MNRAS 000, 1–?? (2022) Preprint 6 August 2022 Compiled using MNRAS LATEX style file v3.0

QPOML: Leveraging Machine Learning to Detect and Characterize Quasi-Periodic Oscillations in X-Ray Binaries

Thaddaeus J. Kiker,[1]★, James F. Steiner[2], Cecilia Garraffo[2]
[1] Sunny Hills High School, Fullerton, CA 92833, USA
[2] Center for Astrophysics | Harvard & Smithsonian, 60 Garden St. Cambridge, MA 02138, USA

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Astronomy is presently experiencing radical growth with the deployment of machine learning in the field to reap the benefits provided by ever-growing massive datasets. ~~However, although the phenomena of Quasi-Periodic Oscillations characteristic to X-ray binaries has been extensively investigated over the last twenty (?) years, to date its origins retain a aura of mystery. Furthermore, this phenomena represents a vast wealth of data heretofore unexplored with machine learning, despite thorough documentation by tens (?) of thousands of detections from numerous space telescopes) (how machine learning could help shed some light on the subject?). In light of this, w~~e have developed a Python library, dubbed `QPOML`, to empower the community to make machine learning enabled discoveries about QPOs. In doing so, we propose and synthesize various best-practices and methods while demonstrating `QPOML`'s capabilities in the first ever detection and characterization of QPOs with machine learning.

Emphasize focus on interpretable Machine Learning

Key words: accretion, accretion disks — black hole physics — stars: individual (MAXI J1535+571) — X-rays: binaries

★ E-mail: thaddaeus@emailforpublicto.see

1 INTRODUCTION

At the ends of their lives, massive stars “do not go gentle into that good night” (Thomas 1952). Instead, if their initial mass exceeds that of ~ 8 $M_\odot$, core collapse leads to spectacular Type II supernovae (Shull 1995). Depending on this initial mass, such a conflagration event can result in a Neutron Star or Black Hole remnant (Gilmore 2004). In special cases, this object maintains a non-degenerate partner, and together these may form an X-Ray Binary (XRB) system, in which the non-degenerate star engages in mass-exchange with its compact partner (Tauris & van den Heuvel 2006). Such systems are characterized by accretion from the donor star, which leads to features like accretion disks (Shakura & Sunyaev 1973), jets (Gallo et al. 2005; van den Eijnden et al. 2018), and winds (Neilsen 2013; Castro Segura et al. 2022) associated with the compact component. Additional exotic phenomena like thermonuclear surface burning (Bildsten 1998) have also been observed with neutron stars, and although black holes are not physically capable of manifesting this latter process, many similarities still exist between XRBs of black hole and neutron star nature (Van Der Klis 1994), despite differences regarding predominant transience of black holes XRBs (Belloni & Motta 2016) versus widespread persistence of NS XRBs (Done et al. 2007). For example, these systems are both observed to emit thermal X-Ray radiation with temperatures ~ 1 keV that is understood to arise from the conversion of gravitational potential to radiative energy in an optically thick, geometrically thin accretion disk (Shakura & Sunyaev 1973). Black hole and neutron star XRBs both also show hard X-Ray flux coming from compton up-scattering of thermal disk flux by a cloud of electrons around the compact source known as the corona (Galeev et al. 1979; White & Holt 1982), and this component commonly modeled by a power law relationship $N(E) \propto E^{-\Gamma}$, where Γ is the photon index (McClintock & Remillard 2006). Additionally, spectra from both neutron stars and black holes exhibit reflection features like a fluorescent, relativistically broadened 6.4 keV Fe Kα line (Fabian et al. 1989) and ~ 30 keV compton hump (Ross & Fabian 2005). Furthermore, systems from both of these classes characteristically evolve through analogous spectral states (Gardenier & Uttley 2018), ranging from hard, to intermediate, and to soft (McClintock & Remillard 2006). These states are parameterized by changes in accretion rate and thus luminosity (Done & Gierliński 2004), spectral hardness or thermal dominance, and thereby position on a Hardness-Intensity or Color-Color Diagram track (Ingram & Motta 2019), and the presence/absence of mysterious Quasi-Periodic Oscillations (QPO) of the observed X-Ray radiation (McClintock & Remillard 2006), which are detected as narrow ($Q = \nu/2\Delta \geq X$) peaks in power-density spectra (Homan & Belloni 2005).

Over the past thirty years, numerous theories, including but not limited to relativistic precession (Stella & Vietri 1998), precesssing inner flow (Ingram et al. 2009), corrugation modes (Kato & Fukue 1980), accretion ejection instability (Tagger & Pellat 1999), and propagating oscillatory shock (Molteni et al. 1996) have been advanced to explain the occurrence of QPOs in NS and BH XRB systems, yet providing a conclusive and let alone universal explanation for low frequency QPOs (LFQPO) and QPOs in general has proven to be a very challenging endeavor. For context, in black hole systems QPOs are most frequently observed as LFQPOs with centroid frequencies ranging from 0.1-30 Hz (Belloni et al. 2020), yet some BH-XRBs have exhibited high-frequency QPOs (HFQPO). The former is further subdivided canonically into three classes ~~of alphabetic flavor~~ (Casella et al. 2005): Type-A QPOs are the rarest, ~~sometimes~~ appearing in the soft state as broad, low amplitude features centered between 6-9 Hz and usually lacking harmonic companions (Ingram & Motta 2019). Type-B QPOs are more common, and can be seen during the short Soft Intermediate State and have shown some connection with Jet behavior (Gao et al. 2017). Finally, type C QPOs are the most common, and can be detected with harmonic companions as narrow (Q>8) features in nearly all states (Fragile et al. 2016). Their centroid frequencies range from 0.1-30 Hz depending on state, and almost always correlate strongly with spectral features like Γ (Motta et al. 2015). As for the rare latter HFQPO phenomena, HFQPOs exhibit centroid frequencies $\nu \geq 100$ Hz (Motta et al. 2011), often in resonant pairs, and are of additional particular interest given the fact that their timescales show similarity to Keplerian orbital periods near/at the Innermost Stable Circular Orbit in accretion disks (Stella & Vietri 1999) and that they can be used to calculate black hole spin (Abramowicz & Kluźniak 2001).

QPOs are also observed ~~as low- and high-frequency signals~~ in neutron star systems (Wang 2016). Neutron star LFQPOs are further sub-classified as horizontal branch oscillations, normal branch oscillations and flaring branch oscillations, which together correlate with spectral states while ranging in frequency from 1-70 Hz, broadly corresponding to the different classes of LFQPOs in black hole systems (Homan et al. 2015). On the other hand, HFQPOs in neutron stars, known as kilohertz or kHz QPOs, range in frequency from 200 - 1300 Hz and represent the fastest variability components in XRBs (Wang 2016). ~~When kHz QPOs are detected in pairs, the features are distinguished as the lower or higher kHz QPO based on frequency (Belloni et al. 2005).~~ Notably, the similarities between black hole HFQPOs and neutron star kHz QPOs are suggestive of common origin (Psaltis et al. 1999; Bhargava et al. 2021). For further discussion of QPOs in XRBs, including discussion of 1 Hz and hectohertz QPOs in Neutron Stars, as well as mHz QPOs, we recommend readers to Ingram & Motta (2019), Jonker et al. (1999), Kato (2005), and Revnivtsev et al. (2001), ~~respectively~~.

Though difficult at times, studying QPOs and their host XRB systems with instruments like RXTE (Levine et al. 1996), Chandra (Weisskopf et al. 2000), and others has proven to be a very fruitful endeavor, for beyond answering numerous important questions about these systems themselves (and raising or re-contextualizing yet many more), such studies have also provided deep insights into our understanding of the universe on a fundamental level, particularly through investigations of general relativity in strong gravity (van der Klis 2004; Maselli et al. 2017). Furthermore, among other insights, the study of XRBs has lead to a deeper understanding of accretion physics relevant to its varied manifestation from young stellar objects to white dwarfs (Done et al. 2007; Scaringi et al. 2015).

All in all, hundreds of XRBs have been observed since the discovery of Cygnus X-1 (Bolton 1972; Liu et al. 2007; Corral-Santana, J. M. et al. 2016). Although machine learning has been used to classify accretion states (Sreehari & Nandi 2021), predict compact object identity (Pattnaik et al. 2021), and study gravitational waves (Schmidt et al. 2021), this represents tens of thousands of observations that have heretofore not been explored with machine learning to the ends of detecting QPOs. Therefore, in this work we seek to develop a methodology for using machine learning to detect QPO ~~for two reasons: first,~~ we believe that within the context of astrophysics, our theoretical understanding of QPOs and their exotic progenitor systems could benefit from insights derived from this ~~non-traditional~~ approach (Fudenberg & Liang 2020). ~~Perhaps machine learning may contribute to a conclusive understanding of this decades old mystery, and in contributing to the potential solution to this mystery, we anticipate the fields of astrophysics and physics in their entirety to benefit (Karniadakis et al. 2021; Meng et al. 2022). Second, we believe that as a field, computer science itself can benefit greatly from~~

Table 1. Description of Sources Included in this Study

Source	Class	Instrument	Number of Observations
MAXI J1535-571	BH	NICER	658
GRS 1915+105	BH	RXTE	620
Total			1278

Figure 1. Example light curve...include a cool NICER heatmap looking plot too?..include an additional "cartoon" XRB for STS version?

~~the utilization of such a rich data treasure-trove for further technical development (Zhu et al. 2016; Roh et al. 2018). In doing this, we strive to adhere to the standards of interpretable machine learning (Murdoch et al. 2019; Belle & Papantonis 2021; Molnar 2022), and we also employ techniques from other fields like game theory to frame our work in an interdisciplinary context.~~

We structure the remainder of this paper as follows: in Section 2 we describe the observations and data analysis upon which we base our work. Following this, in Section 3 we introduce our methods and models, after which we present their results in Section 4. We discuss these results contextually in Section 5, and finally, we conclude in Section 6. Additional work is presented in following appendices.

2 OBSERVATIONS AND DATA ANALYSIS

The energy spectra were fit with `XSPEC` version 12.12.0 using the model `tbabs*(diskbb+nthcomp)`, which represents an absorbed multi-temperature blackbody and thermally Comptonized continuum (Cúneo et al. 2020; Mitsuda et al. 1984; Kubota et al. 1998; Zdziarski et al. 1996; Życki et al. 1999). We fixed the equivalent hydrogen column densities to canonical values from the literature, tied seed photon temperature to $T_{\rm in}$ (of `diskbb`), and computed hardness values as the ratio of the background-subtracted count rates in 4-10 keV / 2-4 keV bands, after subtracting the sum of the ~~temporally scaled~~ background counts in these ranges. ~~Unfortunately, as in Miller et al. (2018), we~~ noticed severe instrumental residuals in the $1.7 - 2.1$ keV range, likely related to NICER's Au mirror coating and residual in the Si K α fluorescence peak, and similarly, we excluded the $1.5 - 2.3$ keV energy band from the spectral fitting process.

The Lorentzians are parameterized as in Equation 1, where ν is frequency in Hertz, σ is full width at half maximum (FWHM), and K is normalization (Arnaud, Gordon & Dorman Arnaud et al.).

We test both raw spectrum input as well as feature vector inputs, as we are interested in seeing ...

$$A(\nu) = \frac{K(\frac{\sigma}{2\pi})}{(\nu - \nu_L)^2 + (\frac{\sigma}{2})^2} \quad (1)$$

2.1 Neutron Star Sources

discuss atoll vs z sources

2.2 Black Hole Sources

Low Mass XRBs (LMXRBs) are defined

3 METHODS

3.1 Model Selection

In machine learning, models can be broadly classified by two categories: whether they are built for classification or regression, and whether they operate in a supervised or unsupervised manner (Bruce & Bruce 2017). ~~Since we are providing our models with explicit targets for loss minimization, o~~ur approach falls under the umbrella of supervised learning (Singh et al. 2016), and as we are attempting to connect spectral information about XRBs with real-valued output vectors that describe QPOs in their power-density spectra, we also fall under multi-output regression (Xu et al. 2019). In selecting our machine learning models for regression, we sought those that natively supported multi-output regression, incorporated capabilities for mitigating overfitting, had precedents of working successfully with medium to small sized data sets, and natively communicated feature importances. Additionally, we sought to evaluate a collection of models against each other in light of the No-Free-Lunch-Theorem (NoF NoF; Lones 2021)

Based on these criteria, we settled on a set of tree-based regression models and their descendants, specifically Decision Trees (Breiman 1984), Random Forests (Breiman 2001), Extremely Randomized Trees (Geurts et al. 2006), and XGBoost (Chen & Guestrin 2016). Decision Trees are the original tree-based regression model and they operate by inferring discriminative splits in data and making predictions via a series of if-then-else decisions (Breiman 1984). Random Forests are more powerful derivatives of Decision Trees, and are based on an ensemble of Decision Trees trained via bootstrap aggregation (Breiman 1996, 2001). By incorporating predictions from such an ensemble, Random Forests reduce prediction variance while increasing overall accuracy when compared to a single Decision Tree (Lakshminarayanan 2016). Extremely Randomized Trees (also known as Extra Trees) are similar to Random Forests in this respect but operate with more randomization during the training process, as instead of employing the most discriminative thresholds within feature spaces for splits, Extremely Randomized Trees select the best performing randomly drawn thresholds for splitting rules (Geurts et al. 2006; Pedregosa et al. 2011). Finally, XGBoost builds upon stochastic gradient boosting for improved performance in terms of speed and efficiency compared to its predecessors like AdaBoost (Chen & Guestrin 2016; Azmi & Baliga 2020). One important distinction between XGBoost and Random Forests/Extremely Randomized Trees is that only the former employs boosting, which is the practice of successively fitting models to training cases with large errors (Friedman 2002), and consequently, in the absence of proper hyperparameter optimization, stands at a greater risk of overfitting, hence Section 3.4 (Bruce & Bruce 2017).

Together, these represent some of the most powerful yet lightweight machine learning models available, and meet our criteria for multi-output regression (Xu et al. 2019), robustness to overfitting (Boinee et al. 2008; Ampomah et al. 2020), success with small/medium sized datasets (Floares et al. 2017), and feature importances (Yasodhara et al. 2021). An additional benefit of these models is that they are natively supported by the `TreeExplainer` method in the `SHAP` Python package (Lundberg & Lee 2017), which frees us from common pitfalls related to impurity and permutation based feature importances, which we discuss in more detail in Section 5.

3.2 Feature Selection

During feature selection, it is generally important to deal with potential multicollinearity by calculating Variance Inflation Factors (VIF) and removing features with VIF values $\gtrsim 5$ (Kline 1998; Sheather 2008). However, we have chosen not to remove potentially collinear features prior to regression for the following reasons. First, a tree based model like Random Forest (on which we focus) is by design robust from the effects of multicollinearity (Strobl et al. 2008; Chowdhury et al. 2021). Second, since multicollinearity only affects the estimated coefficients of linear models, but not their predictive ability, applying a linear model to potentially collinear data is perfectly reasonable in our case, as we are using the linear model solely as a baseline against which we will compare the predictive capabilities of the more complicated Random Forest model; i.e. as we are applying the linear model, we are not interested in its components (Lieberman & Morris 2014; Mundfrom et al. 2018). We will, however, revisit multicollinearity when we interpret feature importances in Section **??**.

3.3 Feature Engineering

As Casari & Zheng (2018) detail, feature engineering refers to the crucial process of transforming raw data to maximize predictive performance. After experimenting with different formats, we settled on the following for using fitted `XSPEC` and raw spectral data as predictors and timing features as outcomes.

When using scalar values for inputs, we format our input data as a ~~two dimensional tensor (i.e.~~ matrix~~)~~ composed of vectors containing the hardness ratio, $T_{\rm in}$, Γ, `nthcomp` normalization, `diskbb` normalization, and net count rate values for every observation. This input structure is visualized in Equation 2 (where h is hardness, nn is `nthcomp` normalization, dn is `diskbb` normalization, and ncr is net count rate):

$$\mathrm{IN}_{n\times 6} = \begin{bmatrix} T_{\mathrm{in}_1} & \Gamma_1 & h_1 & nn_1 & dn_1 & ncr_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{\mathrm{in}_n} & \Gamma_n & h_n & nn_n & dn_n & ncr_n \end{bmatrix} \tag{2}$$

Similarly, when working with raw spectral data, we re-bin the energy spectra into 25 equal-width channels in the energy range **update**, similar to Pattnaik et al. (2020), and use the 25 derived count rate values directly as the input vectors within the input tensor.

The QPO output ~~tensor~~ is similarly formatted as a two dimensional tensor composed of vectors that match by index to vectors in the input ~~tensor~~, but with a twist. A significant challenge related to the prediction of not only the presence of QPOs in a given PDS, but also the number of QPOs, as well as the physical parameters of each QPO present, is that over the course of an outburst (or multiple), the number of detectable QPOs is known to change (cite two). We account for this challenge of variable output cardinality by first identifying all QPO occurrences associated with an observation. Then, we order these occurrences and their features in a vector of length $L = N_f \times \max(N_s)$, where N_f is the number of features describing every QPO (e.g. $N_f = 2$ for frequency and width), and N_s is the maximum number of simultaneous QPOs observed in a PDS in the dataset.

We then structure each output vector as a repeating subset of features for every QPO contained, and order these internal QPO parameterizations with the sub-harmonic ~~features first, followed by the fundamental features, which are finally followed by the first harmonic features~~. If one or more of these occurrences are not detected in a PDS, their feature spaces in the vector are populated with zeros. This allows us to circumvent the aforementioned issue with variable output cardinality, because as our models will learn during training to associate indices populated with zeros as ~~being QPO~~ non-detections (Chollet 2017), Hence the prepossessing results in a feature space spanning [0.1, 1.0], rather than [0.0, 1.0].

~~As in the case of input features,~~ Equation 3 provides a visualization of a potential output matrix returned by our model, where each row corresponds to one observation matched with a row in the input matrix (both out of n total observations).

$$\mathrm{OUT}_{n\times 9} = \begin{bmatrix} \nu_{1,1} & \Delta_{1,1} & a_{1,1} & \cdots & \nu_{1,3} & \Delta_{1,3} & a_{1,3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \nu_{n,1} & \Delta_{n,1} & a_{n,1} & \cdots & \nu_{n,3} & \Delta_{n,3} & a_{n,3} \end{bmatrix} \tag{3}$$

In this example, the maximum number of QPOs simultaneously observed in a PDS is three, and each QPO is described in terms of its frequency, width, and amplitude, so the output matrix takes the shape OUT $= n\times 9$. Prior to reformatting the data in this manner, we applied a min-max standardization to the `XSPEC`, and hardness input features, as well as the QPO lorentzian output features, which linearly transformed each distribution into a [0.1, 1] range while preserving their shapes, according to Equation 4 (Kandanaarachchi et al. 2019).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{4}$$

This step is necessary to prevent features with larger ranges receiving greater weight than they necessarily deserve, and it also frees the models from dependency on measurement units (Akanbi et al. 2015; Han et al. 2012). We did not apply this standardization step to channel count and net count rate input features, however, as the imposition of *a priori* ~~theoretical~~ limits to these features is not as readily justifiable (Pattnaik et al. 2020). [1]

3.4 Training, Validation, and Hyperparameter Tuning

To better understand our models in different data combinations and minimize statistical noise, while guaranteeing every observation gets included in a training, as well as at a separate time, test instance, we employ a repeated k-fold cross-validation strategy (Vanwinckelen & Blockeel 2012) for model evaluation (as opposed to a default

[1] Standardization prior to splitting data into train and test sets does not impair our model's predictive validity because its pre-adjusted inputs will always be constrained within the ~~theoretical~~ bounds applied during standardization for each feature (e.g. Γ will always initially range between $x - y$).

Figure 2. A flowchart visualization of our QPOML routine method.

proportion-based train-test split). ~~According to this procedure,~~ our data is ~~first~~ randomly split into $k = 10$ folds. Then, every model is evaluated on each unique fold ~~thus generated~~ after being trained on the remaining folds, with the individual k-fold performance taken as the mean of these evaluations across the ten folds. We repeat this process ten times (randomly shuffling the data between each iteration), and the final performance for each model is calculated as the mean performance across the ten k-fold instances (Kuhn & Johnson 2019). To ensure fair comparison between these algorithms, each underwent automatic and individualized hyperparameter tuning via grid search prior to this evaluation (Dangeti 2017). In the case of XGBoost, for example, this included modulation of learning rate η, ℓ_1 and ℓ_2 regularization, the number of boosting iterations, and maximum tree depth to minimize overfitting and maximize predictive performance (see Appendix B. for complete discussion).

4 RESULTS

5 DISCUSSION

5.1 Feature Importances and Interpretation

Although it is common to discuss default impurity-based feature importances, this approach is flawed as these are both biased towards high-cardinality ~~numerical input~~ features, as well as computed on training set statistics, which means they may not accurately generalize to held-out data (Pedregosa et al. 2011). Additionally, although permutation importances are often forwarded as a superior alternative, these suffer from multicollinearity, as in the process of permutating single features a feature can be erroneously calculated as having little to no effect on model performance because its information can be inferred from a different, correlated feature (Strobl et al. 2007; Nicodemus et al. 2010; Hooker et al. 2019). Therefore, we chose to to determine feature importances for our XXXXX with the state of the art TreeSHAP algorithm as implemented in the Python package `shap` by Lundberg & Lee (2017). This extends game theoretic coalitional Shapley values to calculate SHapley Additive exPlanations (SHAP) in the presence of multicollinearity by incorporating conditional ex-

Figure 3. Violin plot of predictive performance for the source ******* across $k = 10$ validation folds repeated 10 times.

Figure 4. Example detector spectrum and PDS for the source ***** with predicted QPOs overplotted.

Figure 5. A results regression plot for all QPOs from a given fold for the source *****, as predicted by ordinary linear regression (left) and ***** (right). As this demonstrates, linear regression ...

Figure 6. Feature importances for the ***** model predicting for the source ********* derived from the absolute value of the tree-shape calculated SHAP values.

pected predictions (Shapley 1952; Lundberg & Lee 2017; Molnar 2022). As hinted earlier and detailed in Lundberg & Lee (2017) and Molnar (2022), an additional benefit of using Tree based models in this context is that through tree traversal and dynamic programming the computational cost for computing SHAP values is brought down from exponential $O(2^n)$ ~~time inherent to the brute force approach (defined in Equation 6)~~ to polynomial time $O(n^2)$.

To determine global feature importances, we take the average of the absolute values per feature throughout the data, according to Equation 5,

$$I_j = \frac{1}{n}\sum_{i=1}^{n}|\phi_j^{(i)}| \tag{5}$$

where TreeSHAP estimates $\phi_i(f,x)$ from Equation 6.

$$\phi_i(f,x) = \sum_{z'\subseteq x'} \frac{|z'|!(M-|z'|-1)!}{M!}\left[f_x(z') - f_x(z' \setminus i)\right], \tag{6}$$

which represents the weighted average of differences in model performance when a feature is present and absent for all subsets $z' \subseteq x'$. *discuss feature importances, but note distinction between correlation and causation and caution in this context...maybe in conclusion*

5.2 Statistical Model Comparison

As mentioned in Section 3, we included an ordinary least squares model in addition to the others to serve as a justification for their utilization. As Figure 3 and Figure 5 demonstrate, every model ~~vastly~~ outperforms linear regression, suggesting their use is justified. ~~However, this is not adequate grounds for conclusion, and we therefore~~ turn to more sophisticated methods to determine whether or not these differences in performance are significant. To do so, we first employ the Nadeau & Bengio (2004) formulation of the frequentist Diebold-Mariano corrected paired t-test (Diebold & Mariano 1995)~~, which is shown in Equation 7~~:

$$t = \frac{\frac{1}{k\cdot r}\sum_{i=1}^{k}\sum_{j=1}^{r} x_{ij}}{\sqrt{\left(\frac{1}{k\cdot r} + \frac{n_{\text{test}}}{n_{\text{train}}}\right)\hat{\sigma}^2}} \tag{7}$$

where $k = 10$ and represents the ~~number of~~ folds in our aforementioned k-fold validation, $r = 10$ and equals the number of times we repeated the k-fold procedure, x is the performance difference between two models, and $\hat{\sigma}^2$ represents the variance of these differences (Pedregosa et al. 2011). It is necessary to correct the $t-$values in this manner because the performance of the models are correlated with each fold upon which they are tested, as some folds may make it harder for one of or all of the models to generalize, whereas others make it easier, and thus the collective performance of the models co-vary. Based on this test, we are able to calculate whether or not a model is significantly better than another. The results of these pairwise tests for all permutations of two model combinations on XXX sources is shown in Table 2.

Table 2. Pairwise ... p is the ... A full, machine-readable, version of this table is available electronically.

Source	M_1	M_2	p	% M_1 Better
nan	nan	nan	nan	nan
nan	nan	nan	nan	nan

Following this, we implement the Bayesian Benavoli et al. (2016) approach, which allows us to calculate the *probability* that a given model is better than another~~, using the distribution formulated in Equation (8)~~:

$$\text{St}(\mu; n-1, \bar{x}, (\frac{1}{n} + \frac{n_{\text{test}}}{n_{\text{train}}})\hat{\sigma}^2) \tag{8}$$

where n is the total number of samples, $\bar{x}$ is the mean score difference, and $\hat{\sigma}^2$ is the Nadeau & Bengio (2004) corrected variance in differences (Pedregosa et al. 2011). These results of these pairwise tests are also shown in Table 2.

6 CONCLUSION

Although we focused our validation on black hole LFQPOs and neutron star kHz QPOs in low mass X-Ray binary systems ... *generalize*

We note that in `QPOML` our entire methodological process, from input and output tensor construction and prepossessing, to hyperparameter tuning, model evaluation, and plot generation, is conveniently both 1. executed internally "under the hood" (the user can implement all these steps in less than five lines of code, as shown in the Appendix A.), and 2. extendable to any number of QPOs and any number of scalar observation features for any number of observations.

ML (read augment physics papers) should clearly not be a replacement for traditionally modeling...we hope it augments creative

Future applications/extensions:

- QPO phase lags
- HMXRBs?
- spin/mass info (our models are easily extensible from)
-

We are currently working on a follow up piece to this paper in which we are training a hopefully inter-source and inter-class ... welcome collaborators ... assembling a large standardized database of QPO and spectral data ...

7 ACKNOWLEDGEMENTS

We thank V.A. Cuneo for helpful conversation early in this work. We thank (two guys at XSPEC who helped me) for assistence with XSPEC issues. We thank Travis Austen with help recovering lost

... from damaged virtual machine disk, and Brandon Barrios for thereafter recommending the Windows Subsystem for Linux (WSL) alternative.

Facilities:

Software:

REFERENCES

Abramowicz M. A., Kluźniak W., 2001, A&A, 374, L19
Akanbi O. A., Amiri I. S., Fazeldehkordi E., 2015, in , A Machine-Learning Approach to Phishing Detection and Defense. Elsevier, pp 45–54, doi:10.1016/b978-0-12-802927-5.00004-6, https://doi.org/10.1016/b978-0-12-802927-5.00004-6
Ampomah E. K., Qin Z., Nyame G., 2020, Information, 11
Arnaud K., Gordon C., Dorman B., , An X-Ray Spectral Fitting Package, https://heasarc.gsfc.nasa.gov/xanadu/xspec/manual/XspecManual.html
Azmi S. S., Baliga S., 2020.
Belle V., Papantonis I., 2021, Frontiers in Big Data, 4
Belloni T. M., Motta S. E., 2016, in Bambi C., ed., Astrophysics and Space Science Library Vol. 440, Astrophysics of Black Holes: From Fundamental Aspects to Latest Developments. p. 61 (arXiv:1603.07872), doi:10.1007/978-3-662-52859-4_2
Belloni T., Méndez M., Homan J., 2005, A&A, 437, 209
Belloni T. M., Zhang L., Kylafis N. D., Reig P., Altamirano D., 2020, MNRAS, 496, 4366
Benavoli A., Corani G., Demsar J., Zaffalon M., 2016, arXiv e-prints, p. arXiv:1606.04316
Bhargava Y., Belloni T., Bhattacharya D., Motta S., Ponti. G., 2021, MNRAS, 508, 3104
Bildsten L., 1998, in Buccheri R., van Paradijs J., Alpar A., eds, NATO Advanced Study Institute (ASI) Series C Vol. 515, The Many Faces of Neutron Stars.. p. 419 (arXiv:astro-ph/9709094)
Boinee P., Angelis A. D., Foresti G. L., 2008, International Journal of Computer and Information Engineering, 2, 2246
Bolton C. T., 1972, Nature, 235, 271
Breiman L., 1984, Classification and Regression Trees. (The Wadsworth statistics / probability series), Wadsworth International Group, https://books.google.com/books?id=ZxPvAAAAMAAJ
Breiman L., 1996, Machine learning, 24, 123
Breiman L., 2001, Machine learning, 45, 5
Bruce P., Bruce A., 2017, Practical Statistics for Data Scientists: 50 Essential Concepts. O'Reilly Media, https://books.google.com/books?id=ldPTDgAAQBAJ
Casari A., Zheng A., 2018, O'Reilly Media, Inc., p. 218
Casella P., Belloni T., Stella L., 2005, ApJ, 629, 403
Castro Segura N., et al., 2022, Nature, 603, 52
Chen T., Guestrin C., 2016, arXiv e-prints, p. arXiv:1603.02754
Chollet F., 2017, Deep Learning with Python. Manning
Chowdhury S., Lin Y., Liaw B., Kerby L., 2021, arXiv e-prints, p. arXiv:2111.02513
Corral-Santana, J. M. Casares, J. Muñoz-Darias, T. Bauer, F. E. Martínez-Pais, I. G. Russell, D. M. 2016, A&A, 587, A61
Cúneo V. A., et al., 2020, MNRAS, 496, 1001
Dangeti P., 2017, Statistics for Machine Learning. Packt Publishing, https://books.google.com/books?id=C-dDDwAAQBAJ
Diebold F. X., Mariano R. S., 1995, Journal of Business Economic Statistics, 13, 253
Done C., Gierliński M., 2004, Progress of Theoretical Physics Supplement, 155, 9
Done C., Gierliński M., Kubota A., 2007, A&ARv, 15, 1
Fabian A. C., Rees M. J., Stella L., White N. E., 1989, MNRAS, 238, 729
Floares A., Ferisgan M., Onita D., Ciuparu A., Calin G., Manolache F., 2017, Int J Oncol Cancer Ther, 2, 13
Fragile P. C., Straub O., Blaes O., 2016, MNRAS, 461, 1356
Friedman J. H., 2002, Computational Statistics & Data Analysis, 38, 367
Fudenberg D., Liang A., 2020, SIGecom Exch., 18, 4–11
Galeev A. A., Rosner R., Vaiana G. S., 1979, ApJ, 229, 318
Gallo E., Fender R., Kaiser C., 2005, in Burderi L., Antonelli L. A., D'Antona F., di Salvo T., Israel G. L., Piersanti L., Tornambè A., Straniero O., eds, American Institute of Physics Conference Series Vol. 797, Interacting Binaries: Accretion, Evolution, and Outcomes. pp 189–196 (arXiv:astro-ph/0501374), doi:10.1063/1.2130232
Gao H. Q., et al., 2017, MNRAS, 466, 564
Gardenier D. W., Uttley P., 2018, MNRAS, 481, 3761
Geurts P., Ernst D., Wehenkel L., 2006, Mach. Learn., 63, 3–42
Gilmore G., 2004, Science, 304, 1915
Han J., Kamber M., Pei J., 2012, in , Data Mining. Elsevier, pp 83–124, doi:10.1016/b978-0-12-381479-1.00003-4, https://doi.org/10.1016/b978-0-12-381479-1.00003-4
Homan J., Belloni T., 2005, Ap&SS, 300, 107
Homan J., Fridriksson J. K., Remillard R. A., 2015, ApJ, 812, 80
Hooker G., Mentch L., Zhou S., 2019, arXiv e-prints, p. arXiv:1905.03151
Ingram A., Motta S. E., 2019, New Astronomy Reviews
Ingram A., Done C., Fragile P. C., 2009, MNRAS, 397, L101
Jonker P. G., van der Klis M., Wijnands R., 1999, ApJ, 511, L41
Kandanaarachchi S., Muñoz M. A., Hyndman R. J., Smith-Miles K., 2019, Data Mining and Knowledge Discovery, 34, 309
Karniadakis G. E., Kevrekidis I. G., Lu L., Perdikaris P., Wang S., Yang L., 2021, Nature Reviews Physics, 3, 422
Kato S., 2005, PASJ, 57, L17
Kato S., Fukue J., 1980, PASJ, 32, 377
Kline R., 1998, Principles and Practice of Structural Equation Modeling. Methodology in the Social Sciences, Guilford Publications, https://books.google.com/books?id=JGVuQgAACAAJ
Kubota A., Tanaka Y., Makishima K., Ueda Y., Dotani T., Inoue H., Yamaoka K., 1998, PASJ, 50, 667
Kuhn M., Johnson K., 2019, Applied Predictive Modeling. Springer New York, https://books.google.com/books?id=L_gewAEACAAJ
Lakshminarayanan B., 2016, PhD thesis, UCL (University College London)
Levine A. M., Bradt H., Cui W., Jernigan J. G., Morgan E. H., Remillard R., Shirey R. E., Smith D. A., 1996, ApJ, 469, L33
Lieberman M., Morris J., 2014, 40, 5
Liu Q. Z., van Paradijs J., van den Heuvel E. P. J., 2007, A&A, 469, 807
Lones M. A., 2021, arXiv e-prints, p. arXiv:2108.02497
Lundberg S. M., Lee S.-I., 2017, in Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., eds, Vol. 30, Advances in Neural Information Processing Systems. Curran Associates, Inc., https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
Maselli A., Gualtieri L., Ferrari V., Pani P., Stella L., 2017, in 14th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories. pp 1850–1853, doi:10.1142/9789813226609_0193
McClintock J. E., Remillard R. A., 2006, in , Vol. 39, Compact stellar X-ray sources. pp 157–213
Meng C., Seo S., Cao D., Griesemer S., Liu Y., 2022, arXiv e-prints, p. arXiv:2203.16797
Miller J. M., et al., 2018, The Astrophysical Journal, 860, L28
Mitsuda K., et al., 1984, PASJ, 36, 741
Molnar C., 2022, Interpretable Machine Learning, 2 edn. https://christophm.github.io/interpretable-ml-book
Molteni D., Sponholz H., Chakrabarti S. K., 1996, ApJ, 457, 805
Motta S., Muñoz-Darias T., Casella P., Belloni T., Homan J., 2011, Monthly Notices of the Royal Astronomical Society, 418, 2292
Motta S. E., Casella P., Henze M., Muñoz-Darias T., Sanna A., Fender R., Belloni T., 2015, MNRAS, 447, 2059
Mundfrom D., Smith M., Kay L., 2018, General Linear Model Journal, 44, 24
Murdoch W. J., Singh C., Kumbier K., Abbasi-Asl R., Yu B., 2019, Proceedings of the National Academy of Sciences, 116, 22071
Nadeau C., Bengio Y., 2004, Machine Learning, 52, 239
Neilsen J., 2013, Advances in Space Research, 52, 732

Nicodemus K. K., Malley J. D., Strobl C., Ziegler A., 2010, BMC Bioinformatics, 11

Pattnaik R., Sharma K., Alabarta K., Altamirano D., Chakraborty M., Kembhavi A., Méndez M., Orwat-Kapola J. K., 2020, Monthly Notices of the Royal Astronomical Society, 501, 3457

Pattnaik R., Sharma K., Alabarta K., Altamirano D., Chakraborty M., Kembhavi A., Méndez M., Orwat-Kapola J. K., 2021, MNRAS, 501, 3457

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Psaltis D., Belloni T., van der Klis M., 1999, ApJ, 520, 262

Revnivtsev M., Churazov E., Gilfanov M., Sunyaev R., 2001, A&A, 372, 138

Roh Y., Heo G., Euijong Whang S., 2018, arXiv e-prints, p. arXiv:1811.03402

Ross R. R., Fabian A. C., 2005, MNRAS, 358, 211

Scaringi S., et al., 2015, Science Advances, 1, e1500686

Schmidt S., et al., 2021, Phys. Rev. D, 103, 043020

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

Shapley L. S., 1952, A Value for N-Person Games. RAND Corporation, Santa Monica, CA, doi:10.7249/P0295

Sheather S. J., 2008, A modern approach to regression with R, 2009 edn. Springer Texts in Statistics, Springer, New York, NY

Shull J. M., 1995, in Haas M. R., Davidson J. A., Erickson E. F., eds, Astronomical Society of the Pacific Conference Series Vol. 73, From Gas to Stars to Dust. pp 365–386

Singh A., Thakur N., Sharma A., 2016, in 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom). pp 1310–1315

Sreehari H., Nandi A., 2021, MNRAS, 502, 1334

Stella L., Vietri M., 1998, ApJ, 492, L59

Stella L., Vietri M., 1999, Phys. Rev. Lett., 82, 17

Strobl C., Boulesteix A.-L., Zeileis A., Hothorn T., 2007, BMC Bioinformatics, 8

Strobl C., Boulesteix A.-L., Kneib T., Augustin T., Zeileis A., 2008, BMC Bioinformatics, 9

Tagger M., Pellat R., 1999, A&A, 349, 1003

Tauris T. M., van den Heuvel E. P. J., 2006, in , Vol. 39, Compact stellar X-ray sources. pp 623–665

Thomas D., 1952, In Country Sleep: And Other Poems. New Directions book, James Laughlin

Van Der Klis M., 1994, Neutron Stars and Black Holes in X-Ray Binaries. Springer Netherlands, Dordrecht, pp 265–275, doi:10.1007/978-94-011-0794-5_27, https://doi.org/10.1007/978-94-011-0794-5_27

Vanwinckelen G., Blockeel H., 2012.

Wang J., 2016, International Journal of Astronomy and Astrophysics, 6, 82

Weisskopf M. C., Tananbaum H. D., Van Speybroeck L. P., O'Dell S. L., 2000, in Truemper J. E., Aschenbach B., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4012, X-Ray Optics, Instruments, and Missions III. pp 2–16 (arXiv:astro-ph/0004127), doi:10.1117/12.391545

White N. E., Holt S. S., 1982, ApJ, 257, 318

Xu D., Shi Y., Tsang I. W., Ong Y.-S., Gong C., Shen X., 2019, arXiv e-prints, p. arXiv:1901.00248

Yasodhara A., Asgarian A., Huang D., Sobhani P., 2021, arXiv e-prints, p. arXiv:2110.00086

Zdziarski A. A., Johnson W. N., Magdziarz P., 1996, MNRAS, 283, 193

Zhu X., Vondrick C., Fowlkes C. C., Ramanan D., 2016, International Journal of Computer Vision, 119, 76

Życki P. T., Done C., Smith D. A., 1999, MNRAS, 309, 561

van den Eijnden J., Degenaar N., Russell T. D., Wijnands R., Miller-Jones J. C. A., Sivakoff G. R., Hernández Santisteban J. V., 2018, Nature, 562, 233

van der Klis M., 2004, in Kaaret P., Lamb F. K., Swank J. H., eds, American Institute of Physics Conference Series Vol. 714, X-ray Timing 2003: Rossi and Beyond. pp 371–378, doi:10.1063/1.1781057

[H]

```
from qpoml import qpo, context

context = None

qpo_csv_url = "https://www.overleaf.com/"

context_csv_url = "https://www.overleaf.com/"

context.load(qpo_csv)
context.evaluate(context_csv, model="random-forest")

context.plot_correlation_matrix(dendrogram=True)
context.plot_results_regression()
context.plot_feature_importances()
```

Figure 7. A demonstration of `QPOML` which replicates three of the figures from the main paper.

APPENDIX

A. `QPOML` Demonstration

Contents: one figure (code), about one page long (in text discuss all the capabilities QPOML has and the things it handles under the hood)

B. Methodological Experimentation

7.0.1 Hyperparameter Tuning and Grid Search

Content: eight figures, eighteen pages

7.0.2 `eurostep` *Architecture*

During the process of model and configuration exploration, we also envisioned a completely different ensemble architecture, which we dubbed `eurostep` after the (in)famous two-step move in Basketball, and we conclude this subsection with a brief summary. In contrast to the one model approach we used elsewhere (which handled the amount of QPOs to predict via zero padding), the eurostep approach operates across two phases: a primary classification step followed by a secondary regression step. For any given train-test split, the classification model is trained on all available training data; however, individual regressors are trained for each potential outcome; i.e. if the observations in a dataset only have two or three QPO occurrences per PDS, one regressor would be trained on the training instances with one QPO occurrence, and the other regressor would be trained on the training instances with two QPO occurrences. In practice, during the first step the classifier would predict how many QPO occurrences are present for a observation, and then based on this prediction, the algorithm would select which regressor to use. Although this architecture provides an alternative to zero-padding as a solution to the variability in the cardinality of the multiple-output feature space, we concluded that this benefit is outweighed by `eurostep`'s much more complicated nature.

C. Injected QPO Recovery

Content: one figure (two part: left is simulated qpo example, right is something CNN related) and one page total (text is about CNN, etc.)

Luckily, one advantage allotted to recovery of QPOs from power density spectra is that PDS are (or can be) identically structured in frequency space between observations, in contrast to time series, which are often sparsely and non-uniformly sampled. Thus, an entirely different approach can be taken to detect QPOs from PDS if PDS are treated as images, and a model is trained to detect these features from raw PDS unaided. An appropriate model for this case would be a convolutional neural network, which **DESCRIBE** . One draw back of using a CNN for this purpose is that CNNs require **NUMBER** of data for training, and there are not enough QPO observations for any singular source to adequately meet this requirement. However, in this appendix we take a page from exoplanet science and "inject" QPOs into XXXX PDS as an initial proof-of-concept exploration to see if the approach is tenable, and if so, open the door to a future exploration which would investigate the merits of training a CNN on Monte-Carlo generated PDS and QPOs tailored to a configurations of a particular detector, and then attempt to recover real QPOs.

NOTES

We account for X-Ray absorption on the part of the Interstellar Medium with the Tuebingen-Boulder ISM absorption model ...

$$N_{\mathrm{H}} = \frac{\tau_{\mathrm{grains}}}{\sigma_{\mathrm{grains}}} \tag{9}$$

Where

$$\sigma_{\mathrm{grains}} = \xi_g \int_0^{\infty} \frac{dn_{\mathrm{gr}}(a)}{da} \sigma_{\mathrm{geom}} \times [1 - \exp() - \langle\sigma\rangle\langle N\rangle)]da \tag{10}$$