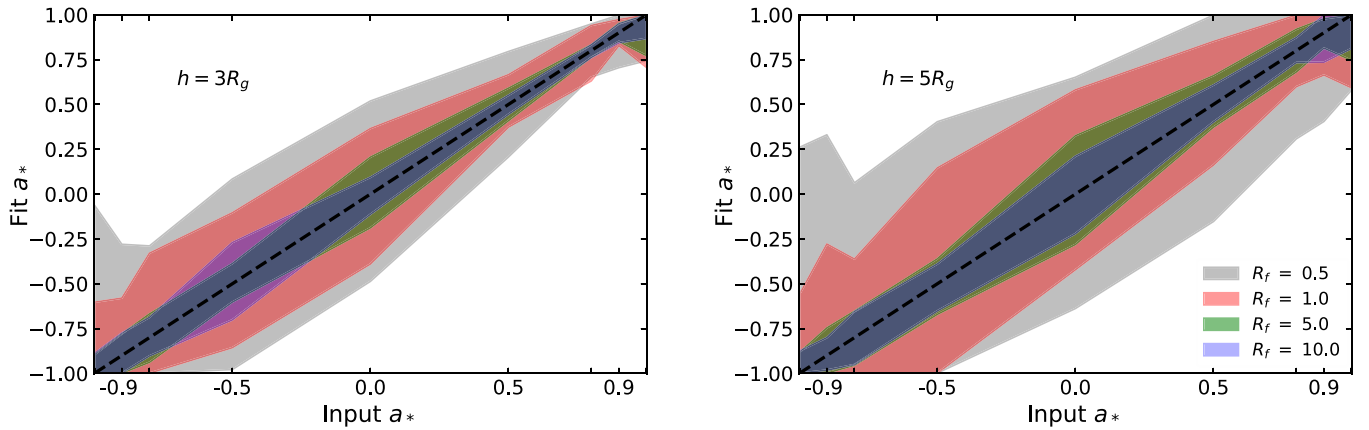[11] Point-source simulation files, http://www.nustar.caltech.edu/page/response_files.

Figure 1. Spectral fit results for our simulated data measuring spin a_* from all 5832 simulations of a bright X-ray source, each with an exposure of 100 ks. The fit parameters are summarized in Table 1. Each panel shows results for a different lamp-post height. The solid-colored intervals depict the 90% dispersion among the simulations for a given value of input parameter, centered around the mean. The typical fit statistic for each simulation shown has $\chi_\nu^2 \leqslant 1.01$. The dashed line depicts the simulation input value for spin.

Table 1
Parameter Values Used in `RELXILLLP` in Combination
with Nine Input a_* Values

Parameter	Input Values	Fit Parameter	Units
Source height (h)	3, 5	No	$R_g\left(=\frac{GM}{c^2}\right)$
Inclination (i)	15, 45, 75	Yes	deg
Inner disk radius (R_{in})	1	No	R_{ISCO}
Outer disk radius (R_{out})	400	No	R_g
Redshift (z)	0.0	No	...
Photon index (Γ)	1.4, 2.0, 2.6	Yes	...
Log ionization ($\log \xi$)	1.0, 2.3, 3.6	Yes	...
Iron abundance (A_{Fe})	1, 5, 10	Yes	solar
High-energy cutoff (E_{cut})	300	No	keV
Reflection fraction (R_f)	0.5, 1.0, 5.0, 10.0	Yes[a]	...

Note. Inner and outer disk radii, source redshift, and high-energy cutoff were kept at default values. Except source height, the remaining were used as fit parameters.
[a] We checked the estimates on R_f based on the dispersion in best-fit data. Not only are the error bars on R_f low, but also there is no significant difference in recovering all R_f values between both h. The 90% confidence is very narrow, and thus, R_f can be assumed to be a noncontributing fit parameter in terms of underlying physics rather than the statistics involved.

across the allowed parameter space for spin: ± 0.998, ± 0.9, ± 0.8, ± 0.5, and 0.0, and used values shown in Table 1 for the other `relxilllp` parameters. We explored all combinations in our sampling grid, creating ~5800 distinct simulations, representing a unique source condition in each. For the parameters Γ, $\log \xi$, and A_{Fe}, the input values have been chosen to represent the physical conditions in most observed systems, as mentioned in García et al. (2013). While the range for Γ in AGNs is generally narrower, that for binaries span a wider stretch with very low/hard $\Gamma \sim 1.4$–1.5 to very high/ soft $\Gamma \sim 2.5$–2.6 (e.g., Remillard & McClintock 2006). Similarly, AGN disks tend to be less hotter than those in binaries. Not only do spectra with ionization profiles with $\xi \sim 1$–20 appear similar, but also those with $\xi \gtrsim 10^4$ tend to be featureless (closely representing the incident power-law continuum). Iron abundances here are kept at solar and super-solar levels to have prominent Fe emission line features, also considering that a sub-solar abundance would not be enough to compensate against the stripping of most of the iron available at

higher ξ. Over-abundance of iron for explaining X-ray reflection spectra is not uncommon (e.g., Fabian 2006; Reynolds et al. 2012, etc.). The inner disk inclination (i) space has been selected to represent systems from being oriented almost face-on to having high relativistic smearing at an almost edge-on view, bearing in mind that the majority of the observed sources can be represented on average between the chosen extremes here.

We let the reflection fraction fit independently of the lamp-post-prescribed value (i.e., we set `fixReflFrac = 0` in `relxilllp`). The fitting procedure was designed such that the parameters in each fit initialized at random values that were required to lie beyond $X \pm 3\sigma_X$, assuming a Gaussian distribution based on the fit covariance around the true value X of each parameter, and where σ_X was estimated from preliminary fits. This approach was adopted to avoid fits becoming stuck in local minima, while at the same time allowing the model to explore the parameter space. We examine this procedure's effect on the overall results in Section 4.

3. Results

In Figure 1, we show the fitted versus inputted values of spin. Colored regions depict 1.645σ (90%) intervals showing the dispersion of the simulation fit dispersion about the average value for spin grid point. All simulations were fitted employing the χ^2-statistic. We present a parallel set of results showing the fitted values of the ionization parameter for the three input values from Table 1 in Figure 2. The reported χ_ν^2 is the reduced χ^2, where ν stands for the degrees of freedom in the fit.

Figure 1 shows that the input values are recovered in the fits on the whole. At a weak reflection signal of $R_f = 0.5$ and a corona located farther from ($h = 5R_g$) the compact object, the confidence, for example, on an intermediate spin $a_* = 0.5$ is weaker by a factor of 2 compared to when the corona is at low h. Larger reflection fraction and larger spin values both lead to tighter parameter constraints. Tighter constraints are likewise achieved at the lower source height. This can be understood as each effect separately produces an increase in the amount of relativistic reflection signal. We find that at sufficiently large values of R_f ($\gtrsim 5$), there is negligible change in the confidence interval. This is because the simulations depict a fixed number
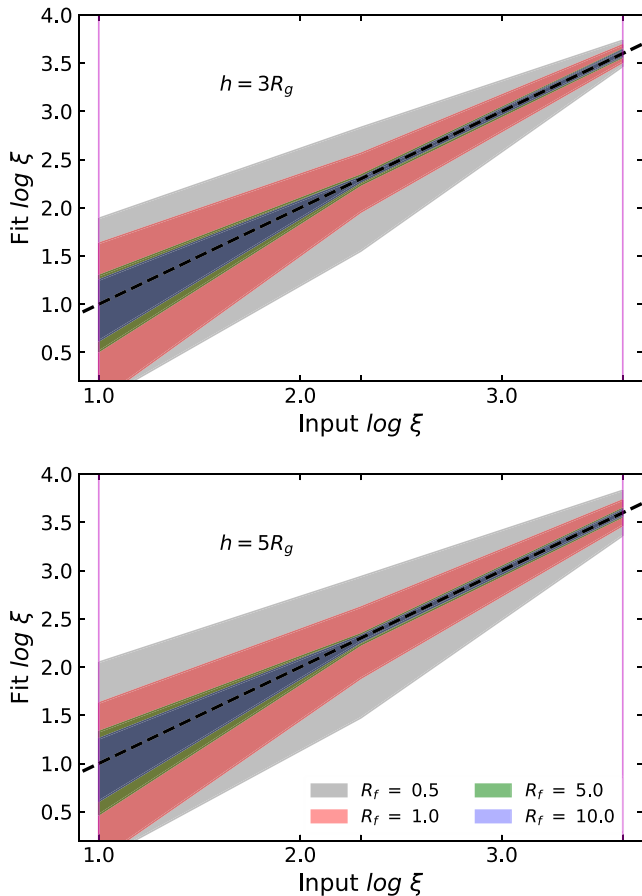
Figure 2. Spectral fit results for our simulated data measuring $\log\xi$ from all 5832 simulations of a bright X-ray source, each with an exposure of 100 ks. The fit parameters are summarized in Table 1 with nine values of a_*. Each panel shows results for a different lamp-post height. The solid-colored intervals depict the 90% dispersion among the simulations for a given value of input parameter, centered around the mean. The typical fit statistic for each simulation shown has $\chi^2_\nu \leqslant 1.01$. The dashed line depicts the simulation input value for ionization.

of counts, and the proportion of the signal that is conveyed in the reflection component is approximately $\frac{R_f}{R_f+1}$. For any large value of R_f, the reflection signal is essentially the same.

Similar conclusions can be drawn for the ionization parameter from examination of Figure 2. For instance, the ionization constraints between $h = 5R_g$ and $h = 3R_g$ degrade with a ratio of $\sim$1.1–1.7 (progressing from low to high ξ, in logarithm scale) at $R_f = 0.5$. It is interesting that higher dispersion seems to occur at low ionization ($\log\xi \lesssim 2$). This is explained by García et al. (2013), where it can be seen that the difference between reflection spectra in the *NuSTAR* band are quite minor at such low values of ionization. By contrast, at high ionization, the spectra are more readily distinguished, as reflected in the fits. It is, however, harder to recover a hotter disk at a higher source height at a weak reflection signal (wider relative confidence region for hotter disk) compared to when the primary source is closer because of a lesser reflection signal received from the inner disk as the point source moves higher up and away. In general, for the model parameter the dispersion intervals here, again, are slightly narrower for lower h, and they reduce appreciably with growing $\log\xi$ or increasing R_f.

4. Discussion

A similar investigation to ours was recently presented by Bonson & Gallo (2016; BG16 hereafter). BG16 analyzed over 4000 simulated spectra of AGN using RELXILL, with $3.5 \times 10^5 \pm 10^3$ net counts, each between 2.5 and 10 keV with *XMM Newton* (EPIC-pn; Strüder et al. 2001) and 10–70 keV with *NuSTAR*. They determined the effects on estimates of relativistic reflection parameters with Fe-line fitting in AGN X-ray astronomy and concluded that measuring such parameters accurately can be a great challenge, especially the spin (a_*) parameter, which they found can be decisively recovered for $a_* > 0.8$ with an accuracy of ± 0.1.

Like us, BG16 explored RELXILL behavior over a wide range of parameter space by simulating high-signal observations. However, unlike us, they initialized all their model fits at a single starting set of conditions (which could be arbitrarily far from or really close to the inputs). This type of fitting could be interpreted as what a particular user would do when there is no information regarding the parameters that characterize the system. On the contrary, our approach tries to simulate the other extreme possibility, in which the user has a good rough estimate of the parameters, based on previous analyses or measurements performed with other techniques. They report to have performed error checks on all fit parameters for one set as a verification step in the 2.5–70 keV band and that there were no significant differences found from their published results that were obtained only by running the fit. The effect of doing so without enforcing rigorous exploration of parameter space (e.g., via ERROR, STEPPAR, or MCMC approaches) can lead to erroneous results because the fit procedure can easily become trapped in a local minimum, which BG16 mention as well. Indeed, this shortcoming is not unique to BG16's approach, but is a caution that should be marked by observers to ensure that parameter space is explored intensively and sufficiently. In Section 2, we also used small step sizes (equal to or smaller than the predetermined σ_X) for the parameters of interest so that the fit can explore the parameter spaces better, although leading to the code running comparatively slower. We have attempted to elucidate the differences, with lamp-post geometry, between results employing our fitting approach and those from replicating the BG16 procedure as a representation of what is being done in day-to-day X-ray data analyses.

We chose to use our well-recovered set with $R_f = 5.0$ and $h = 3R_g$ for reference, and adopt all nine of our values of spin parameter. We simulated and fitted the same number of observations as in Section 2 for a single R_f and h combination with our source counts. The primary difference between this examination and BG16's is that we kept the input spin values discretized, whereas they selected values at random (but from the same range we adopt). When fitting, all parameters were initialized according to the BG16 approach (i.e., their "Test A" set: $a_* = 0.5$, $i = 30°$, $\Gamma = 2.0$, $\xi = 75\ \mathrm{erg\ cm\ s^{-1}}$, and $A_{\mathrm{Fe}} = 3$). We again binned the simulated data to three-times oversample the detector energy resolution. We ran errors on all fit parameters.

Figure 3 shows the results for recovering black hole spin with the representational fitting approach above. We show the $\chi^2_\nu \leqslant 1.1$ good fits separate from the color-coded scatter. We can clearly see here that only a minority ($\approx$35%) are actually good fits in the high-signal sample considered, and those that we further inspected are not stuck in the local minimum beyond a representational 10σ statistical significance over all fit
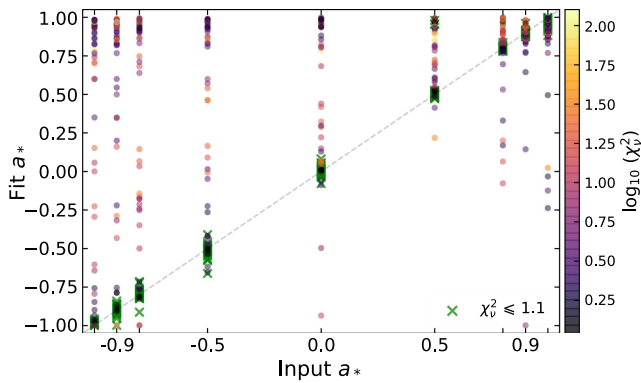
Figure 3. Spin fitting results following the approach of BG16 at our total counts. We show the fitted spin (a_*) vs. its input for model settings of $R_f = 5.0$ and $h = 3R_g$. Each data point is color-coded according to the logarithm of its reduced χ^2. Good fits with $\chi_\nu^2 \leqslant 1.1$ have been shown separately to compare with BG16's work.

parameters. Here, error checks push to span the parameter space better and are able to converge to good best-fit statistics. However, evidently, there is a much larger scatter of fits with $\chi_\nu^2 > 1.1$ and many spurious results appear to muddy the determination of a_*. We find that it is much more likely for the fit to miss entirely and when it does to overestimate the spin. These are the larger share of bad fits stuck in local minima, and can be compared for a contrast with results from our fitting approach with similar parameter combination in Figure 1. Figure 4 shows the scatter in the other parameters, and we can see that both Γ and $\log \xi$ have large scatter similar to that shown by BG16. The good fits show no under/overestimation in the photon index, the iron abundance and the inner disk inclination parameters. We can also see a few good fits with input $\log \xi < 2.0$ being underestimated, but it can be explained because of the spectral similarity at low ionization mentioned in Section 3.

As apparent in Figure 4, and confirmed by exploration of the data, the primary problem in question with the fits in local minima arises due to incorrect measurements of $\log \xi$ pushing spin toward poor estimates. The nonlinear behavior exhibited by ionization can well be a problem in effectively analyzing data without a sound fitting procedure. To test this hypothesis, we performed a qualitative examination on all four parameters shown in Figure 4. We excluded all fits $\leqslant |2\sigma|$ confidence around an ideal recovery (i.e., fit = input) for each input value for a_*, and mapped the remainder (outliers all) to the Figure 4 fit parameters. The results confirmed that the worst spin fits were associated with very large mismatches in the ionization parameter. Iron abundance, followed by Γ and i, were minimally associated with causing outliers in spin. This is likely because the spectral changes associated with evolving ionization are not necessarily smooth and continuous, unlike, e.g., changes associated with i or A_{Fe} (see, e.g., García et al. 2013).

4.1. Line and Continuum Features Trading Off

Having noticed that Γ's distribution of spurious fits in Figure 4 is strongly skewed to fitting larger values of Γ, and revealing surprisingly large discrepancies up to $\gtrsim 0.1$ (far larger than statistical errors), we explore the relationship between Γ and the nonlinear parameter $\log \xi$ from our results with the representational fitting approach.

To examine this behavior, in Figure 5, we have color-coded the relationships of $\log \xi$ and Γ from Figure 4. Blue depicts cases with $\Delta\Gamma > 0.01$, green shows $\Delta\Gamma < -0.01$, and red gives the remaining cases (all good fits) in which the fit is close to the input value. Note that fits that greatly underestimate $\log \xi$ correspond to fits that overestimate Γ, and vice versa.

This correlation can be understood again from the reflection principles outlined in García et al. (2013). Low values of $\log \xi$ emit less at soft X-rays and so produce harder X-ray spectra. In order for the fit to compensate for the harder signal in the reflection, Γ is increased. In this way, there is correlated trade-off between the parameters describing narrow reflection features (i.e., $\log \xi$), and the X-ray continuum (Γ).

In order to see how the results fare at lower counts, we reproduced the dispersion shown in Figures 1 and 2 for the set with $R_f = 5.0$ and $h = 3R_g$, the scatter in Figures 3 and 4, and the mapping in Figure 5—all at BG16's total counts. As expected, the dispersion in spin and ionization increased at all input values of a_* and $\log \xi$ with our fitting approach, although not deviating anywhere from the input line but having noticeably wider confidence at lower input values. The trend in the scatter plots with the representational fitting approach at lower counts was similar to those shown in this work at higher counts. The percentage of good fits not stuck in local minima increased (to $\approx 48\%$) but had much larger deviations from the input lines. We observed that most of the good fits sticking around the input line in this case could be attributed to well-recovered i and Γ, while $\log \xi$ again remained the primary culprit in creating spurious results with a higher number of good ξ fits being underestimated. This ascertains the need to adopt a proper fitting approach when testing model parameters at any range of total counts.

4.2. Bias, Binning, and Statistical Methods

As seen from Section 4.1, Γ seems to be overestimated in the case of assessing blurred reflection using RELXILL over a broad range of counts with the representational fitting approach. While we described above the way in which an offset in $\log \xi$ can introduce a correlated shift in the continuum Γ, this effect (also reported by BG16 for Γ at $\chi_\nu^2 \leqslant 1.1$) is more subtle and is rooted in the data treatment because in our case it shows up, for both ranges of counts, only at $\chi_\nu^2 > 1.1$. In this section, we illustrate how the use of χ^2 rather than Poisson statistics can fully account for this reported bias.

To demonstrate the problem, we have produced 1000 simulations using a pure power-law model for the same *NuSTAR* FPMA response to look into the photon index independently, with randomly sampled values of Γ ranging between 1.4 and 2.6. Each spectrum contains 3.5×10^5 counts (within Poisson limits) across the useful 3–79 keV band.

These simple simulations were then binned using two approaches: (1) the data were binned according to our default procedure, oversampling the detector resolution by a factor of 3, or (2) the data were grouped to achieve a minimum of 25 counts per bin, and the detector resolution was not taken into consideration. We also test the importance of the Gaussian approximation inherent in χ^2 fitting, in both cases, by comparing results achieved by χ^2 minimization with parallel results achieved using the proper Cash statistic (Cash 1979; termed "cstat" in XSPEC), which is appropriate for Poisson-distributed data.

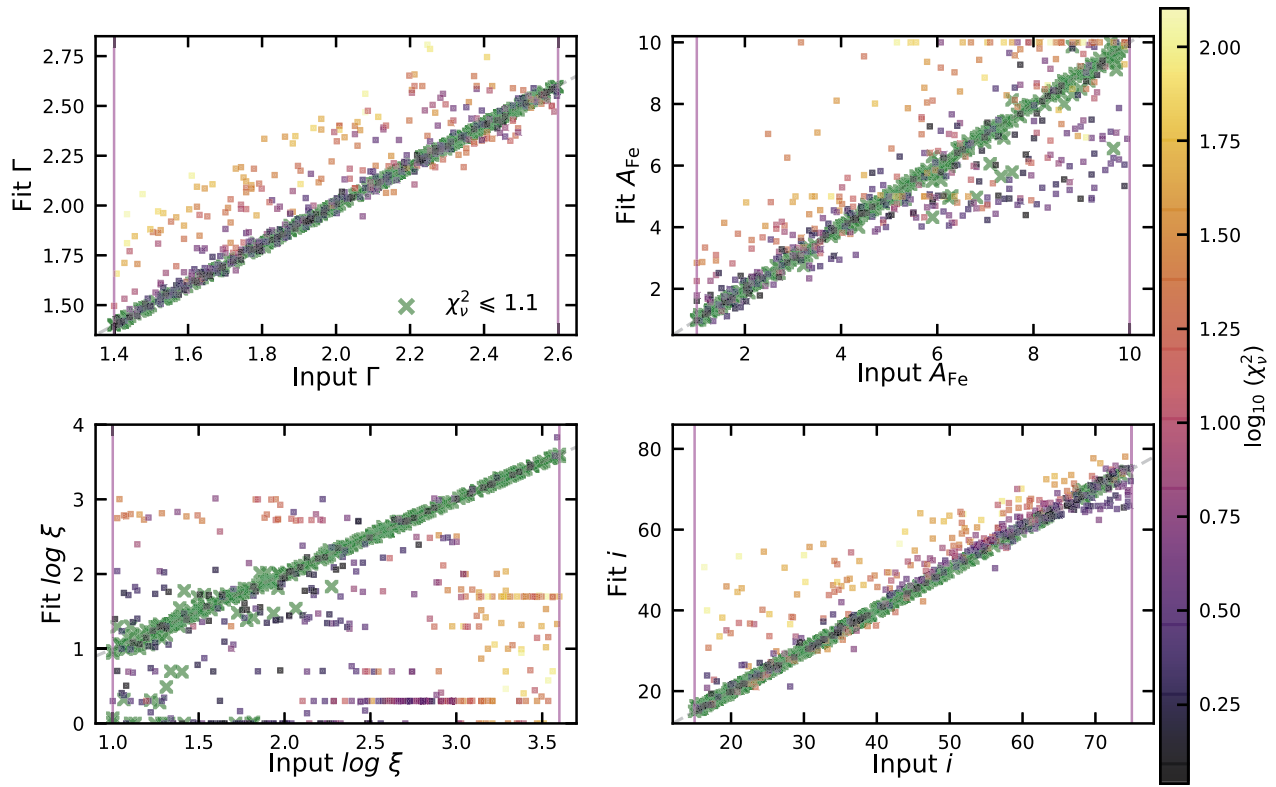Figure 6 illustrates our results. The leftmost pair of panels depicts two sets of χ^2 fits using the two binning options,

Figure 4. Results using the approach of BG16, as in Figure 3, but showing Γ, A_{Fe}, $\log \xi$, and i. Although the color-coding indicates overestimation in Γ and i and underestimation of $\log \xi$, we find that none of the good fits of these parameters, other than a few $\log \xi_{inp} = 2.0$ underestimations, are under/overestimated at our signal and with our grouping approach.
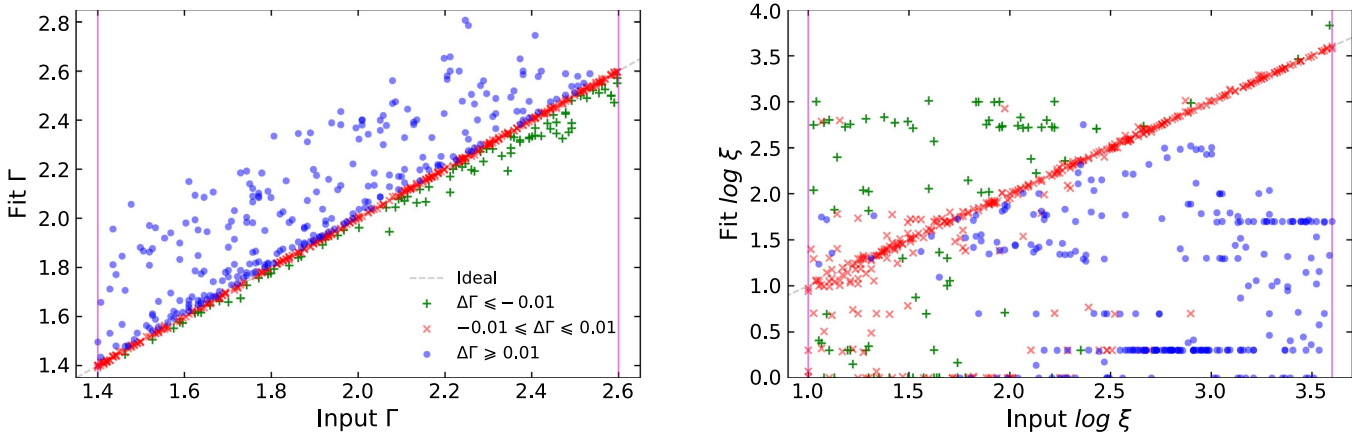

Figure 5. Left upper and lower panels of Figure 4 color-coded based on the difference $\Delta\Gamma \equiv \Gamma_{fit} - \Gamma_{input}$. There is clear correspondence between the differences in Γ, and in $\log \xi$.

presenting the difference between fitted and inputted Γ. A strong bias in Γ is clearly evident in the "minimum binning" panel, which relates to the aforementioned binning option (2). In this instance, one infers that Γ is generally too high, by a factor of ~0.005. The precise value of this offset turns out to be a function of the binning and also of the total signal in the spectrum, with the bias becoming worse when the signal is sparser and the counts per bin fewer. As can be seen, nearly all fits in the minimum binning case have good converging statistics. However, the majority appear to be stuck in local minima, largely showing overestimate in Γ owing to the sampling performed. Importantly, our adopted resolution binning appears comparatively immune to this bias and is clearly preferable.

However, the rightmost pair of panels show that when taking a more careful look, as revealed here by scaling the difference in Γ by the statistical error, even the resolution-binning case suffers from a small bias when employing the approximate (but most widely used) χ^2-statistic compared to the furthest right panel, which shows the same using the Cash statistic. Note the Cash-statistic distribution is centered on zero with good converging statistics, while the χ^2 case is very slightly offset to larger fitted Γ (a minor shift of order ~0.2σ). Given the very minor scale of the bias associated with our adopted approach, our earlier results from Section 3 are quite acceptable.

We further illustrate the dependence of the bias in Γ on the total counts and degree of binning in Figure 7. Here, we select a
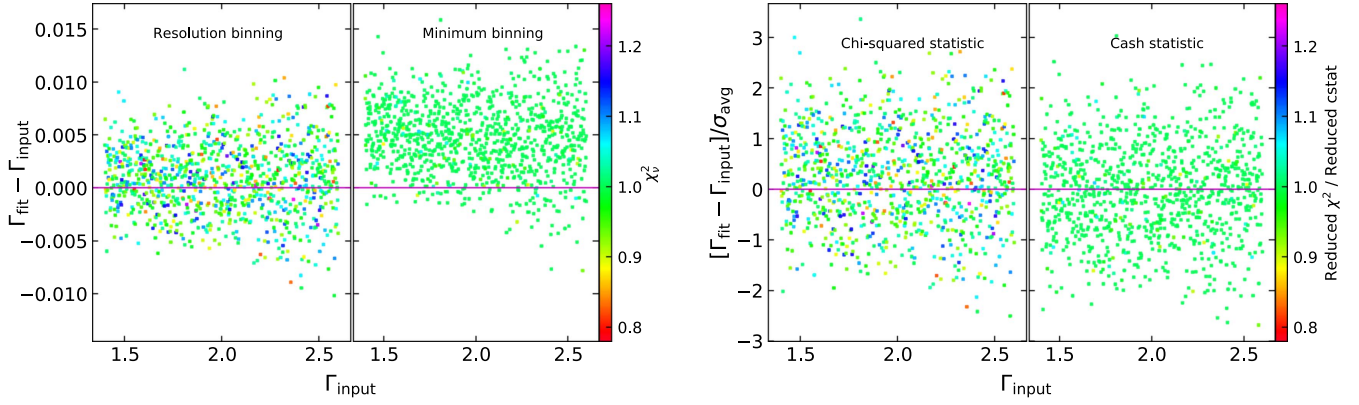
Figure 6. Color-coded fits from 1000 Seyfert 1 spectra simulated with the POWERLAW model with *NuSTAR*. Left: scatter plot for the difference $\Gamma_{\rm fit} - \Gamma_{\rm input}$ ($=\Delta\Gamma$) for two sets, one with data oversampled according to our resolution-binning method (left pane), and the other with a minimum grouping of 25 cts bin^{-1} (right pane). Right: scatter plot for the difference $\Delta\Gamma$ in terms of σ for the same set of resolution-binned data, fit one time each using χ^2 statistic and cstat. $\sigma_{\rm avg}$ refers to the average statistical error bar for each point in each scatter plot. The scatter $>|2\sigma|$ is negligible. The color-bar represents χ^2_ν and reduced-cstat.
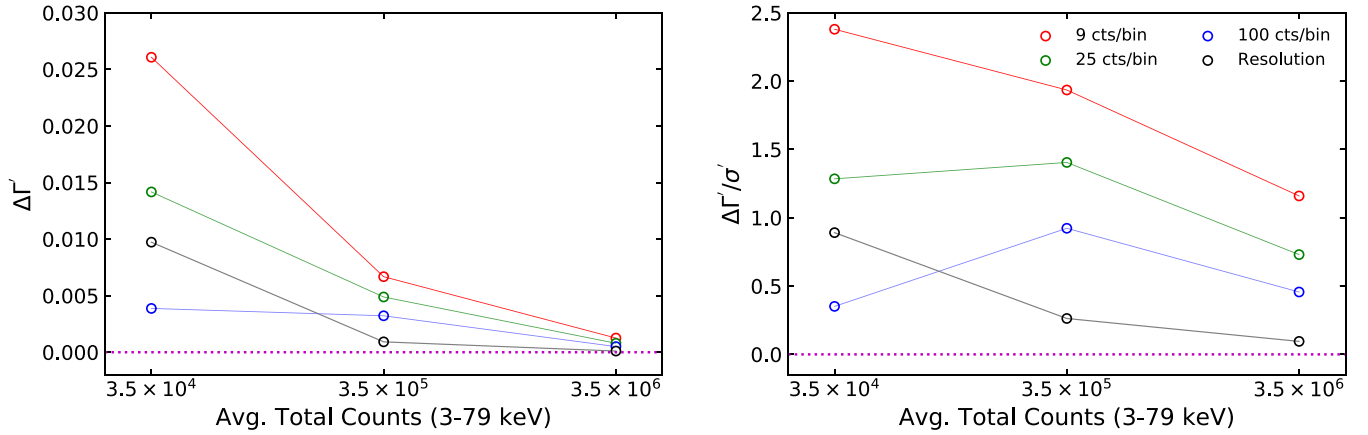

Figure 7. Results from pure power-law simulations illustrating the effect of varying the total signal and the grouping. Each point is obtained from a χ^2-minimizing fit, and the results are compared to a reference analysis using the Cash statistic. The left panel shows the absolute offset in Γ ($\Delta\Gamma' \equiv \Gamma_{\rm (fit,\chi^2)} - \Gamma_{\rm (fit,cstat)}$), and the right panel depicts the bias in terms of its statistical significance.

single representative value of $\Gamma = 2.0$ as a fiducial value and have varied the counts per spectrum by an order of magnitude in either direction in a smaller set of simulations that depict different data groupings, where the minimum number of counts per bin is 9, 25, or 100. Using a fit with the Cash statistic as our reference point, we show the resulting bias in Γ in the left panels and compare that to its significance in the right panels. We present the results for our adopted resolution-binning as well.

This minimum binning approach is commonly adopted for the analyses of observational X-ray data in order to have ample counts in each data bin, and overlooks the detector resolution. The overestimation found in Γ can be evidently seen due to this. Apart from insufficient binning of data, using the Gaussian-assuming χ^2-statistic to model Poisson-distributed data could also be a cause. Accordingly, we recommend using the Cash statistic when possible, and suggest further that it is advisable to bin the data according to the detector resolution (refer to page 2 for references on the choice of statistics and data binning).

5. Conclusions

We have simulated *NuSTAR* data of a bright X-ray source to determine how well RELXILL can practically determine spin and ionization parameters. We have adopted the simplest case of lamp-post geometry. Our results are summarized below.

We find that all model parameters are well-recovered in the fits to our reflection simulations, and that the precision is improved at higher R_f, higher spin, and lower h, because each increases the signal-to-noise in relativistic reflection features. Recovering retrograde spin is more difficult compared to their prograde counterparts owing to the lower reflection signal received because of the position of the inner disk radius $R_{\rm in}$, which we have fixed at the innermost stable circular orbit (ISCO). $R_{\rm ISCO}$ for a counter-rotating black hole tends to increase outwards with more retrograde values, thereby limiting reflection from the inner regions of the disk. Further complications may arise because of lowered relativistic bending at higher source heights.

We carried out an examination of adopting fitting methods while analyzing X-ray data based on results from a similar work done by Bonson & Gallo (2016), across our selected 10^6–10^7 range of total counts for testing the estimates in *NuSTAR*'s effective energy range. BG16 show that the model RELXILL yields poor estimates for the relativistic and disk parameters when a representational fitting approach like they adopt is employed. We also performed the examination at lower ($\sim 10^5$) counts and, to our expectations, obtained similar results. We find that the intrinsic bias imposed by such a fitting approach, extended to any range of detected counts, can be minimized by instead adopting to fit data with some

predetermined estimates around the true parameter configuration, owing to the complexity of the model itself. This points to the efficiency of the fitting method we adopt in Section 2, and also stands to support the concern expressed by BG16 over such common X-ray fitting methods. Furthermore, parameters like $\log \xi$, which has a nonlinear impact on spectra, require thorough stepping through their parameter spaces in order to avoid internal subtle adjustments among parameters that can easily make the user believe in misleading good fits.

On top of the fitting method chosen, picking the data binning scheme also seems to impact results when using the Gaussian-assuming χ^2-statistic, which can be seen from the affect on the global parameter Γ in Section 4.2. In the case of data being Poisson-distributed, it is crucial to employ a sound binning methodology. Then again, using the Cash statistic is preferably the better choice as it is suitable for treating Poissonian data. We therefore advise that data should be binned according to the detector resolution and the Cash statistic employed when possible (ensuring at least 1 count per bin at very faint detections when the noise is purely Poissonian).

The results put forth in this work have been determined in the most ideal case: using one module and the broad bandpass of *NuSTAR* with Poissonian background, devoid of effects like interstellar medium absorption <2 keV, to assess blurred reflection data using lamp-post irradiation. While the results depict the efficiency of RELXILL in being able to constrain spin and ionization very well, the conditions can be conservative not only because of the geometry adopted but also in the added sense that there is no unique "correct" method when adopting a fitting approach in order to probe a model's efficacy. In addition to the selection of a sound fitting methodology, the extent to which the user can avoid local minima relies on the fitting algorithm, data sampling, and the degrees of freedom in the fit that may complicate the situation altogether. The user, however, needs to be cautious since results obtained may in fact shadow the true nature of parameter constraints and degeneracies (which we have not probed into for the current work) involved intrinsically, a common problem while analyzing real data.

Simultaneous fitting with simulations from instruments operating at lower energies, like *XMM Newton* or *Suzaku*, can significantly improve our constraints, and may serve as an extension to the current work. Similar work in García et al. (2015) showed that the constraints in $E_{\rm cut}$ in RELXILL improved with the use of a broader waveband encompassing soft energies. An alternative in the case of data analyzed from observations could be "pgstat," which reads Gaussian background with Poissonian detection (see Appendix B in XSPEC manual). The implementation effect of this fit statistic can be tested with real data or a simulated Gaussian noise since background can definitely be not just Poisson-distributed in reality. But this is beyond the goals and scope of this work.

ORCID iDs

Javier A. García ⓘ https://orcid.org/0000-0003-3828-2448
James F. Steiner ⓘ https://orcid.org/0000-0002-5872-6061
Cosimo Bambi ⓘ https://orcid.org/0000-0002-3180-9502

References

Arnaud, K. A. 1996, in ASP Conf. Ser. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes (San Francisco, CA: ASP), 17
Arnaud, K. A. 2016, in AAS/High Energy Astrophysics Division 15, poster, 115.02
Bambi, C., Jiang, J., & Steiner, J. F. 2016, CQGra, 33, 064001
Bautista, M. A., & Kallman, T. R. 2001, ApJS, 134, 139
Beuchert, T., Markowitz, A. G., Dauser, T., et al. 2017, arXiv:1703.10856
Bonson, K., & Gallo, L. C. 2016, MNRAS, 458, 1927
Brenneman, L. W., & Reynolds, C. S. 2006, ApJ, 652, 1028
Cash, W. 1979, ApJ, 228, 939
Dauser, T., Garcia, J., Wilms, J., et al. 2013, MNRAS, 430, 1694
Dauser, T., Svoboda, J., Schartel, N., et al. 2012, MNRAS, 422, 1914
Dauser, T., Wilms, J., Reynolds, C. S., & Brenneman, L. W. 2010, MNRAS, 409, 1534
Dovčiak, M., Karas, V., & Yaqoob, T. 2004, ApJS, 153, 205
Duro, R., Dauser, T., Wilms, J., et al. 2011, A&A, 533, L3
El-Batal, A. M., Miller, J. M., Reynolds, M. T., et al. 2016, ApJL, 826, L12
Fabian, A. C. 2006, AN, 327, 943
Fabian, A. C., Rees, M. J., Stella, L., & White, N. E. 1989, MNRAS, 238, 729
Fabian, A. C., Zoghbi, A., Wilkins, D., et al. 2012, MNRAS, 419, 116
Frank, J., King, A., & Raine, D. J. 2002, Accretion Power in Astrophysics (3rd ed.; Cambridge: Cambridge Univ. Press), 398
García, J., Dauser, T., Lohfink, A., et al. 2014, ApJ, 782, 76
García, J., Dauser, T., Reynolds, C. S., et al. 2013, ApJ, 768, 146
García, J., & Kallman, T. R. 2010, ApJ, 718, 695
García, J., Kallman, T. R., & Mushotzky, R. F. 2011, ApJ, 731, 131
García, J. A., Dauser, T., Steiner, J. F., et al. 2015, ApJL, 808, L37
George, I. M., & Fabian, A. C. 1991, MNRAS, 249, 352
Guilbert, P. W., & Rees, M. J. 1988, MNRAS, 233, 475
Harrison, F. A., Craig, W. W., Christensen, F. E., et al. 2013, ApJ, 770, 103
Humphrey, P. J., Liu, W., & Buote, D. A. 2009, ApJ, 693, 822
Iwasawa, K., Fabian, A. C., Young, A. J., Inoue, H., & Matsumoto, C. 1999, MNRAS, 306, L19
Kaastra, J. S., & Bleeker, J. A. M. 2016, A&A, 587, A151
Kerr, R. P. 1963, PhRvL, 11, 237
Laor, A. 1991, ApJ, 376, 90
Lightman, A. P., & White, T. R. 1988, ApJ, 335, 57
Magdziarz, P., & Zdziarski, A. A. 1995, MNRAS, 273, 837
Marin, F., Goosmann, R. W., Dovčiak, M., et al. 2012, MNRAS, 426, L101
Martocchia, A., & Matt, G. 1996, MNRAS, 282, L53
Matt, G., Perola, G. C., & Piro, L. 1991, A&A, 247, 25
Miller, J. M., Tomsick, J. A., Bachetti, M., et al. 2015, ApJL, 799, L6
Miller, L., Turner, T. J., & Reeves, J. N. 2008, A&A, 483, 437
Parker, M. L., Tomsick, J. A., Kennea, J. A., et al. 2016, ApJL, 821, L6
Remillard, R. A., & McClintock, J. E. 2006, ARA&A, 44, 49
Reynolds, C. S. 2014, SSRv, 183, 277
Reynolds, C. S., Brenneman, L. W., Lohfink, A. M., et al. 2012, ApJ, 755, 88
Reynolds, C. S., & Fabian, A. C. 2008, ApJ, 675, 1048
Ross, R. R., & Fabian, A. C. 2005, MNRAS, 358, 211
Ross, R. R., Fabian, A. C., & Young, A. J. 1999, MNRAS, 306, 461
Steiner, J. F., McClintock, J. E., Reis, R. C., Gou, L., & Remillard, R. A. 2011, BAAS, 43, 404.02
Strüder, L., Briel, U., Dennerl, K., et al. 2001, A&A, 365, L18
Tanaka, Y., Nandra, K., Fabian, A. C., et al. 1995, Natur, 375, 659
Walton, D. J., Brightman, M., Risaliti, G., et al. 2017a, arXiv:1706.02088
Walton, D. J., Mooley, K., King, A. L., et al. 2017b, ApJ, 839, 110
Walton, D. J., Tomsick, J. A., Madsen, K. K., et al. 2016, ApJ, 826, 87
Wilkins, D. R., & Fabian, A. C. 2011, MNRAS, 414, 1269