**disnht: modeling X-ray absorption from distributed column densities**

Nicola Locatelli\(^1\), Gabriele Ponti\(^2\), Stefano Bianchi\(^3\)

\(^1\) Max-Planck-Institut für Extraterrestrische Physik (MPE), Giessenbachstrasse 1, 85748 Garching bei München, Germany
\(^2\) INAF - Osservatorio Astronomico di Brera, via E. Bianchi 46, 23807 Merate (LC), Italy
\(^3\) Dipartimento di Matematica e Fisica, Università degli Studi Roma Tre, via della Vasca Navale 84, I-00146 Roma Italy

Accepted ???. Received ??; in original form ??

**ABSTRACT**

Collecting and analysing X-ray photons over either spatial or temporal scales encompassing varying optical depth values requires knowledge about the optical depth distribution. In the case of sufficiently broad optical depth distribution, assuming a single column density value leads to a misleading interpretation of the source emission properties, nominally its spectral model. We present a model description for the intermediate column absorption in X-ray spectra at moderate energy resolution, extracted over spatial or temporal regions encompassing a set of independent column densities. The absorption model (named disnht) approximates the distribution with a lognormal one and is presented in table format. The solution table and source code are made available and can be further generalized or tailored for arbitrary optical depth distributions encompassed by the extraction region. The disnht absorption model presented and its generalized solution are expected to be relevant for present and upcoming large angular scale analyses of diffuse X-ray emission, such as the ones from the extended ROentgen Survey with an Imaging Telescope Array (eROSITA) and the future Athena missions.

**Introduction**

Neutral and partially ionized material can efficiently absorb ultra-violet light and X-rays (Wilms et al. 2000). Absorption is observed as an exponential decrease in the received intensity of the light $I \propto e^{-\tau}$, quantified through the optical depth $\tau = \sigma N_H$, where $N_H$ is the integral of its neutral hydrogen along the line of sight in cm\(^{-2}\). The abundances of all the other metals contributing to the absorption in the intervening medium are also taken into account through the absorption cross section $\sigma(X(Z), E)$, in cm\(^2\), as a function of the photon energy $E = h\nu$, where $X(Z)$ maps the relative abundance of the element $Z$ with respect to the hydrogen column density $N_H$ (Rybicky et al. 1982). The interstellar medium (ISM) hosts large amounts of such an absorbing material in the form of a complex multiphase collection of ions, atoms, molecules and dust grains diffused all across the thin stellar disk of our Galaxy, encompassing a wide range of densities ($10^4 < n_{\text{ISM}} < 10^6$ cm\(^{-3}\)) and temperatures ($T_{\text{ISM}} < 10^5 K$) and contributing differently to the X-ray absorption (e.g. Wolfire et al. 2003; Ferrière 2001, for a review). When analysing the light spectrum along one line of sight the absorption can be easily detected and accounted for, thanks to its known characteristic energy dependence (namely well known cross sections, Balucinska-Church & McCammon 1992). However, if the light collected into an instrument and summed into a spectrum encompasses several independent $N_H$ elements (in space or time), an appropriate description of the absorption necessarily requires a description of the $N_H$ distribution of values, as a single column density value describing the absorption within a given (and satisfactory) accuracy may simply not exist. This fact follows from the mathematical evidence that a sum over exponential functions can be approximated by a single exponential function only in the case of a very peaked distribution of exponents around one value. An incomplete but instructive list of cases in which this assumption is broken include the spectral analysis of: faint diffuse emission for which collection of photons over a large angular scale is performed to increased the significance of the spectral features (e.g. Munu et al. 2009; Masui et al. 2009; Yoshino et al. 2009; Gupta et al. 2012; Ponti et al. 2015; Cheng et al. 2021; Wang et al. 2021); extraction regions holding high ($\sim 10^{24}$ cm\(^{-2}\)) and very high ($\sim 10^{25}$ cm\(^{-2}\)) absorption such as toward dense clouds in the Galactic disc, for which collecting the photons over even a few resolution elements can significantly affect the X-ray spectrum (e.g. Molinari et al. 2011); pointlike sources holding column densities variable over the observation timescale, such as for dipping low mass X-ray binaries binaries (Díaz Trigo et al. 2006; Ponti et al. 2016; Bianchi et al. 2017), changing-look active galactic nuclei (Matt et al. 2003; Puccetti et al. 2007; Bianchi et al. 2009; Risaliti et al. 2009; Marchese et al. 2012; LaMassa et al. 2015) and magnetic accreting white dwarfs (Done & Magdziarz 1998). Modeling of the spectral absorption through distribution of column densities has already successfully performed in the past using a power law column density distribution (pwab model, presented in Done & Magdziarz 1998) and more recently using a lognormal distribution (Cheng et al. 2021; Wang et al. 2021). These examples show the increasing need in the X-ray community of careful and systematic treatment of the absorption when its properties change in a non-negligible way across space or time. For these and all the other cases mentioned above, in this work we aim to provide the community with a method and precomputed tables of absorption factors $-\ln A_\odot(E)$ able to properly reproduce the absorption curve. The model produced will be useful in general to any X-ray spectral analysis that involves a set of independent column densities broadly distributed around an expected value.

Send offprint requests to: E-mail: nlocat@mpe.mpg.de
The DIStributed NH Table model (disnht) is written into an ascii readable table and is already implemented in fits format readable from the X-ray SPECTral fitting software XSPEC as an exponential table model (etable). All the tables and materials used to create them, including the source code are publicly available 1. The code can be used to compute the absorption spectrum for any arbitrary column density distribution. In Section 2 we describe the model creation and output format and usage. We present different realistic scenarios in Sec. 3 and we draw our conclusions in Sec. 4.

2. Model

For any given set of M column densities $\Theta_i(N_H)$, the exact overall absorption spectrum can be computed as

$$A_0 = \sum_i \exp\left[-\sigma(X(Z), E)N_{H,i}\right]$$

where the cross section $\sigma$ is derived from Balucinska-Church & McCammon (1992) and assuming Lodders 2003 as our benchmark model for the metals in the ISM. Cross sections from different abundance sets commonly used in the XSPEC software are also provided for completeness (Anders & Grevesse 1989; Asplund et al. 2009).

In the lack of an exact model for the absorption by a distribution of column densities, it is common to simply assume an average value $\langle N_H \rangle$ or to fit for a single value, so that the absorption spectrum is

$$A_0 = \exp\left[-\sigma(X(Z), E)\langle N_H \rangle\right]$$

The discrepancy introduced in $A_0$ with respect to the exact absorption spectrum $A_0$ can be significant. We show an example in Fig. 1, where we accumulate the emission from a sample of absorbed spectra $p_{\text{abs}}(1\sim M)$ with $F_{\nu} \propto \nu^{-2}$ and fixed (arbitrary) normalization. We assume the distribution of column density $\Theta_i(N_H)$ to be lognormal with expected value $\text{avg}_{\Theta} = 21, 22, 23, 24 \text{dex}$ and standard deviation of $\text{std}_{\Theta} = 0.01, 0.2, 0.3, 0.5 \text{dex}$. Every spectrum $p_{\text{abs}}$ is absorbed by a column density in the set $\Theta_i$ opacity equal to the one used by the tbabs model (Wilms et al. 2000). The absorbed spectra are then summed together. The sum spectrum $A_0$ (Eq. 1) is then compared with a single spectrum absorbed by the average value of $\Theta_i$, $\bar{A}_0$ (Eq. 2). We note that assuming a shape for the column density distribution $\Theta_i$ is necessary to make the data fit to converge. An arbitrary distribution can instead be modeled whenever precisely know, regardless of its shape. We find that a lognormal distribution for the column density is a reasonable assumption in real cases where the discrepancy between the exact and average-value solutions for the absorbed spectrum becomes relevant (see Sec. 3 for further details). The lognormal distribution is parameterized by its expected value $\text{avg}_{\Theta}$ and standard deviation $\text{std}_{\Theta}$.

The discrepancy between the exact and average column density (blue dashed line in Fig. 1) solutions can be computed in terms of the ratio $A_0/\bar{A}_0$. In the example shown in the upper left panel $A_0/\bar{A}_0$ is in excess of 1% at $E \leq 1 \text{ keV}$, amounting to 10% at 0.3 keV. The discrepancy increases exponentially at lower energies, so that it is already a few orders of magnitudes at 0.1 keV. This situation worsens for all values of $\text{avg}_{\Theta} > 21.0$ and $\text{std}_{\Theta} > 0.01 \text{ dex}$.

1 http://www.brera.inaf.it/hotmilk/models.html

| comp | param | tbabs | disnht |
|------|-------|-------|--------|
| absorption | $N_H, \text{avg}_{\Theta}$ | 3.29 $\pm$ 0.06 $\times 10^{22}$ | 23.27 $\pm$ 0.03 |
| absorption | $\text{std}_{\Theta}$ | - | 0.340 $\pm$ 0.002 |
| APEC | $kT$ | 1.77 $\pm$ 0.04 | 1.04 $\pm$ 0.02 |
| APEC | $Z/Z_\odot$ | 1.7 $\pm$ 0.2 | 1.13 $\pm$ 0.15 |
| APEC | norm | 8.7 $\pm$ 0.7 $\times 10^{-4}$ | 8.1 $\pm$ 1.1 $\times 10^{-3}$ |

Table 1. Best fit parameters from the spectral fit of the mock data. d.o.f. degrees of freedom.

We compute and tabulate the correct absorption $A_0$ for all values of $\text{avg}_{\Theta} \in [19.5, 25.5]$ and $\text{std}_{\Theta} \in [0, 1]$. The table is multiplicative so that entries, as a function of energy, are to be multiplied to any unabsorbed spectrum. The XSPEC table provides instead the quantity $-\ln A_0$ so that, for instance, for $A_0 = e^{-2}$ the XSPEC entry provides the value 2. The XSPEC class of this table is an etable and it also can be used as a multiplicative model when loaded through the command etable{<TABLE_NAME>}.

3. Discussion

We demonstrate the efficacy of the disnht model in retrieving the correct values by simulating a set of 100 lines of sight encompassing a distribution of column densities $N_H$. The $N_H$ distribution $\Theta$ is drawn from a lognormal with mean value $\text{avg}_{\Theta} = 23.30$ and standard deviation $\text{std}_{\Theta} = 0.35$. The $\text{avg}_{\Theta}$ and $\text{std}_{\Theta}$ are multiplied after the extraction are 23.25 and 0.34 dex respectively. For each $N_H$, we generate a mock spectrum using the fakeit tool of XSPEC. We build the spectrum from a collisionally ionized plasma model (APEC hereafter, Smith et al. 2001) at a temperature $T_{\text{APEC}} = 1 \text{ keV}$ and ISM metal abundances (taken from Lodders 2003), absorbed by neutral material, aiming to emulate the observation of a large region of diffuse emission from the hot phase of the ISM in the midplane of the Milky Way. We set the exposure time of each observation to 1 ks. The instrumental response matrix and ancillary files have been taken from XMM-Newton observations of the Galactic center region with the EPIC-PN camera (Ponti et al. 2015). After simulating each spectrum individually, we sum them. The resulting spectrum is shown in Fig. 2. The best-fit values of the disnht*APEC model are reported in Tab. 1. Both the simulated distribution of column densities and the emission model are reproduced by the disnht model best-fit parameters within the statistical uncertainties, holding a C statistics of $\sim 964$ over 975 degrees of freedom. The tbabs*APEC best-fit model instead provides a column density of the absorbing material (Wilms et al. 2000) of $3.29 \times 10^{22} \text{ cm}^{-2}$, about an order of magnitude lower than the mean of the simulated (log-)distribution. Furthermore, when the disnht model is used, the retrieved temperature and metallicity of the emission component are consistent with a $kT = 1 \text{ keV}$ collisionally ionized plasma with solar abundances, whereas by using the tbabs model, the plasma characterization ($kT = 1.56 \pm 0.03 \text{ keV}$, $Z = 1.7 \pm 0.2 Z_\odot$) significantly differs from the origin set. The C statistics held by the best-fit tbabs*APEC model is 1594.4 over 976 degrees of freedom. There is thus an overall improvement of the disnht model with respect to tbabs, both in terms of describing the absorption and (consequently) the emission properties of the spectrum, and of the statistical goodness of the fit, at the expense of an additional degree of freedom.
3.1. Lognormal distribution approximation

The disnht model is also already provided and precomputed as an interpolation between the discrete entries of a two dimensional table. Each dimension of this table corresponds to one degree of freedom of the model. The physical distribution of column density in space or time can be arbitrarily complex to model, in terms of number of parameters required to describe it analytically. The (log-)normal distribution, if applicable, offers the advantage of providing physical insight rather than just a mathematical description and also to require a minimal amount of free parameters: a mean and a dispersion value. Is a (log-)normal distribution a sufficiently good representation of the column density distribution? We consider the assumption as sufficiently good by testing different observed column density distributions against a lognormal with the same mean, standard deviation and number of elements as the real distribution, through a Kolmogorov-Smirnov test. The approximation is valid for rejection probabilities of the null hypothesis (of the two distributions to have been drawn by the same one) larger than 0.05. We draw real distributions from the all sky column density maps of HI4PI Collaboration et al. (2016). We consider this map representative of the physical column density at high Galactic latitudes $b > 10$°. We build column density sets extracted over circular regions of different dimension to test our assumption over different angular scales. For each extraction radius we draw 1000 sets centered at random positions in the sky at $b > 10$°. From every set we obtain a probability of the set to be described by a lognormal distribution. We reject the hypothesis for p-values smaller than 0.05. The distributions of p-values obtained over sets at different angular scales are shown in Fig. 3, for all p-values larger than $10^{-5}$. The original $N_H$ map has a resolution of ≃ $3.4'$ pixels. Larger angular scales then encompass a larger number of set elements. For instance, the $\sim 0.1 (1.0) \pi \text{deg}^2$ extraction region encompasses 30 (300) pixels. As shown in Fig. 3, the column density distributions are generally consistent with a lognormal one up to $\sim 1$ deg radius, with the corresponding distribution of p-values peaking at $p > 0.05$. For larger regions, the real distribution starts to significantly deviate from lognormal. From a deeper inspection, this is due in most cases to a $N_H$ distribution resembling a sum of
Fig. 2. Simulated spectrum absorbed by a lognormal distribution of column densities. The emission model is an APEC at $T = 1$ keV and solar abundances (Lodders 2003). The mock data are the sum of 100 simulated APEC absorbed by different column density values of the neutral material extracted from a lognormal distribution. The solid lines show distinct models with avg $\Theta$ values: 23.25 (best-fit) and different std $\Theta$ 0.50 (green), 0.33 (red, best-fit), 0.20 (yellow) and 0.01 (cyan). The blue dashed line show the same emission model absorbed by a single column density equal to avg $\Theta = 23.25$ (with best-fit $kT = 1.5$ keV). Residuals $\chi = (\text{data} - \text{model})/\text{error}$ and ratios data/model are shown in the central and lower panel respectively.

3.2. Deviations from single column density absorption models

So far we have demonstrated that: i) there are several cases of significantly broad column density distributions, produced by averaging over a large enough sky area; ii) in many of such cases, the distribution of column densities can be approximated with a lognormal one; iii) such an approximation leads to a more precise characterization of the absorption spectrum in terms of statistical power and accuracy of the best-fit values. To further demonstrate the importance of adopting a more complex model than only a single exponential function of the column density and also to assess when the latter is still a good enough approximation to the spectrum, we computed the discrepancy between the two methods, in terms of the quantities $A_0$ and $\tilde{A}_0$ already introduced in Sec. 2. Provided that the limit for small std $\Theta$ of the disnht model gives the tbabs solution, the fractional (percent) deviation of $\tilde{A}_0$ with respect to $A_0$ is

$$\left(\frac{A_0}{\tilde{A}_0} - 1\right) \times 100 \quad (3)$$

and provides a mean to distinguish when the two models actually describe a significantly different absorption spectrum and to what extent. We quantify the fractional deviation at energy 0.3 keV. At lower (higher) energies the models will deviate more (less). We study the deviation for a grid of parameters in the ranges avg $\Theta \in [19.5; 23.0]$ using a 0.25 dex step and std $\Theta \in [0.02; 0.22]$ using a 0.02 dex step. The results are shown in Fig 4. Because of the exponential form, the difference between a negligible correction and an important correction is very large, even in log scale. We thus compress the color scale to highlight corrections between 1 and 30% (+100% corresponding to a factor 2). A $\sim$ 1% correction may in fact be statistically significant in a high signal-to-noise spectrum. White color regions
Fig. 3. Goodness of the approximation with a lognormal distribution in terms of Kolmogorov-Smirnov p-value distributions (log$_{10}$ scale). The vertical red line shows $p_{KS} = 0.05$. The Kolmogorov-Smirnov test is run over column density sets extracted from circular region of different areas in the $N_H$Pro all sky maps of HI4PI Collaboration et al. (2016).

Fig. 4. Percent deviation ($A_0/\tilde{A}_0 - 1 \times 100$ (log scale)) between the absorption at 0.3 keV produced by a lognormal distribution of column densities and a single-parameter absorption model (i.e. $tbabs$ ($N_H = av_{g0}$)), as a function of the free parameters $av_{g0}$ and $std_0$. Blue and red points show the scatter plot for distributions extracted respectively from $6' \times 6'$ random regions in the Galactic center column density map (Molinari et al. 2011) and $1$ deg $\times 1$ deg random region at Galactic latitudes $b > 10$ deg from the HI column densities of HI4PI Collaboration et al. (2016).

4. Conclusions

We have presented a model for the absorption of soft X-ray spectra in the case that photons are collected from different lines of sight holding a broad distribution of column density values. After showing the efficacy of the model in retrieving the absorption properties of the spectrum, we proved that several cases exist for which it is relevant to consider it. In such cases the assumption of an absorption spectrum described by a single column density value breaks. We have proven that in such cases the lognormal distribution is a fair approximation to the physical one. The absorption model proposed is computed by assuming a constant layer of emission to be absorbed by a distribution of column densities. This is a rather simple assumption compared to a general case in which emission from a plasma is expected to deviate from homogeneity regardless of absorption. However, our goal is to focus on the modeling of the absorption of the spectrum, rather than on its emission. In general, whenever the source emission can be considered smooth with respect the angular scale of the absorption medium fluctuations, the $\text{disnht}$ model is valid and provides a better description of the absorbing medium with respect to a single-valued column density model. However, in case of a non smooth underlying emission, a correction for the column density distribution is still to be preferred to a simple average-value analysis of the absorption, in terms of result accuracy, for large enough column density distributions. We expect the proposed model to be relevant for present and future generation of X-ray imaging telescopes furnished with survey capabilities and large fields of view, such as eROSITA (Merloni et al. 2012; Predehl et al. 2021) and Athena (Nandra et al. 2013).

The model is made available either through a python code to compute the absorption spectrum given an arbitrary column density distribution, or precomputed under the assumption of a
lognormal distribution both in ascii and .fits tables (etable, XSPEC readable) formats.

Acknowledgements

We thank our anonymous reviewer for the helpful comments on the manuscript. NL and GP acknowledge financial support from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program Hot-Milk (grant agreement No. [865637]). SB acknowledges financial support from the Italian Space Agency (grant 2017-12-H.0).

References

Anders, E. & Grevesse, N. 1989, Geochim. Cosmochim. Acta, 53, 197
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Balucinska-Church, M. & McCammon, D. 1992, ApJ, 400, 699
Bianchi, S., Piconcelli, E., Chiaberge, M., et al. 2009, ApJ, 695, 781
Bianchi, S., Ponti, G., Muñoz-Darias, T., & Petrucci, P.-O. 2017, MNRAS, 472, 2454
Cheng, Y., Wang, Q. D., & Lim, S. 2021, MNRAS, 504, 1627
Díaz Trigo, M., Parmar, A. N., Boirin, L., Méndez, M., & Kauffmann, J. S. 2006, A&A, 445, 179
Done, C. & Magdziarz, P. 1998, MNRAS, 298, 737
Ferrière, K. M. 2001, Reviews of Modern Physics, 73, 1031
Gupta, A., Mathur, S., Krongold, Y., Nicastro, F., & Galeazzi, M. 2012, ApJ, 756, L8
HI4PI Collaboration, Ben Bekhti, N., Flöer, L., et al. 2016, A&A, 594, A116
LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, ApJ, 800, 144
Lodders, K. 2003, ApJ, 591, 1220
Marchese, E., Baito, V., Della Ceca, R., Caccianiga, A., & Severgnini, P. 2012, MNRAS, 421, 1803
Masui, K., Mitsuda, K., Yamasaki, N. Y., et al. 2009, PASJ, 61, S115
Matt, G., Guainazzi, M., & Maiolino, R. 2003, MNRAS, 342, 422
Merloni, A., Predehl, P., Becker, W., et al. 2012, arXiv e-prints, arXiv:1209.3114
Molinari, S., Bally, J., Noriega-Crespo, A., et al. 2011, ApJ, 735, L33
Muno, M. P., Bauer, F. E., Baganoff, F. K., et al. 2009, ApJS, 181, 110
Nandra, K., Barr, D., Barcons, X., et al. 2013, arXiv e-prints, arXiv:1306.2307
PONTI, G., BIANCHI, S., MUÑOZ-DARIAS, T., et al. 2016, Astronomische Nachrichten, 337, 512
Ponti, G., Morris, M. R., Terrier, R., et al. 2015, MNRAS, 453, 172
Predehl, P., Andritschke, R., Arefieva, V., et al. 2021, A&A, 647, A1
Puccetti, S., Fiore, F., Risaliti, G., et al. 2007, MNRAS, 377, 607
Risaliti, G., Miniutti, G., Elvis, M., et al. 2009, ApJ, 696, 160
Rybicky, G., Lightman, A. P., & Domke, H. 1982, Astronomische Nachrichten, 303, 142
Smith, R. K., Brickhouse, N. S., Liedahl, D. A., & Raymond, J. C. 2001, The Astrophysical Journal, 556, L91
Wang, Q. D., Zeng, Y., Bogda, A., & Ji, L. 2021, arXiv e-prints, arXiv:2110.06995
Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914
Wolffire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, ApJ, 587, 278
Yoshino, T., Mitsuda, K., Yamasaki, N. Y., et al. 2009, PASJ, 61, 805

2 http://www.brera.inaf.it/hotmilk/models.html