The Astrophysical Journal, 983:124 (8pp), 2025 April 20 https://doi.org/10.3847/1538-4357/adc256

Hyperluminous Supersoft X-Ray Sources in the Chandra Catalog

Andrea Sacchi[1], Kevin Paggeot[1], Steven Dillmann[2], Juan Rafael Martínez-Galarza[1,3], and Peter Kosec[1]
[1] Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 20138, USA; andrea.sacchi@cfa.harvard.edu
[2] Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA
[3] AstroAI, Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 20138, USA
Received 2025 January 31; revised 2025 March 16; accepted 2025 March 17; published 2025 April 15

Abstract

Hyperluminous supersoft X-ray sources (HSSs), such as bright extragalactic sources characterized by particularly soft X-ray spectra, offer a unique opportunity to study accretion onto supermassive black holes in extreme conditions. Examples of hyperluminous supersoft sources are tidal disruption events (TDEs), systems exhibiting quasiperiodic eruptions, changing-look active galactic nuclei, and anomalous nuclear transients. Although these objects are rare phenomena among the population of X-ray sources, we developed an efficient algorithm to identify promising candidates exploiting archival observations. In this work, we present the results of a search for HSSs in the recently released Chandra catalog of serendipitous X-ray sources. This archival search has been performed via both a manual implementation of the algorithm we developed and a novel machine learning–based approach. This search identified a new TDE, which might have occurred in an intermediate-mass black hole. This event occurred between 2001 and 2002, making it one of the first TDEs ever observed by Chandra.

Unified Astronomy Thesaurus concepts: Tidal disruption (1696); Supermassive black holes (1663); X-ray active galactic nuclei (2035)

1. Introduction

We define hyperluminous supersoft X-ray sources (HSSs) as bright extragalactic sources of X-rays characterized by a supersoft spectrum. Their X-ray luminosity exceeds $10^{41}\,\mathrm{erg\,s^{-1}}$, and their spectra are extremely steep. When modeled with a power-law profile, the typical photon indices are $\Gamma \gtrsim 3$, if a blackbody model is adopted, typical temperatures are $kT \lesssim 100\,\mathrm{eV}$, and with very few exceptions, all of their emission is below 2 keV. HSSs are usually spatially coincident with the nuclear region of their host galaxies. This, coupled with their high luminosities and steep spectra, indicates that the origin of their emission is accretion onto supermassive black holes (SMBHs).

The best-studied and best-represented subclass of HSSs is that of tidal disruption events (TDEs). TDEs occur when a star orbit happens to plunge deep enough into the potential well of an SMBH to be disrupted by tidal forces. In the standard scenario, roughly half of the resultant stellar debris remains on bound orbits and eventually falls back on the SMBHs and later is accreted after forming a bright accretion disk. The electromagnetic emission generated by this newly formed accretion disk is predicted to peak in the ultraviolet (UV) band, while the steep Wien tail of the spectrum would fall in the X-ray band. This emission is expected to rapidly rise and then slowly decay over timescales of days to months. These phenomena were first theorized (J. G. Hills 1975; M. J. Rees 1988; E. S. Phinney 1989) and first observed by ROSAT (N. Bade et al. 1996). The properties of the first TDEs observed were exactly the ones predicted: bright and soft X-ray transients from otherwise quiescent galactic nuclei. The importance of these phenomena was immediately clear. They represent unique opportunities to study newly formed accretion flows on otherwise quiescent and relatively "light" SMBHs ($\sim 10^6\,M_\odot$). This is because the tidal radius for a Sun-like star is smaller than the Schwarzschild radius for an SMBH with a mass $\gtrsim 10^8\,M_\odot$, and hence a star would not get disrupted but rather directly swallowed by any more massive SMBH: one such encounter would not generate a TDE.

After the first TDEs were discovered in the X-ray band, these phenomena were found to be bright across the full electromagnetic spectrum, and today we know of about 100 TDEs observed in every wavelength, from radio (Y. Cendes et al. 2024) to infrared (M. Masterson et al. 2024), optical/UV (S. Gezari et al. 2012; S. van Velzen et al. 2021; E. Hammerstein et al. 2023), X-ray (R. Saxton et al. 2020; S. Sazonov et al. 2021), and gamma ray (D. N. Burrows et al. 2011). The simple picture described above struggles to explain all of the data accumulated on these phenomena in the last three decades. Ongoing debates revolve around where, when, and how the multiwavelength properties of TDEs originate and are correlated (e.g., L. Dai et al. 2018; C. Bonnerot et al. 2021). To further complicate this scenario, new classes of bright and supersoft transients from SMBHs have emerged in recent years. The discovery of quasiperiodic eruptions (QPEs), spectacular X-ray flashes from galactic nuclei recurring over hourly timescales (G. Miniutti et al. 2019; M. Giustini et al. 2020; R. Arcodia et al. 2021; E. Quintin et al. 2023; M. Nicholl et al. 2024), repeated/partial TDEs (A. V. Payne et al. 2021; A. Malyali et al. 2023; T. Wevers et al. 2023; M. Guolo et al. 2024; Z. Lin et al. 2024), long-lived TDEs (J. S. He et al. 2021; D. Lin et al. 2017b; A. J. Goodwin et al. 2022), and active galactic nuclei (AGN) in a purely thermal state (J. S. He et al. 2021; D. Lin et al. 2017a) calls for further theoretical and observational efforts in order to explain the nature and characteristic of these phenomena. The properties of this plethora of sources are similar enough to those of "classical" TDEs that a comprehensive model of the latter should also be able to somewhat account for the former. Although all of these classes of sources exhibit different temporal evolutions, their

X-ray properties, in terms of spectrum and overall luminosity, are strikingly similar; hence we collect them all and define the HSS category.

The path to advance our understanding of these phenomena, along with further modeling and theoretical investigations, is to identify new HSSs. Today's standard way of detecting these sources is to exploit their extreme variability, and repeated scans of the sky, performed at different wavelengths, have proven to be highly effective in this task. However, no such scan is currently being performed in the X-ray band, and overlooked HSSs may still be awaiting identification in archival data sets, particularly in available catalogs. In a recent publication (A. Sacchi et al. 2023; hereafter Sa23), we scanned the XMM-Newton source catalog (N. A. Webb et al. 2020) for HSSs. Rather than relying on the variability of the sources, we exploited the unique combination of high-luminosity and supersoft spectral properties of HSSs to identify them in the catalog, which allows us to include in our search sources for which only single-epoch observations are available. The results described in Sa23 are promising: we successfully identified nine previously unknown HSSs, among which four are candidate TDEs. A follow-up observation performed with XMM-Newton confirmed the nature of one of these candidates, making it the first ever X-ray TDE identified solely through X-ray spectroscopy.

In this paper, we repeat this process on the recently released updated version of the Chandra source catalog. This time, alongside the algorithm employed to search the XMM-Newton source catalog, we also adopt an unsupervised machine learning (ML) method by S. Dillmann et al. (2025b) that relies on learning a physically meaningful representation of the individual X-ray photon events for each catalog source. This accomplishes multiple goals at once. We successfully identify a new TDE, validate the ML algorithm, and explore its potentiality for future applicability to other catalogs as well as other types of sources.

This paper is organized as follows: in Section 2 we describe how we identified HSSs in the Chandra source catalog via our "classic" algorithm, in Section 3 we discuss the properties of the new TDE we identified, in Section 4 we present the comparison of our algorithm with the ML one, and in Section 5 we draw our conclusions.

2. Classical Approach

2.1. Sample Selection

Our process for identifying new HSSs in the Chandra source catalog (CSC 2.1; I. N. Evans et al. 2024) was similar to the process described in Sa23 for the XMM-Newton source catalog. We proceeded to download the full catalog, which amounts to more than 1.3 million individual detections of about 407,000 unique sources observed before the end of 2021. We accessed the catalog through the application that represents its main interface.[4] We used individual detections rather than stacked data sets of single sources. This is because HSSs might exhibit extreme variability, and different observations of the same source can have significantly different properties.

For all detections, we retrieved their fluxes, hardness ratios (HRs), significances (Ss) in different energy bands, and errors in each of these quantities. We adopted aperture flux, to be as independent as possible from particular choices of spectral models, and flux Ss, which is a simple estimate of the ratio of the flux measurement to its average error. We selected our sources based on their S in the soft (0.5–1.2 keV) band, HR in the soft-to-medium (1.2–2 keV) band, and luminosity in the broad (0.5–7 keV) band. We compute the absorbed luminosities from the X-ray fluxes using redshift measurements. We crossmatched the Chandra catalog with all the redshift catalogs available in SIMBAD (M. Wenger et al. 2000). We performed the crossmatch using the tool TOPCAT (M. B. Taylor 2005), with a matching radius corresponding to the error on the source position (provided by the Chandra catalog at 90% confidence level) and considering the best optical match for each X-ray source, as multiple X-ray sources might be hosted in the same galaxy. Knowing the redshift, the luminosity values are computed by assuming a standard flat ΛCDM cosmology with $\Omega_M = 0.3$ and $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. We note that, given that the emission from HSSs is extremely soft, the natural choice would be to employ also the ultrasoft (0.2–0.5 keV) band. However, the progressive and dramatic decrease in effective area of Chandra's detectors, especially at ultrasoft energies (P. P. Plucinsky et al. 2018; P. P. Plucinsky et al. 2022), implies that the ultrasoft band is not reliable for source selection.

To build our sample we discarded all sources with a broadband luminosity $L_X < 10^{41}\,\mathrm{erg\,s^{-1}}$, HR > -0.55 in the soft-to-medium band, and $S > 7$ in the soft band. The choice of HR corresponds to a photon index $\Gamma \gtrsim 3$ if the X-ray spectra of the sources were modeled with a power-law profile. These three criteria ensure that we are excluding sources powered by stellar processes (which rarely exceed the chosen luminosity threshold), the bulk of the AGN population (which usually exhibits much flatter spectra), and that the source statistics is high enough to perform a reliable spectral and timing analysis. This resulted in 222 detections from 125 individual sources. Figure 1 shows the X-ray luminosity as a function of the HR for all selected detections.

These 222 detections represent about 4 times as many HSS candidates as were present in the XMM catalog from Sa23, even though the initial catalogs are similarly sized. This is because we are not imposing any conditions on the ultrasoft (0.2–0.5 keV) band, and hence we expect more spurious sources to be present in this sample. To further clean our selection we thus exploit the optical classification of the sources, which we retrieved along with the redshift information from SIMBAD. Straightaway we identified and excluded 24 spurious sources (86 detections): 10 sources associated with stars, 1 located behind a supernova remnant, 4 associated with a foreground galaxy, 7 whose soft X-ray photons are associated with the emission from galaxy clusters, and 2 whose emission is extended and attributable to a pair of interacting galaxies. Similarly to what we found in Sa23, at this stage, the bulk of our sample is composed of sources optically classified as AGN, 75 (96 detections) as broad-line AGN, and 15 (24 detections) Seyfert 2 galaxies (mostly classified by M. P. Véron-Cetty & P. Véron 2010 and I. Pâris et al. 2014). The X-ray emission of these sources passes our supersoft criterion as it is characterized by a strong soft excess dominating the energy range below 2 keV. However, they also exhibit hard X-ray emission, which can be well reproduced by a power-law model (seen through different ranges of absorption material). Overall, the X-ray emission of

[4] The CSC 2.1 application is available at http://cda.cfa.harvard.edu/cscview/.

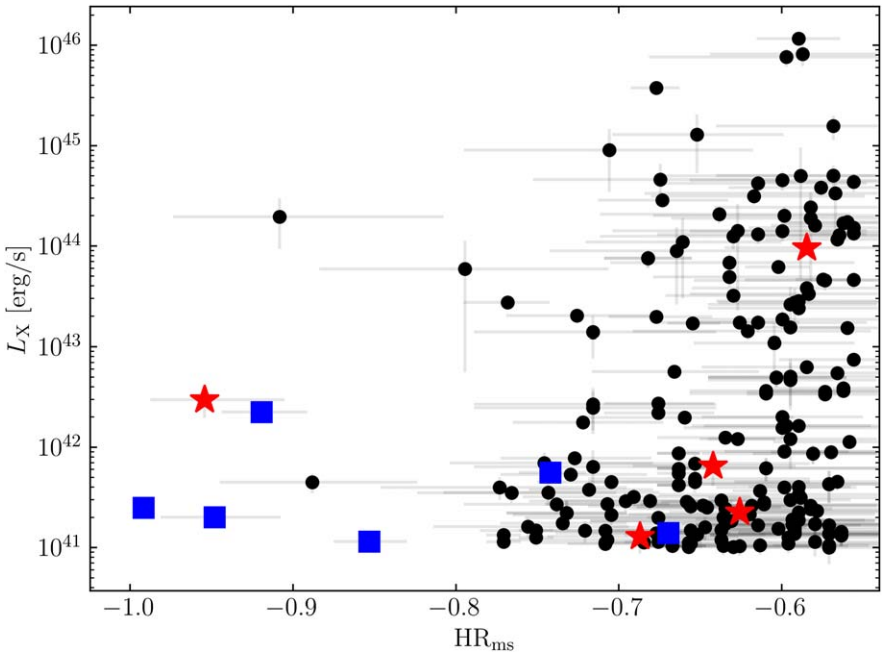

Figure 1. X-ray luminosity as a function of the HR in the soft-to-medium band for the source selected via their X-ray properties. In black are the sources we then excluded based on their optical classification, in blue are the HSSs known in the literature, and in red are the sources classified as nonactive galaxies whose X-ray spectrum we finally analyzed. The error bars correspond to 1σ uncertainties.

Table 1
Properties of the Five Sources We Selected as Candidate HSSs (Indicated by Red Stars in Figure 1)

ID	R.A.	Decl.	z	HR_{ms}	L_X (erg s^{-1})	Classification
2XMMi J215039.5–055335	327.6646	–5.8934	0.392	–0.58	9.59×10^{44}	AGN
IC 1792	34.7544	34.4623	0.035	–0.63	2.26×10^{41}	Collisionally ionized gas
2MASX J23272553+2635075	351.8564	26.5854	0.058	–0.69	1.29×10^{41}	Collisionally ionized gas
ESO 152–37	29.1705	–55.9271	0.091	–0.64	6.51×10^{41}	Collisionally ionized gas
SDSS J103557.83+572500.2	158.9909	57.4167	0.102	–0.95	2.96×10^{42}	TDE

Note. Identifier and redshift pertain to the optical counterpart from SIMBAD. Coordinates, HR, and luminosity (computed from the broad 0.5–7 keV band flux) refer to the X-ray source. The reported classification is the one decided upon the analysis of the X-ray spectrum of each source.

these sources is compatible with their optical classification, and we will not address them further.

This cleaning process leaves us with a sample of 11 candidate sources. Six of these are well-known HSSs. RX J1301.9+2747 and GSN 069 are peculiar AGN showing QPEs (G. Miniutti et al. 2019; M. Giustini et al. 2020). WINGS J134849.88+263557.5 and ASASSN-14li are two famous TDEs, the former discovered in the galaxy cluster A1795 (W. P. Maksym et al. 2013) and the latter often referred to as a "textbook" TDE, given that its multi-wavelength emission follows precisely all the predictions for these phenomena (W. P. Maksym et al. 2014). SDSS J150052.07+015453.8 is a long-lived TDE discovered by D. Lin et al. (2017b), and [FWB2009] HLX-1 is a hyperluminous X-ray source hosted in ESO 243-49, which is thought to be powered by accretion over a bona fide intermediate-mass black hole (IMBH; S. A. Farrell et al. 2009). These sources are exactly the target of our search and testify to the reliability of our algorithm and cleaning procedure.

This leaves five sources, which are classified as nonactive galaxies based on their optical emission. The properties of these five sources are listed in Table 1. To investigate their nature further, we performed a full X-ray spectral analysis of all the publicly available data sets for each source.

2.2. X-Ray Spectral Analysis

All Chandra data sets were reprocessed and reduced with the Chandra Interactive Analysis of Observations software package (`CIAO`, v.4.12; A. Fruscione et al. 2006) and the `CALDB` 4.9.0 release of the calibration files. For each source, spectra, spectral redistribution matrices, and ancillary response files were generated using the `CIAO` script `specextract`. Spectra were extracted from regions encompassing a point-spread function fraction of about 95%. Local backgrounds were estimated from source-free annular regions, centered on the source position. These annular regions have inner radii of $\approx 10''$ and outer radii of $\approx 30''$. If these regions are too close to the detector borders, a circular region of $\approx 20''$ radius was adopted instead.

One source, 2XMMi J215039.5-055335 has been observed by Chandra twice. In both observations, the X-ray spectrum of the source is well modeled by a power law with a photon index $\Gamma \approx 2$ when the full 0.5–7 keV energy band is considered. This suggests that the X-ray emission from this source is attributable to a typical X-ray corona and, coupled with its high X-ray

luminosity ($\approx 10^{44}$ erg s^{-1}), implies that it is a misclassified AGN. Three sources, IC 1792, 2MASX J23272553+2635075, and ESO 152–37, exhibit X-ray spectra that can be well reproduced by an `apec` model, suggesting that the origin of their emission is powered by collisionally ionized gas, heated up when falling in the gravitational potential well of the galaxy itself. This is also compatible with their luminosity, which barely exceeds the 10^{41} erg s^{-1} threshold. The X-ray spectra and spectral parameters of these sources are reported in the Appendix. However, although passing our cuts, these sources are not HSSs: they are either not powered by accretion of gas onto a massive BH or their X-ray spectrum is not supersoft when the full X-ray energy band is inspected.

Finally, one source, SDSS J103557.83+572500.2, exhibits a supersoft thermal X-ray spectrum coupled with a luminosity exceeding 10^{42} erg s^{-1}. This source is a bona fide HSS and represents a strong candidate TDE. Its multiwavelength emission is described at length in the next section.

3. A New "Old" TDE

SDSS J103557.83+572500.2 (J1035+57 hereafter) is a galaxy at spectroscopic redshift $z = 0.10203$ (R. Ahumada et al. 2020), with a stellar mass of $6.45 \times 10^{10}\,M_{\odot}$ and a star formation rate (SFR) of $4.24\,M_{\odot}\,\mathrm{yr}^{-1}$ (Y.-Y. Chang et al. 2015), and it is cataloged as a star-forming galaxy by Y. Toba et al. (2014). The infrared colors (W1 − W2 = 0.258, from AllWISE, reported by D.-W. Kim et al. 2023), as well as the optical spectroscopy from the Sloan Digital Sky Survey (SDSS; R. Ahumada et al. 2020), suggest no ongoing nuclear activity. The X-ray luminosity from unresolved X-ray binaries, based on the reported values of stellar mass and SFR, amounts to $L_{\mathrm{X,XRB}} \approx 1.3 \times 10^{40}$ erg s^{-1}, estimated employing the B. D. Lehmer et al. (2010) relation. From the stellar mass of J1035+57, exploiting the M_{BH}–$M_{\star}$ relation for nonactive galaxies (A. E. Reines & M. Volonteri 2015), we infer a SMBH mass of $4.8^{+9.4}_{-3.2} \times 10^{8}\,M_{\odot}$.

J1035+57, due to its proximity to Lockman Hole, was detected by Chandra during a serendipitous visit on 2002 April 30 (ObsID 3346, P.I. Barger). The effective exposure time of this visit is 38.2 ks, and the source position lies on the edge of the detector, at about 16′ of off-axis angle. Considering this, the X-ray source position is uncertain at about 2″, and it is marginally compatible with its host galaxy center. Figure 2 shows an optical image of J1035+57 with the superposed location and uncertainty of the X-ray source.

The X-ray spectrum of J1035+57, reported in Figure 3, can be well reproduced by a simple blackbody model (redshifted at the source distance, `zbbody` in `XSPEC`). Given the high signal-to-noise ratio of the detection, we rebinned the spectrum to have at least 20 counts in each bin and adopted the standard χ^2 statistic. Considering the supersoft nature of the source's spectrum and the fact that the data set was acquired in 2002 when the Chandra effective area in the soft X-ray band was not degraded by molecular contamination, we also include the counts collected in the 0.3–0.5 keV band in our analysis. To the blackbody model, we added a layer of neutral absorber (`TBabs`) with N_{H} fixed to the Galactic value 5.46×10^{19} cm^{-2} (HI4PI Collaboration et al. 2016), using abundances from J. Wilms et al. (2000), with the photoelectric absorption cross sections from D. A. Verner et al. (1996). We obtained a satisfactory fit ($\chi^2/\nu = 36.3/29 \approx 1.24$, where χ^2 is the value of the statistic and ν the degrees of freedom) for a temperature of 85 ± 3 eV

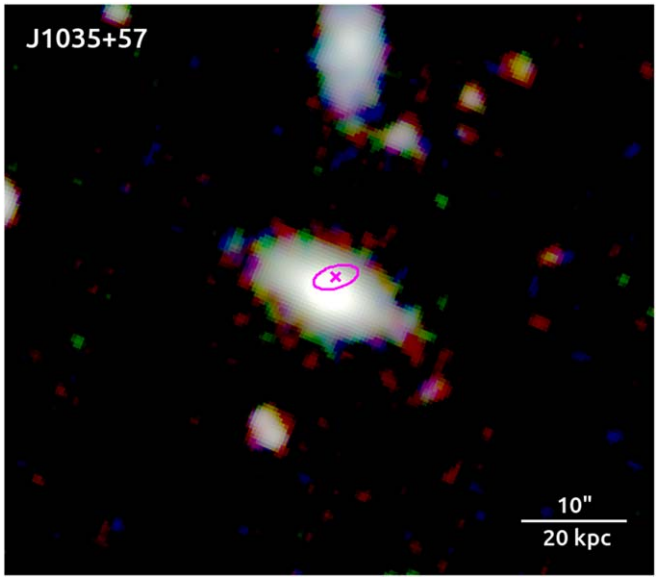

Figure 2. Three-color composite image of J1035+57 obtained with *SDSSgri* filters. The cross and the magenta ellipse indicate the centroid and positional uncertainty of the X-ray source detected by Chandra.

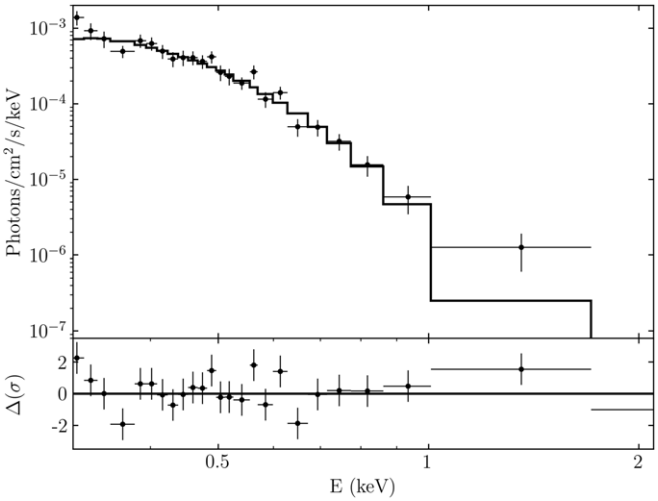

Figure 3. X-ray spectrum (upper panels) and residuals (lower panels) of J1035+57. The solid lines show the best-fitting models described in the text. Only the 0.3–2 keV energy range is shown as above 2 keV the spectrum is background dominated.

and an unabsorbed luminosity of $1.28 \pm 0.07 \times 10^{42}$ erg s^{-1} in the 0.5–2 keV band. The spectrum residuals offer hints of excess counts in the 1–2 keV band, but adding a power law to reproduce this excess does not significantly improve the fit's quality. Likewise, the quality of the fit is not improved by adding a free-to-vary layer of neutral absorption. The 3σ upper limit on the intrinsic absorption at the source location (or within the host galaxy) is $N_{\mathrm{H,intrinsic}} < 3 \times 10^{20}$ cm^{-2}. We also attempted fitting the spectrum with alternative models, such as a power-law and an `apec` model, and in both cases, we obtained unacceptable fits ($\chi^2/\nu = 57.5/29 \approx 1.98$ and $\chi^2/\nu = 127.9/29 \approx 4.41$, respectively).

We inspected the X-ray light curve of J1035+57 in different energy bands and detected no significant short-term variability.

The location of J1035+57 fell in the field of view of Chandra 1 yr before the reported detection, on 2001 May 16 (ObsID 1697, P.I. Mushotzky), for a total exposure time of

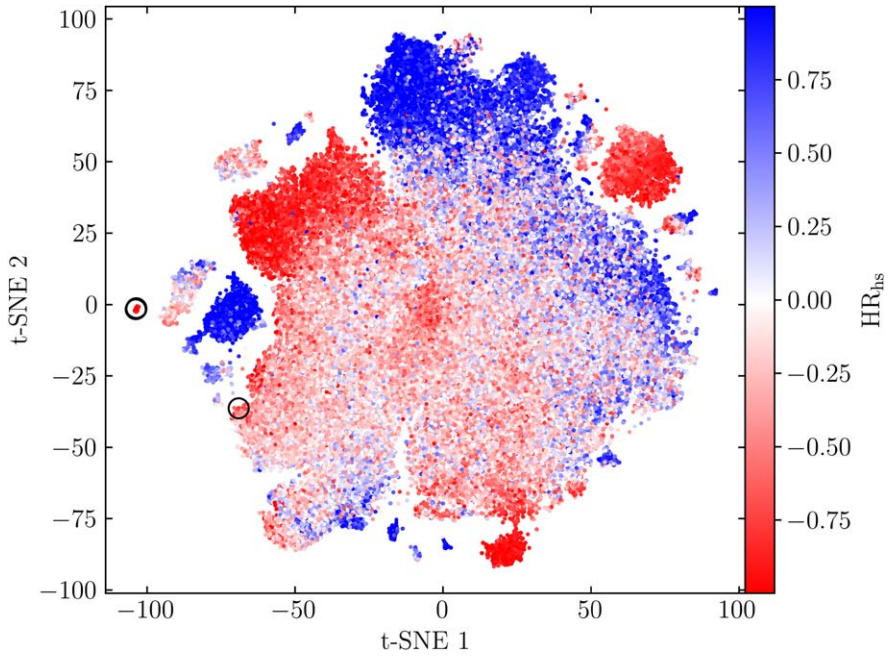

Figure 4. t-SNE visualization of the autoencoder embedding representations for the sources considered in S. Dillmann et al. (2025b), color coded by HR_{hs}. The black empty circles indicate the location of known HSSs.

43.7 ks, and in this case too its position is about 16' off axis. In this visit the source was not detected, and the 3σ upper limit we computed with the `CIAO` script `srcflux` amounts to 1.46×10^{41} erg s^{-1}, a value a factor of about 9 lower than the detection 1 yr later. No other X-ray telescope visited the source location prior to or since, and the ROSAT All Sky Survey has not detected J1035+57, although the inferred flux upper limit is too high to be meaningful.

At the time of the 2002 X-ray detection, the source location was not covered in any other band (optical or infrared) and has not shown significant variability in any observation since. The source is also not detected in any radio band.

The host galaxy properties, X-ray source location, long-term variability, spectral shape, and luminosity all suggest that J1035+57 emission is due to a TDE. The nondetection dating 1 yr before the detection places the date of the TDE in the second half of 2001 or the first half of 2002.

We note that the SMBH mass we inferred from the galaxy stellar mass is unconventionally heavy for a TDE; in fact, the Schwarzschild radius of a $4.8 \times 10^8\,M_\odot$ SMBH is larger than its tidal radius for a Sun-like star, and the Wien tail of multicolor disk spectrum would not fall in the soft X-ray band but rather on the far-UV band. This opens two possible scenarios: either the inferred SMBH mass is heavily overestimated, making J1035+57 an outlier of the $M_{\rm BH}$–$M_\star$ relation, or this TDE occurred on an IMBH, hosted in the same galaxy.

4. ML Approach

In the previous sections, we manually applied the algorithm developed in Sa23 and successfully identified a previously overlooked TDE in the Chandra catalog. In this section, we explore the applicability and efficiency of an ML technique developed in S. Dillmann et al. (2025b) for this task.

In ML, representation learning (Y. Bengio et al. 2013), refers to the process of training a deep neural network to learn a low-dimensional informative representation of the input data that can be used for various downstream tasks, such as classification or regression on relevant parameters. It can also be used for similarity searches and anomaly detection in large data sets characterized by high dimensionality and represents a more flexible and scalable approach with respect to classical methods.

In the X-ray band, the application of classification algorithms based on ML techniques has only been recently explored (H. Yang et al. 2022; V. S. Pérez-Díaz et al. 2024). Most efforts have focused on the identification of variability patterns and in particular of fast X-ray transients (S. Dillmann et al. 2025a, 2025b). Representation learning algorithms build compressed, low-dimensional representations of high-dimensionality objects. In these learned embedding spaces, anomalies are typically isolated from the bulk of "normal" objects, and HSSs, the targets of our search, owing to their spectral properties, are expected to be anomalies in this representation space.

We adopt the framework developed by S. Dillmann et al. (2025b), which learns a representation of X-ray sources by training a sparse autoencoder to reconstruct the 2D distribution of arrival times and energies of the photons associated with those sources. After training, the learned representations encode meaningful physical information about the sources, such as variability and spectral properties, that are directly learned from the data rather than computed using relatively complicated pipelines.

This approach, which led to the discovery of the new fast extragalactic transient XRT 200515 (S. Dillmann et al. 2025b), was trained using 95,473 Chandra filtered event files corresponding to 58,932 sources. This is a smaller data set than the one we analyzed in the previous section, but notably, it included three HSSs: SDSS J150052.07+015453.8, WINGS J134849.88+263557.5, and two observations of ASASSN-14li. We investigate if the learned representations of these known TDEs are isolated in the representation space with respect to the rest of the X-ray sources. If so, this would indicate the algorithm's ability to identify new HSS candidates.

Figure 4 shows a visualization of the autoencoder latent space after training. In addition to the autoencoder, the high-

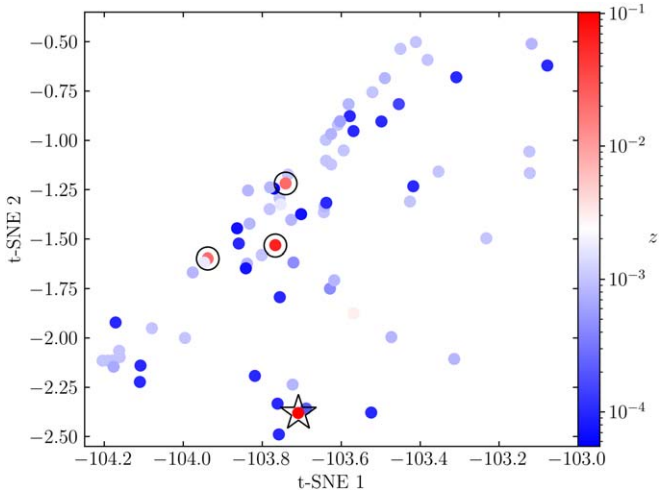

Figure 5. Zoom-in of the previous embedding representations, this time color coded by redshift. The black empty circles indicate the location of known HSSs. The black empty star indicates the location of the newly identified TDE.

dimensional input data (a vector representing the 2D distribution of photon arrival times and energies) is further reduced to the shown 2D representation using a t-distributed stochastic neighbor embedding (t-SNE; L. van der Maaten & G. Hinton 2008). The embeddings are color coded by spectral hardness, and we have highlighted the embedding locations of the known HSSs with the black circles. Three out of the four HSSs' observations, the two of ASASSN-14li and the single of WINGS J134849.88+263557.5, cluster in the same "island" of the 2D map. In this pilot study, we focus on using this representation learning approach for similarity searches around the known HSSs. The embedding region in which they live is well isolated and might harbor additional HSS candidates.

We retrieved optical information (redshift and classification) for each of the 40 sources (with 75 corresponding observations) in the selected region or island of the embedding space where three of the HSSs' observations cluster. Figure 5 shows a zoom-in of the selected region, where each source is color coded by redshift. While the vast majority of the objects in the selected regions belong in the Milky Way or nearby galaxies ($z \lesssim 0.002$), four objects have redshift $z \gtrsim 0.02$. Three of these are known HSSs (two observations of ASASSN-14li), and the fourth source is our newly found TDE, J1035+57, highlighted in Figure 5 by a black empty star and described at length in Section 3.

This is a promising result that highlights the applicability and efficiency of ML and representation learning techniques to find and classify X-ray sources of all kinds, not limited to objects exhibiting peculiar short-term variability. An in-depth analysis of all the sources in the selected region, as well as possible interesting neighbors of the HSS falling outside of this cluster, is beyond the scope of this paper, and it will be investigated in a future publication.

5. Conclusions

In this paper, we mined the most recent release of the Chandra source catalog searching for HSSs. The study of these peculiar sources allows us to investigate the properties of the less-massive SMBHs and to probe accretion processes on timescales accessible to human beings.

Starting from ≈1.3 million detections of 407,000 sources, we selected, owing to their X-ray properties, a parent sample of 125 sources, which we further filtered based on their optical classification to a final sample of 11 sources. Out of these, 6 are known HSSs: TDEs, QPEs, and a bona fide accreting IMBH. These 6 objects are an encouraging confirmation that our filtration process indeed successfully selects HSSs.

We performed a complete X-ray spectral analysis of the remaining five sources and found that one is a misclassified AGN, three are not powered by accretion onto SMBHs, and finally one whose emission is perfectly compatible with being a TDE. This newly found TDE, J1035+57, exploded between late 2001 and early 2002, making it one of the first TDEs ever observed by Chandra. The relatively large mass of the host galaxy implies that either the SMBH in its center is a low outlier of the $M_{\rm BH}$–$M_\star$ relation or that this TDE occurred on an IMBH hosted in the same galaxy.

Finally, we explored the applicability of ML and representation learning techniques in identifying HSSs. Exploiting reduced-dimensionality embeddings of part of the Chandra catalog, we identified a region of this parameter space within which some known HSSs group together. After retrieving redshift information for the neighboring sources of the clustered HSSs, we noticed that almost all sources are hosted in the Milky Way or in other nearby galaxies, except for one: this source is the newly identified TDE J1035+57. This exciting result proves that representation learning can be employed for X-ray source classification, and its power and versatility can be harnessed to identify new HSSs.

Acknowledgments

We thank the anonymous referee for the insightful comments that improved the paper. A.S. acknowledges support from a Scholarly Studies Award. P.K. was supported by NASA through the NASA Hubble Fellowship grant HST-HF2-51534.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Incorporated, under NASA contract NAS5-26555. This research has made use of data obtained from the Chandra Data Archive and the Chandra Source Catalog, both provided by the Chandra X-ray Center (CXC). This paper employs a list of Chandra data sets, obtained by the Chandra X-ray Observatory, available at DOI:10.25574/cdc.334.

Facility: CXO.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), Matplotlib (J. D. Hunter 2007), NumPy (C. R. Harris et al. 2020), Topcat (M. B. Taylor 2005), Ds9.

Appendix
X-Ray Spectral Analysis of Excluded Sources

In this section, we present the X-ray spectral analysis of the four sources we excluded as their X-ray emission is not compatible with that of an HSS.

A.1. 2XMMi J215039.5–055335

This source, located at spectroscopic redshift $z = 0.3928$ (J. L. Connelly et al. 2012), is not classified as an AGN. It has been observed twice with Chandra, in 2006 (ObsId 6791, P.I. Mulchaey) and 10 yrs later in 2016 (ObsID 17862, P.I. Lin) for 100 and 77 ks of observing time, respectively. The source

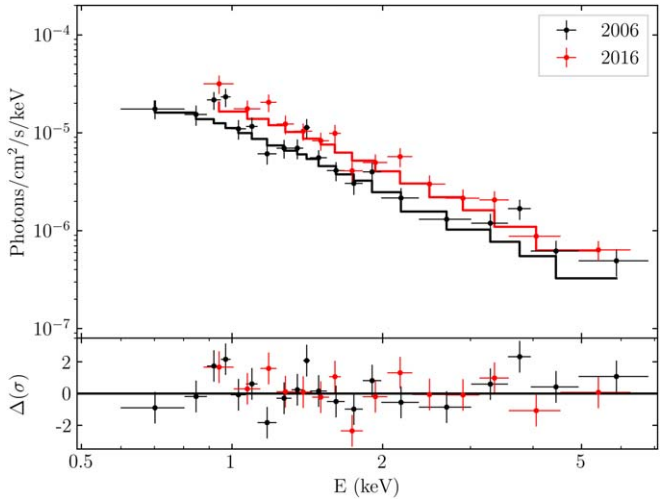

Figure 6. X-ray spectrum (upper panels) and residuals (lower panels) taken with Chandra in 2006 (in black) and 2016 (in red) of 2XMMi J215039.5–055335. The solid lines show the best-fitting models described in the text.

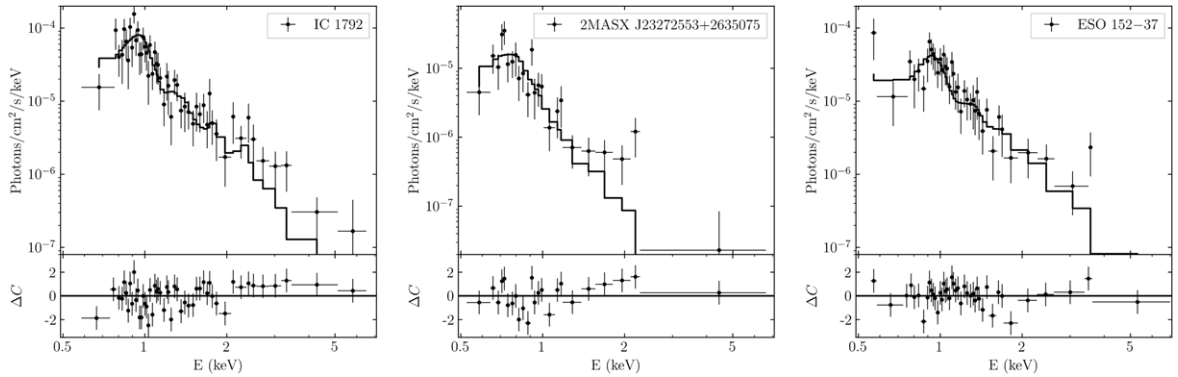

Figure 7. X-ray spectra of, from left to right, IC 1792, 2MASX J23272553+2635075, and ESO 152−37. The X-ray spectrum is shown in the upper panels and residuals in the lower ones. The solid lines show the best-fitting models described in the text.

location has also been visited by XMM-Newton, but the analysis of that data set is beyond the scope of this paper.

The X-ray spectra of the source were regrouped to have at least 20 counts per bin, and χ^2 statistic was employed. The spectra were fitted simultaneously, and we adopted a simple power-law model, with a neutral absorber with $N_{\rm H}$ fixed at the Galactic value. We obtained an acceptable fit ($\chi^2/\nu = 41/31 \approx 1.3$) with a value of photon index $\Gamma = 2.1 \pm 0.1$ and an unabsorbed luminosity in the 0.5–7 keV band of $(3.8 \pm 0.2) \times 10^{43}$ in 2006 and $(5.5 \pm 0.2) \times 10^{43}$ in 2016. We attempted fitting the spectra with a blackbody model but obtained an unacceptable-quality fit ($\chi^2/\nu = 132/31 \approx 4.2$). The X-ray spectrum in the two Chandra visits is reported in Figure 6. The spectral shape of this source excludes it from being an HSS and coupled with its luminosity suggests that it is in fact a misclassified AGN.

A.2. IC 1792, 2MASX J23272553+2635075, ESO 152–37

These three sources, all classified as quiescent galaxies, are located at redshift $z = 0.035$, 0.056, and 0.091, respectively (G. Paturel et al. 2002; D. H. Jones et al. 2009; J. P. Huchra et al. 2012). Each galaxy has been observed once with Chandra IC 1792 in 2009 (ObsID 11575, P.I. Croston), 2MASX J23272553+2635075 in 2010 (ObsID 11830, P.I. Garmire), and ESO 152–37 in 2012 (ObsID 13489, P.I. Benson). The X-ray spectra of each source were regrouped to have at least 1 count per bin, and modified Cash statistic (W-statistic, in the `XSPEC` implementation) was employed. We fit the spectra, shown in Figure 7, with an `apec` model with a neutral absorber with $N_{\rm H}$ fixed at the Galactic value. For IC 1792 we obtained an acceptable fit ($C/\nu = 140/113 \approx 1.2$) corresponding to a plasma temperature of $kT = 1.13 \pm 0.08$ keV and an unabsorbed luminosity in the 0.5–7 keV band of $(2.0 \pm 0.1) \times 10^{41}$ erg s^{-1}. For 2MASX J23272553+2635075 the best fit ($C/\nu = 64.3/60 \approx 1.07$) corresponds to a plasma temperature of $kT = 0.69 \pm 0.07$ keV and an unabsorbed luminosity in the 0.5–7 keV band of $(7.3 \pm 0.7) \times 10^{40}$ erg s^{-1}. Finally, for ESO 152–37 we obtained an acceptable fit ($C/\nu = 147.2/171 \approx 0.86$) with a plasma temperature of $kT = 1.42 \pm 0.09$ keV and an unabsorbed luminosity in the 0.5–7 keV band of $(9.4 \pm 0.5) \times 10^{41}$ erg s^{-1}. We attempted fitting each spectrum with a blackbody model but obtained an unacceptable-quality fit or a much-worse-quality fit. The X-ray spectral properties and luminosity of these sources suggest that rather than being powered by ongoing accretion onto SMBHs, their emission is due to circumgalactic gas warming up while falling in the galactic gravitational potential well.

ORCID iDs

Andrea Sacchi https://orcid.org/0000-0002-7295-5661
Steven Dillmann https://orcid.org/0000-0002-4773-1463
Juan Rafael Martínez-Galarza https://orcid.org/0000-0002-5069-0324
Peter Kosec https://orcid.org/0000-0003-4511-8427

References

Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3
Arcodia, R., Merloni, A., Nandra, K., et al. 2021, Natur, 592, 704
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Bade, N., Komossa, S., & Dahlem, M. 1996, A&A, 309, L35
Bengio, Y., Courville, A., & Vincent, P. 2013, ITPAM, 35, 1798
Bonnerot, C., Lu, W., & Hopkins, P. F. 2021, MNRAS, 504, 4885
Burrows, D. N., Kennea, J. A., Ghisellini, G., et al. 2011, Natur, 476, 421
Cendes, Y., Berger, E., Alexander, K. D., et al. 2024, ApJ, 971, 185
Chang, Y.-Y., van der Wel, A., da Cunha, E., & Rix, H.-W. 2015, ApJS, 219, 8
Connelly, J. L., Wilman, D. J., Finoguenov, A., et al. 2012, ApJ, 756, 139
Dai, L., McKinney, J. C., Roth, N., Ramirez-Ruiz, E., & Miller, M. C. 2018, ApJL, 859, L20
Dillmann, S., Martínez-Galarza, J. R., Soria, R., Di Stefano, R., & Kashyap, V. 2025a, Representation Learning for Time-domain High- Energy Astrophysics: Transient Candidates Catalog of X-ray Flares and Dips, Zenodo, doi:10.5281/zenodo.14589318
Dillmann, S., Martínez-Galarza, J. R., Soria, R., Stefano, R. D., & Kashyap, V. L. 2025b, MNRAS, 537, 931
Evans, I. N., Evans, J. D., Martínez-Galarza, J. R., et al. 2024, ApJS, 274, 22
Farrell, S. A., Webb, N. A., Barret, D., Godet, O., & Rodrigues, J. M. 2009, Natur, 460, 73
Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, Proc. SPIE, 6270, 62701V
Gezari, S., Chornock, R., Rest, A., et al. 2012, Natur, 485, 217
Giustini, M., Miniutti, G., & Saxton, R. D. 2020, A&A, 636, L2
Goodwin, A. J., van Velzen, S., Miller-Jones, J. C. A., et al. 2022, MNRAS, 511, 5328
Guolo, M., Pasham, D. R., Zajaček, M., et al. 2024, NatAs, 8, 347
Hammerstein, E., van Velzen, S., Gezari, S., et al. 2023, ApJ, 942, 9
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
He, J. S., Dou, L. M., Ai, Y. L., et al. 2021, A&A, 652, A15
HI4PI Collaboration, Ben Bekhti, N., Flöer, L., et al. 2016, A&A, 594, A116
Hills, J. G. 1975, Natur, 254, 295
Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, ApJS, 199, 26
Hunter, J. D. 2007, CSE, 9, 90
Jones, D. H., Read, M. A., Saunders, W., et al. 2009, MNRAS, 399, 683
Kim, D.-W., Cassity, A., Bhatt, B., et al. 2023, ApJS, 268, 17
Lehmer, B. D., Alexander, D. M., Bauer, F. E., et al. 2010, ApJ, 724, 559
Lin, D., Godet, O., Ho, L. C., et al. 2017a, MNRAS, 468, 783
Lin, D., Guillochon, J., Komossa, S., et al. 2017b, NatAs, 1, 0033
Lin, Z., Jiang, N., Wang, T., et al. 2024, ApJL, 971, L26
Maksym, W. P., Ulmer, M. P., Eracleous, M. C., Guennou, L., & Ho, L. C. 2013, MNRAS, 435, 1904
Maksym, W. P., Miller, J. M., Cenko, S. B., et al. 2014, ATel, 6834, 1
Malyali, A., Liu, Z., Rau, A., et al. 2023, MNRAS, 520, 3549
Masterson, M., De, K., Panagiotou, C., et al. 2024, ApJ, 961, 211
Miniutti, G., Saxton, R. D., Giustini, M., et al. 2019, Natur, 573, 381
Nicholl, M., Pasham, D. R., Mummery, A., et al. 2024, Natur, 634, 804
Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, A&A, 563, A54
Paturel, G., Dubois, P., Petit, C., & Woelfel, F. 2002, LEDA, 0
Payne, A. V., Shappee, B. J., Hinkle, J. T., et al. 2021, ApJ, 910, 125
Pérez-Díaz, V. S., Martínez-Galarza, J. R., Caicedo, A., & D'Abrusco, R. 2024, MNRAS, 528, 4852
Phinney, E. S. 1989, in IAU Symp. 136, The Center of the Galaxy, ed. M. Morris (Cambridge: Cambridge Univ. Press), 543
Plucinsky, P. P., Bogdan, A., & Marshall, H. L. 2022, Proc. SPIE, 12181, 121816X
Plucinsky, P. P., Bogdan, A., Marshall, H. L., & Tice, N. W. 2018, Proc. SPIE, 10699, 106996B
Quintin, E., Webb, N. A., Guillot, S., et al. 2023, A&A, 675, A152
Rees, M. J. 1988, Natur, 333, 523
Reines, A. E., & Volonteri, M. 2015, ApJ, 813, 82
Sacchi, A., Risaliti, G., & Miniutti, G. 2023, A&A, 671, A33
Saxton, R., Komossa, S., Auchettl, K., & Jonker, P. G. 2020, SSRv, 216, 85
Sazonov, S., Gilfanov, M., Medvedev, P., et al. 2021, MNRAS, 508, 3820
Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
Toba, Y., Oyabu, S., Matsuhara, H., et al. 2014, ApJ, 788, 45
van der Maaten, L., & Hinton, G. 2008, JMLR, 9, 2579, http://jmlr.org/papers/v9/vandermaaten08a.html
van Velzen, S., Gezari, S., Hammerstein, E., et al. 2021, ApJ, 908, 4
Verner, D. A., Ferland, G. J., Korista, K. T., & Yakovlev, D. G. 1996, ApJ, 465, 487
Véron-Cetty, M. P., & Véron, P. 2010, A&A, 518, A10
Webb, N. A., Coriat, M., Traulsen, I., et al. 2020, A&A, 641, A136
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Wevers, T., Coughlin, E. R., Pasham, D. R., et al. 2023, ApJL, 942, L33
Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914
Yang, H., Hare, J., Kargaltsev, O., et al. 2022, ApJ, 941, 104