SAPIENZA
UNIVERSITÀ DI ROMA

Sapienza University of Rome

Physics Department
PhD in Astronomy, Astrophysics and Space Science (AASS)

Case studies of interpretable machine learning in astrophysics

Thesis Advisor
Prof. Giuseppe Bono (Tor Vergata University of Rome)
Dr. Mario Pasquato (Université de Montréal & MILA)

Candidate
Piero Trevisan
19467667

Academic Year MMXXII-MMXXIII (XXXVI cycle)

Where are the turtles?

Abstract

In this work, variable stars (especially RR Lyrae) light curves and the pursuit of Intermediate Mass Black Holes (IMBHs) within globular clusters (GCs), are explored through a data-driven approach using interpretable machine learning (ML) techniques. We initiate the study with the development of an inherently interpretable classifier, utilizing L1 penalization to induce simplicity through sparsity in the model. This approach ensures a straightforward interpretation of astronomical data, providing a transparent model that facilitates the extraction of valuable insights. This penalized classifier, which reaches 90% sparsity in the light-curve features, with a limited trade-off in accuracy performs well both on the Catalina Sky Survey validation set and, remarkably, also on the different ASAS/ASAS-SN light curve test set.

Following this, we apply the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm to analyze light curves of RR Lyrae and δ Scuti stars, uncovering the underlying dynamical systems from observational light curves from the Catalina Sky Survey. This method yields a sparse and interpretable representation. The success rate depends systematically on variable type, with possible implications for variable star classification; however it does not obviously depend on amplitude or period. Successful models can be reduced to the generalized Lienard equation $\ddot{x}+(a+bx+c\dot{x})\dot{x}+x=0$. For $a, b = 0$ the equation can be solved exactly, and it admits both periodic and non-periodic solutions. We find a condition on the coefficients of the general equation for the presence of a limit cycle, which is also observed numerically in several instances. In addition, we employ Dynamic Mode Decomposition (DMD) to investigate the modes of variable stars within the Omega Centauri cluster. This data-driven technique provides insights into the diverse modes of stellar variability, contributing to our understanding of the complex dynamics of these stars.

In the context of IMBH detection in GCs, we apply two ML models: CORELS, an inherently interpretable model, and XGBoost, a black box model elucidated post hoc using local, model-agnostic explanation rules known as anchors. By training these models on simulated GC data and subsequently applying them to actual observational data, we emphasize the importance of interpretability in scientific investigations. Our results demonstrate that simpler, interpretable models can indeed attain accuracy comparable to their more complex counterparts (for the relevant metrics), an approach of significant importance in the field of astronomy where comprehending the model's decision-making process is crucial for establishing trust and facilitating further scientific exploration for domain-field experts.

Our findings challenge the prevalent assumption that complexity is a necessary condition for accuracy, highlighting the existence of interpretable models within the set of accurate predictive models. In the domains of variable stars and identifying intermediate-mass black holes within globular clusters, we demonstrate that the application of machine learning tools can be both reliable and insightful when guided by models that are both interpretable and straightforward. This aligns with the immediate call for transparency and human-understandability in ML applications, extending beyond astronomy and into the broader scientific community.

Contents

List of Figures **v**

List of Tables **xi**

Acronyms **xiv**

1 Introduction **1**
1.1 Interpretability versus explainability . . . 3
1.2 Machine Learning approach . . . 6
1.2.1 The Task . . . 7
1.2.2 The Performance measure . . . 7
1.2.3 The experience . . . 7
1.2.4 Training set . . . 7
1.2.5 Test set . . . 8
1.2.6 Validation set . . . 8
1.2.7 Deployment Dataset . . . 8
1.2.8 Standard metrics . . . 8
1.2.9 k-Fold . . . 9
1.2.10 Stratification . . . 10

2 Variable stars **13**
2.1 The astrophysical context . . . 13
2.1.1 RR Lyrae . . . 14
2.1.2 Eclipsing Binaries . . . 17
2.2 Data . . . 17
2.2.1 Catalina Sky Survey . . . 18
2.2.2 ASAS/ASAS-SN . . . 19
2.2.3 ω Cen data . . . 21
2.3 Methods . . . 22
2.3.1 Gaussian Processes . . . 22
2.3.2 (Regularized) Logistic Regression . . . 25
2.3.3 SINDY . . . 28
2.3.4 Koopman Analysis and DMD . . . 31
2.4 Sparse Logistic Regression for RR Lyrae versus Binaries Classification . . . 34
2.4.1 Introdution . . . 34
2.4.2 Data processing . . . 35
2.4.3 Results . . . 39
2.4.4 Conclusion . . . 46
2.5 Sparse identification of variable-star dynamics . . . 47
2.5.1 Data . . . 48
2.5.2 Analytical solutions for s-0 . . . 51
2.5.3 Limit cycle behavior of the g-0 form . . . 56

2.5.4 Statistics on the full sample . . . 58
2.5.5 Conclusion . . . 61
2.6 Variable Star Light Curves in Koopman space . . . 61
2.6.1 Data processing . . . 62
2.6.2 Results . . . 62
2.6.3 DMD vs Fourier Series: . . . 68
2.6.4 Conclusions . . . 70

3 Intermediate-Mass Black Hole 72
3.1 Introduction . . . 72
3.2 Intermediate Mass Black Holes . . . 73
3.3 Globular clusters . . . 75
3.4 Data . . . 77
3.4.1 Training, validation, and test datasets . . . 77
3.4.2 Deployment Dataset and Data Drift considerations . . . 79
3.5 Methods . . . 80
3.5.1 XGBoost Classifier . . . 80
3.5.2 Anchors . . . 82
3.5.3 CORELS . . . 84
3.5.4 Kernel Density Estimation . . . 86
3.6 Results . . . 87
3.6.1 Kernel Density Estimate and real data . . . 90
3.6.2 Physical interpretation of the CORELS rule list and of the anchors . . . 96
3.7 Discussion and conclusion . . . 96

4 Conclusion and final words 99

Bibliography 102

A Appendix A: Other results from the penalized classifier work 126
A.1 Performance with principal components . . . 126
A.2 Performances with information on period and amplitude . . . 128

B Appendix B: Globular Clusters 130
B.1 Full list of candidates . . . 130

C Appendix C: Explainer Dashboard for XGBoost 133

List of Figures

1.1 Number of refereed publications per month that include the terms “machine learning” or “artificial intelligence” in the title (orange) or abstract (blue) in the top 25 astronomy and astrophysics-related journals. Credits: Huppenkothen et al. (2023) 2

1.2 Growth of papers about explainability and interpretability in AI: number of papers that have abstracts or titles containing one or more of the strings: ‘machine learning’, ‘artificial intelligence’, ‘deep learning’ (in saffron yellow) papers with the same condition AND contains words in the abstract or titles like: "explainability", "interpretability", "XAI", "interpretable machine learning", "interpretable", "explainable" (in blue). Only 2000-2023 period is reported. Only peer-review journals are considered. 3

1.3 Trade-off between Model Flexibility and Interpretability in Astronomy: Highly flexible models can achieve robust fits to complex domain training datasets, though they may be prone to overfitting, require extensive computation time, and exhibit poor generalization, all while lacking interpretability. Conversely, models with lower flexibility may not capture the complexity of certain tasks but are often more suitable for many astronomical applications, particularly when simpler baselines are established. Note that the axes in this figure are qualitative and the optimal balance depends on the specific astronomical problem. The references in order from top left to bottom right are Li et al. (2021), Lanusse et al. (2018), Ishida et al. (2019), Małek et al. (2013), Wadekar et al. (2023), Askar et al. (2019) and Evrard et al. (2014). Credits Huppenkothen et al. (2023). 6

1.4 Simple explanation on the k-fold procedure. *Credits*: `https://docs.cleanlab.ai/v2.0.0/tutorials/pred_probs_cross_val.html` 10

1.5 Example of stratification in our training dataset. In the *left panel* it can be seen that compared to the unstratified split (*right panel*) the ratio between IMBH host and non-host is kept costant in all the folds. 11

2.1 Hertzsprung-Russel diagram (HRD) showing the position of different classes of pulsating stars. The borders of the classical instability strip are plotted with gray lines, while the double line represents the zero-age main sequence (ZAMS). 14

2.2 On the *left panel* an example of bailey diagram (sample of RR Lyrae from Catalina Sky Survey. We can see that RRc are clustered in the region of period 0.2-0.4 days amplitudes of 0.2-0.6 mag, meanwhile the RRab span bigger amplitude and have longer periods. On the *top right panel* we can see clearly the saw-tooth shape of RRab lightcurve compared to the more symmetric sinusoidal behaviour of a RRc lightcurve *bottom right panel*. 16

2.3 Light curves of eclipsing binaries from the Catalina Schmidt Survey (CSS): in the *top panel* a sample lightcurve of a Algol Type binaries. In the *bottom panel* a sample lightcurve of a W UMa type. Note the difference in the depth of the eclipses for the two types. 18

2.4 Positions of RRab/RRc and eclipsing stars in galactic coordinates from the Catalina sample. We can see that this sample does not include regions around the Galactic plane. 19

2.5 Initial adopted sample of CSS. *Top-left panel:* the boxplot of the number of data points. *Top-right panel:* boxplot of the amplitudes. *Bottom-left panel:* boxplot for the period and finally in the *bottom-right panel* the bailey diagram of the initial adopted sample. 20
2.6 Bailey diagram on the left for the ASAS-ASAS-SN sample, some statistics on the period (*center panel*) and amplitude *right panel*). 21
2.7 Left panel: Light curve for star V59, a typical RRab variable from ω Cen sample. Observational data points with error bars are shown in gray, our spline fit is shown as a black solid line. Right: Light curve for star V103, a typical RRc variable. 22
2.8 Adopted sample from the *omega* Cen sample in the period-amplitude plane, RRab and RRc stars are respectively represented by black and red points. 22
2.9 Sigmoid function. 26
2.10 Examples of phase-folded lightcurves from CSS of RRab (*first panel*), RRc stars(*second panel*), EA binaries (*third panel*) and EW binaries (*fourth panel*). 29
2.11 Schematic of the SINDy algorithm, demonstrated on the Lorenz equations. Credits to Brunton et al. (2016) 30
2.12 Schematic of the Dynamic Mode Decomposition on a fluid flow dataset. *Credits:* Kutz et al. (2016) 32
2.13 Period–amplitude plot of our CSS data set. Only the region up to 1.5 mag in amplitude and 1 day in period is shown. The stars labeled as RRc are shown in red; the eclipsing variables are shown in different shades of gray. Substantial overlap between the two groups is readily evident. 35
2.14 *Top-left panel:* Autocorrelation plots for the different kernels. I am plotting the distance vs the covariance for a single row of the covariance matrix. This effectively shows how the covariance between points decreases as the distance between them increases, which is a way to visualize 1D kernel functions. *Top-right panel:* we can see the spatial autocorrelation in two dimensions with the covariance values color coded. *Bottom-right panel:* the covariance matrix is computed by evaluating the kernel function on pairs of points. The range for the colormap is set to be between the 10th percentile and the 90th percentile of the kernel values, to increase the contrast for better visualization. *Bottom-left panel:* 5 samples from the $\mathcal{GP}$ with the combined kernel. Each sample is a possible realization of the $\mathcal{GP}$ over the given grid of x values. 37
2.15 Recap of our data processing with a random EW star. Step 1: folding of the data with the nominal period. Step 2: replicate three times our data to enforce periodicity and remove contour effects. Step 3: gaussian process with the kernel described in the text. Step 4: phase-shift our data in order to have the maximum at the start and selection of only one "phase". Step 5: normalization in magnitude: now we have removed information about the amplitude and period and our classifier will use only the light curve shape. The light blue shaded area is the 95%confidence interval based on the standard deviation. 38
2.16 *Left Panel*: Distribution of our data in principal component space. The first three components are shown. The RRc variables are shown in orange, the eclipsing binaries in light blue. *Right Panel:* Percentage of variance explained by the first 10 principal components. The first four principal components explain 90% of the variance of the dataset. 39
2.17 Confusion matrix for the different models. The results are referring to the validation dataset 41
2.18 Confusion matrix for the penalized logistic regression. The results are referring to the validation dataset 41
2.19 Precision-recall curve colorcoded for the different models for both the RRc and EB classes respectively on the left and right panel. When working with unbalanced dataset a precision-recall curve is preferred to a ROC curve. 42

2.20 Percentage of shared misclassifications between the models. 42
2.21 Example of misclassified light curve by all the models. Everytime, I went back to fix the gaussian processes or the parameters of models in order to address these misclassification, I had obviously a better classifier but at the same time I was slowly overfitting my validation set. 43
2.22 Top: values of the linear coefficients of the logistic regression for the non-penalized model (blue bars). A positive coefficient means that an increased value of the magnitude at the corresponding phase would increase the confidence of classification as RRc. Meanwhile, a negative coefficient increase the confidence of classification as EB. Mean light curve of all RRc in the whole data set (orange), mean light curve of all EW (green) and EA (red) are shown as dots. Bottom: same for the L1 regularized model: the regularization strength was chosen so that 92% of the coefficients became zero. 44
2.23 Performance metrics of our classifier (y axis) as a function of the fraction of coefficients that equal exactly zero (sparsity, x-axis). At over 90% we reach a knee where the recall for the RRc class drops precipitously but for lower sparsity the three metrics are approximately constant. 45
2.24 Feature importance scores for the different features in the decision tree model: feature importance scores are calculated based on the amount that each attribute split point improves the purity of the node. The idea is to measure the effectiveness of each feature in separating the classes. 45
2.25 Left: confusion matrix of our non-penalized classifier on the test dataset; right: confusion matrix of our L1 regularized classifier on the test dataset. 46
2.26 Adopted sample for this study in the period-amplitude plane. 48
2.27 Raw light curves for six stars from our original catalog (gray points with error bars) with our best fit spline superimposed (black solid curve). 49
2.28 Four light curves for stars classified as RRab (thick dotted black lines) and the respective Sparse identification of nonlinear dynamics (SINDy) solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panel show the magnitude derivative versus time and time versus the magnitude, respectively. The name and type of the star are shown in the bottom right panel. 52
2.29 Eight light curves for stars classified as RRc (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panel show the magnitude derivative versus time and time versus the magnitude, respectively. The name and type of the star are shown in the bottom right panel. . 53
2.30 Six light curves for stars classified as δ Scuti (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panel show the magnitude derivative versus time and time versus the magnitude, respectively. The name and type of the star are shown in the bottom right panel. . 54

2.31 Effects of perturbing the initial conditions of the equation learned by SINDy in two different cases. The left panel shows star 3055031035294, which is a δ Scuti variable, whose learned equation is in the simplified $\mathfrak{o}$ form, with $a = 0$ and $b = 0$, the right panel star 3039073051652, which is an RRc variable described by an equation in the full $\mathfrak{o}$ form, with $a \neq 0$ and $b \neq 0$. The dotted gray line is the solution with unperturbed initial conditions, and the solid lines fading from gray to green (blue) correspond respectively to perturbed initial conditions where the magnitude has been increased (decreased). In the left panel these correspond to different, but still closed, phase curves. In the right panel they are not closed, but they evolve slowly towards the limit cycle corresponding to the unperturbed solution. Unlike in the previous figures, here the top right and the bottom left panels show three periods rather than one. 55
2.32 Phase planes for s-$\mathfrak{o}$ with $c = 0.2$ (left) and for g-$\mathfrak{o}$ with $a = -0.01$, $b = c = 0.25$ (right). Color encodes the modulus of the vector tangent to the phase trajectory in each point. 55
2.33 Relation between the a, b, and c coefficients for which $\hat{A} > 0$ (green shaded area), resulting in the existence of a limit cycle for g-$\mathfrak{o}$. 57
2.34 Normalized MSE as a function of the light curve amplitude (panel a) and period (panel b). Big black circles represent RRab variables, small gray circles RRc variables, and blue circles represent δ Scuti variables. Panels c to f show, in order, normalized MSE, a, b, and c color-coded onto the Bailey diagram. Colors scales have been clipped to the first and last decile of each variable. The sizes of circles codify the star type as in the first two panels. 59
2.35 Scatter plots of the a, b, and c coefficients versus each other and versus normalized MSE. Big black symbols represent RRab variables, small gray symbols RRc variables, and blue symbols represent δ Scuti variables. Filled circles correspond to g-$\mathfrak{o}$ stars and crosses to s-$\mathfrak{o}$ stars. 60
2.36 DMD reconstruction (orange dashed line) of the interpolated light curve (blue points) compared with the original (red points) and time shifted (green points) raw data for RRab star V59. 63
2.37 DMD reconstruction (orange dashed line) of the interpolated light curve (blue points) compared with the original (red points) and time shifted (green points) raw data for RRc star V103. 63
2.38 Left panel: nMSE achieved by the DMD reconstruction against the original light curve as a function of the spatial dimension of the HODMD snapshot (d parameter in pyDMD). The black solid line represents the median curve for the RRab stars, the red solid line for the RRc stars. The lower and upper boundaries of the black and red shaded areas represent the first and third quartiles respectively. Right panel: Same as the left panel but nMSE is plotted as a function of the number of eigenvalues. . . 64
2.39 Left panel: histogram of the minimum values of d needed to achieve nMSE = 0.45 (or below) for the RRab (black) and the RRc variables (red). Right panel: same, but for the number of eigenvalues. 64
2.40 Bailey diagram as in the right panel of Fig. 2.7, with the Koopman dimension d needed to attain nMSE = 0.45 color coded. 65
2.41 Left panel: Koopman space dimension d as a function of amplitude for RRab (black points) and RRc (red points). Right panel: Koopman space dimension d as a function of period, same color coding as the left panel. 66
2.42 Left panel: number of eigenvalues as a function of amplitude for RRab (black points) and RRc (red points). Right panel: number of eigenvalues as a function of period, same color coding as the left panel. It is known that short period, large amplitude RRab are harder to fit with fourier and in our case they also need more eigenvalues. 66

2.43 Left panel: DMD eigenvalues (red points) for variable V5 calculated between Julian time 2449858.7698 and 2449869.7601, corresponding to the beginning of the observation period. The green dashed line corresponds to the unit circle in the complex plane. Middle panel: DMD eigenvalues (blue points) for variable V5 calculated between Julian time 2450971.5723 and 2450985.6906, corresponding to the end of the observation period. Right panel: DMD eigenvalues that differ between the two. . . . 67
2.44 Left panel: DMD eigenvalues (red points) for variable V115 calculated between Julian time 2449858.7468 and 2449869.7635, corresponding to the beginning of the observation period. The green dashed line corresponds to the unit circle in the complex plane. Middle panel: DMD eigenvalues (blue points) for variable V115 calculated between Julian time 2450971.5723 and 2450985.6906, corresponding to the end of the observation period. Right panel: DMD eigenvalues that differ between the two. . . . 67
2.45 Left panel: DMD eigenvalues (red points) for variable V120 calculated between Julian time 2449858.7468 and 2449869.7635, corresponding to the beginning of the observation period. The green dashed line corresponds to the unit circle in the complex plane. Middle panel: DMD eigenvalues (blue points) for variable V120 calculated between Julian time 2450970.5019 and 2450985.6906, corresponding to the end of the observation period. Right panel: DMD eigenvalues that differ between the two. . . . 68
2.46 Reconstruction of a light curve with an increasing number d of eigenvectors. Top panel : $d = 10$. Center panel : $d = 20$. Bottom panel : $d = 30$. 69

3.1 Globular cluster, M10. *Credits:* Till Credner and Sven Kohle, Observatorium Hoher List. 75
3.2 Pair plot showing our star cluster simulations in feature space. Intermediate-Mass Black Hole (IMBH) hosts are shown as orange points and non-hosts as blue points. In order, the axis are: Half-mass relaxation time (HRT), total luminosity (TVL), central velocity dispersion (CVD), central surface brightness (CSB), Core radius (OCR) and finally half-light radius (OHLR). All the reported values are at 12 Gyr. 79
3.3 Mass-to-light ratio of the simulated clusters color-coded (IMBH hosts are shown in orange and non-hosts in blue) 80
3.4 LIME vs. Anchors: A Toy Visualization LIME solely learns a linear decision boundary and produces local surrogate models. LIME has a problem of instability of the explanations in addition to the difficult to define the neighbourhood of an explanation. Figure from Ribeiro et al. (2018). 83
3.5 Histograms of features for the training dataset. The acronyms are: half-mass relaxation time (HRT), total luminosity (TVL), central velocity dispersion (CVD), central surface brightness (CSB), Core radius (OCR) and finally half-light radius (OHLR). All the reported values are at 12 Gyr. We used the percentile to create the CORELS features. 88
3.6 Left-hand panel: precision recall curve for our XGBoost classifier on each cross-validation fold (solid lines). The dots represent the recall and precision achieved by a CORELS classifier trained on the same fold. The colours identify the fold for both XGBoost and CORELS. Right-hand panel: Precision recall curve for our XGBoost classifier on the test set. The red dot represents the recall and precision achieved by the CORELS classifier on the same test set. 89
3.7 The rules found by CORELS, displayed on the planes corresponding each to the relevant couple of features. The white area corresponds to the region that respects the rule. The red points correspond to simulated star clusters in the training set that host an IMBH and the blue points correspond to non-hosts. 91
3.8 Pair plot showing test data (a representative sample of the simulations, shown in blue) versus real star cluster data (orange). In order, the axis are: Half relaxation time (HRT), total Luminosity (TVL), central velocity dispersion (CVD), central surface brightness (CSB), Core radius (OCR) and finally half-light radius (OHLR). 92

3.9 *Upper panel*: violin plot for the relevant features. We can see that the central surface brightness span more order of magnitudes.*Lower panel*: Distribution of the KDE score for test data (orange) and actual star cluster data (blue). The vertical line is the cutoff adopted for inclusion in the final sample of candidates, which corresponds to the first decile of the KDE score for test data. 93
3.10 Euler-Venn diagram summarizing the inclusion rules for GCs in the final candidate list. The dark gray set includes all GCs who fall in-distribution with respect to the simulations according to the criterion based on KDE we discussed in the text (112 in total); the light gray set includes the XGBoost candidates (88 in total); and the brown set the CORELS candidates (29 in total). The clusters in the mutual intersection of all three sets are included in the final candidate list (24 in total). 94
3.11 Candidate selection shown in the space of relaxation time, total luminosity, log half-light radius, and log core- over half-light radius. The gray squares represent the training data set, the black dots the data from actual star clusters compiled by Baumgardt and Hilker (2018). Star clusters selected as IMBH host candidates only by XGBoost (CORELS) are shown in blue (red), and candidates selected by both models are shown in purple. Units are shown in square brackets except for dimensionless quantities. 95

A.1 Precision-recall curve colorcoded for the different models with principal components as training data for both the RRc and EB classes respectively on the left and right panel. When working with unbalanced dataset a precision-recall curve is preferred to a ROC curve. 126
A.2 Confusion matrix for the different models trained on the first 4 principal components. The results are referring to the validation dataset. 127
A.3 Confusion matrix for the penalized logistic regression trained on the first 4 principal components. The results are referring to the validation dataset. 127
A.4 Precision-recall curve colorcoded for the different models, denoted in the subplot titles, with principal components as training data for both the RRc and EB classes respectively on the left and right panel. When working with unbalanced dataset a precision-recall curve is preferred to a ROC curve. 128
A.5 Confusion matrix for the different models denoted in the subplot titles. with period and amplitude as features as well. The results are referring to the validation dataset. 129
A.6 Confusion matrix for the penalized logistic regression with period and amplitude as features as well. The results are referring to the validation dataset. 129

C.1 Screenshots from the explainer dashboard applied to XGBoost model to the real clusters: as we can see we can see the dashboard have different sections such as feature importances, classification stats, individual predictions, what if, feature dependence and also decision tree. Shap summary (left): shows summarizes the shap values per feature. Can be either sorted by absolute SHAP value (average absolute impact of the feature on final prediction) or by permutation importance (how much does the model get worse when you shuffle this feature, rendering it useless?). Instead SHAP dependence (right) shows the relation between feature values and SHAP values. This allows to investigate the general relationship between feature value and impact on the prediction. In the dashboard you can check whether the model uses features in line with your intuitions, or use the plots to learn about the relationships that the model has learned between the input features and the predicted outcome. In the example we are looking at feature:HRT and the grey area is the value for index=60 corresponding to M10. 134

List of Tables

2.1 Number of RRc and RRab light curves included in the ω Cen initial and final dataset for each filter. Details concerning the selection are specified below. 21
2.2 Adopted data set summary. The initial number of stars (second column) is shown broken down by variable type (first column). The third column shows the number of stars with a successful Gaussian process fit, and the fourth column the stars with a RUWE<1.4; only these stars were included in the final adopted data set. 37
2.3 Different metrics for different models. The L1 Log Regression is the model that have a 90% of the coefficients equals to zero 40
2.4 Performance metrics broken down by class (first column) of our non-penalized (second column) and L1 regularized model (third column) on the test set. 46
2.5 Relevant parameters for the stars whose light curves are shown in Fig. 2.27. We report the numerical ID (col. 1), right ascension (col. 2) and declination (col. 3), period in days (col. 4), V-band magnitude (col. 5), number of measurements (col. 6), and amplitude (col. 7). 49
2.6 Summary of SINDy results broken down by variable type (columns). The number of governing equation solutions found to match the observed light curve to within a given normalized MSE is shown, with percentage on the sample in parenthesis. . . . 58
2.7 Confusion matrix for a classifier based on the minimum Koopman dimension d needed to achieve nMSE= 0.45. 65
2.8 Spearman correlation coefficients (col. 3, col. 5) between the number of eigenvalues and d with amplitude and period over each class of variable stars and the associated p-values (col. 4, col. 6). 65

3.1 Rules found by CORELS during training in 5-fold cross-validation. The first column reports the number of folds in which a given rule was found. The corresponding rule is reported in the second column. 87
3.2 Sub-rules for $T_r < 2000$ Myr and $R_h < 3.0$ pc and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$ and the corresponding precision-recall 88
3.3 Sub-rules for$T_r < 5000$ and $\mathrm{I}_0 > 1000$ $\mathrm{L}_\odot/pc^2$ and $L > 10000$ and $R_c/R_h > 0.2$ and the corresponding precision-recall 89
3.4 Final table of IMBH candidate hosts. The first column reports the cluster name, columns 3-8 report the (rounded, log) values of the relevant features. Columns 9-11 show the classification prediction by XGBoost, CORELS and the in-distribution criterion based on KDE and column 12 shows the confidence of the classification by XGBoost, based on which the rows are ordered in decreasing order. Column 13 reports the IMBH upper mass limit based on radio observations by T18. The last column lists any previous claims or other relevant literature discussing IMBHs in the cluster. The full table, including non-hosts, is presented in Appendix B.1. 94
3.5 Anchors explaining XGBoost decision for real GCs. The first column reports the GC name, the second column the anchor as a list of conditions (each feature has been given a different colour for ease of comparison), the third column reports the anchor's precision (different from the model precision) and the fourth column its coverage. . . 97

A.1 Different metrics for different models trained on the first 4 principal components. The L1 Log Regression is the model that have C=0.056. 126
A.2 Different metrics for different models with the additional information about amplitude and period. The L1 Log Regression model have the same light-curve features as the ones described in the text with the addition of period and amplitude features. We can see that we have a small increment in performance likely due to the stars not in the critical region shown in Fig.2.13. 128

Acronyms

CORELS Certifiably Optimal RulE ListS

CRTS Catalina Sky Survey

CSS Catalina Schmidt Survey

DL Deep Learning

DMD Dynamic Mode Decomposition

GC Globular Cluster

GPs Gaussian Processes

IMBH Intermediate-Mass Black Hole

ML Machine Learning

RRL RR Lyrae

SINDy Sparse identification of nonlinear dynamics

SVD Singular Value Decomposition

Chapter 1

Introduction

In the contemporary landscape of astronomical research, Machine Learning (ML) and Deep Learning (DL) has emerged as an indispensable tool, demonstrating its versatility across a myriad of applications, ranging from the classification of transient astronomical events to the development of neural network emulators for intricate cosmological simulations. The astronomical community is currently witnessing a paradigm shift towards data-driven methodologies, propelled by the influx of extensive observational datasets (Zwicky Transient Facility Masci et al., 2018, near future Vera C. Rubin Observatory Ivezić et al., 2019 among others), the complexity of observational biases (Razim et al., 2021) and the advent of high-resolution simulations. This resulted in a remarkable growth of peer-reviewed ML research in astronomy (see Fig. 1.1).

A really recent paper (Huppenkothen & Ntampaka et al., 2023) aimed at the astronomical community, emphasizes the need for best practices in building, implementing, and reporting ML methods in astronomy, considering the unique challenges of astronomical data. One of the key points of the article is also the importance of interpretability, especially when ML models are less interpretable than existing methods motivated by underlying physics.

Acknowledging the potential risks associated with uninterpretable models, this thesis focuses on the necessity of incorporating interpretable tools within astronomical ML frameworks. Let ask ourselves:

What is interpretability in astronomical ML problems?

In the proceedings of the IAU Symposium titled "Machine Learning in Astronomy: Possibilities and Pitfalls," (Cabrera et al., 2022) Sungwook E. Hong articulated a perspective in response to this question, stating:

> We astronomers are automatically astrophysicists. We want to *understand* the nature behind astronomical data, not just handle the data. But most ML techniques that guarantee high predictability are very sophisticated, and so become black boxes for most users. This means that, ironically, we now need to put additional efforts into *understanding* the very tool that we introduced to *understand* the nature. That is probably one of the key factors why some notable astronomers are disinclined towards applying MLs to astronomy. While many so-called i*nterpretable* MLs have been suggested in the ML community, I believe that we need to discuss with classical, old-fashioned astronomers to define the *interpretability* in astronomical ML. Their expertise is *interpreting* the astronomical data and theories — their interpretations may not always be valid, but they

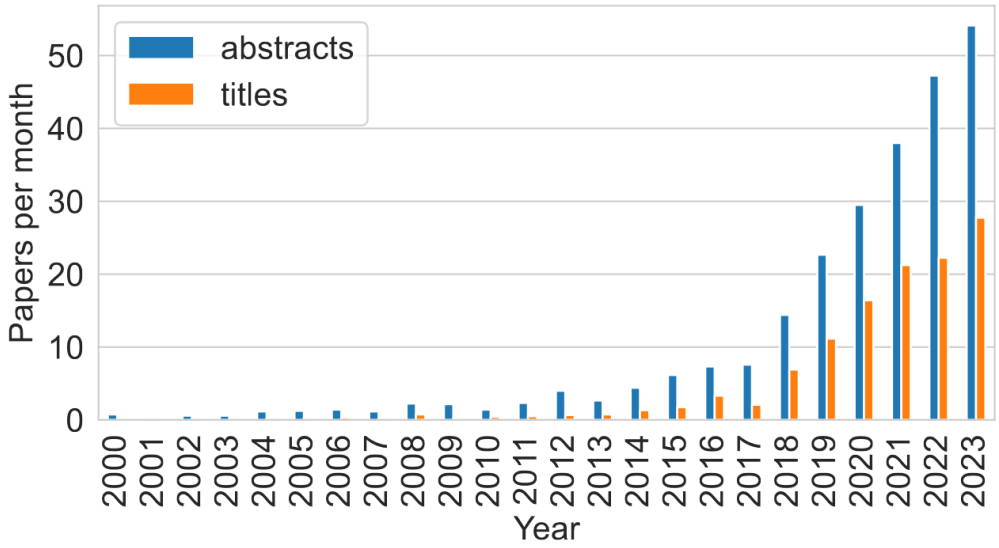


Figure 1.1: Number of refereed publications per month that include the terms "machine learning" or "artificial intelligence" in the title (orange) or abstract (blue) in the top 25 astronomy and astrophysics-related journals. Credits: Huppenkothen et al. (2023)

> have a very high probability of being valid. Therefore, if our definition and application of interpretability do not match with their intuition at all, it will not be accepted in the general astronomical society, no matter how fancy it seems to be.

Echoing his sentiments, I would emphasize that our application of ML also ought to be human-understandable. It is imperative that we prioritize interpretability over sheer predictive power. If a model's outcomes do not align with our intuitive understanding of the data, its predictions inevitably become questionable. Consequently, this research aligns with an emerging trend, that advocates the use of fully data-driven approaches, emphasizing human understandability. For me, a key part in the response of Sungwook was the focus to work with field-domain expert, especially that our definition of interpretability has to match with their intuition. This part was not crystal clear for me at least at the start of the PhD, but as we can see in the second part of this work (see chapter 3) this paradigm shift became prevalent.

As illustrated in Fig. 1.2, a preliminary scan of the ADS (Astrophysics Data System[1]), without the intention of being systematic or comprehensive , reveals an exponential growth trend in "Explainable/Interpretable" topics within the "AI/ML/DL" field in the last years. It is worth noting that some papers, (including our own Pasquato et al., 2022), may utilize interpretable models or explanations without explicitly mentioning this in the abstract. Despite this, the optimistic exponential trend in the literature is evident, although the absolute numbers remain relatively small compared to the vast number of papers published annually in the recent boom of AI in astronomy.

Probably by now the reader will be already confused about the terms interpreability/explainability/Explainable AI/interpretable ML and so on. Do not worry, you are not alone. A lot of papers, not only in astronomy use these words interchangeable. I do not agree: they have nuanced differences. Since, there is not a globally agreed definition following the crystal-clear work by Molnar (2022), I will expand on this subject and at least obtain operational definition of interpretability and explanations.

[1] `https://ui.adsabs.harvard.edu/`

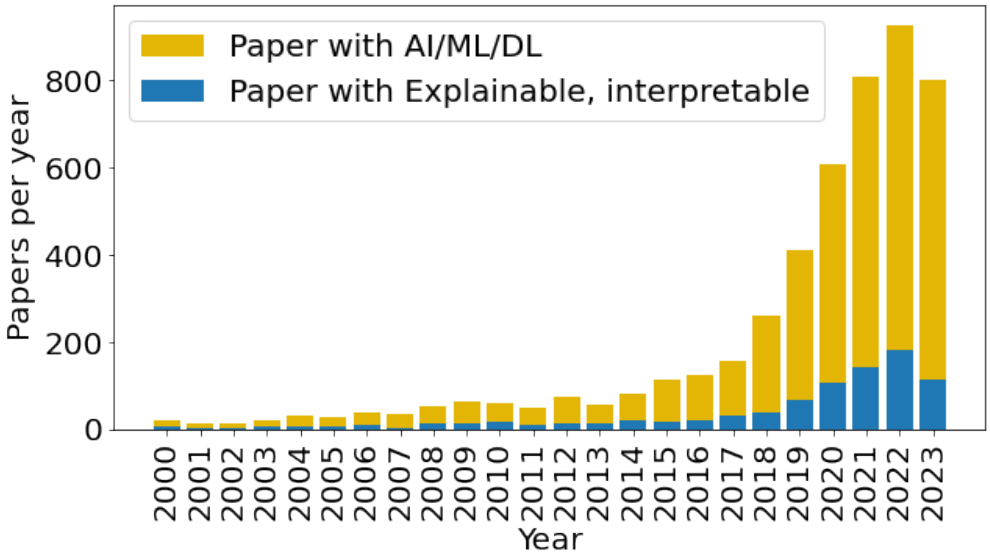


Figure 1.2: Growth of papers about explainability and interpretability in AI: number of papers that have abstracts or titles containing one or more of the strings: 'machine learning', 'artificial intelligence', 'deep learning' (in saffron yellow) papers with the same condition AND contains words in the abstract or titles like: "explainability", "interpretability", "XAI", "interpretable machine learning", "interpretable", "explainable" (in blue). Only 2000-2023 period is reported. Only peer-review journals are considered.

1.1 Interpretability versus explainability

Before going on the distinction of interpretability (Interpretable AI) and explainability (XAI), I will clarify the word black box. A *black box* is a type of system that keeps its internal processes hidden. In the realm of ML, the term black box model is used to characterize models whose workings are not transparent or interpretable through an examination of their parameters, such as a neural network. On the other hand, the counterpart to a black box is often termed a *white box*, and in this work, it is synonymous with an interpretable model.

Interpretable AI focuses on using models that are inherently transparent and easy to understand. The interpretability is built into the model itself. Example of white boxes are Linear Regression, Logistic Regression, Decision Trees and Rule-Based Systems (if-then rules).

XAI, instead, employs a variety of post hoc interpretation methods to explain complex models. These post hoc interpretations are called explanations. Interesting enough, some of the possible explanations of black box are so complex that are not "interpretable". At the same time, the use of a white box does not automatically gives human understandability.

White boxes and good explanations should possess the following characteristics:

- Simplicity: the explanation should be simple enough for the target audience to understand without requiring extensive background knowledge. E.g. the use of a small number of features to explain a prediction, rather than a long list of contributing factors. Simplicity can come from sparsity as shown in this work. When the performance is not sacrified excessively, the sparsified model should be preferred. The classifier explained in Sec. 2.4, the model itself presented in Sec. 2.5 and the penalization of the white box (Sec. 3.5.3) (work presented in chapter 3) have a focus on sparsity. For a review on sparsity see Hoefler et al. (2021).

- Contrastive: humans prefer explanations that highlight the contrast between the actual outcome

and the expected or alternative outcome (Lipton, 1990). In other words, why a precision is made instead of another one.

- Selective: good explanations do not need to cover every detail but should focus on the most important factors that contributed to the outcome. E.g. highlighting the key features that played a significant role in a model's prediction, rather than listing all the features used by the model.

- Feature Importance: the explanation should provide information on the importance or contribution of each feature to the model's prediction. E.g. using techniques like LIME by Ribeiro et al. (2016) (or even better anchors by Ribeiro et al., 2018 as explained in 3.5.2) or SHAP (Lundberg and Lee, 2017 based on the idea in Shapley, 1951) to show how much each feature contributed to a specific prediction.

- Consistent: the explanations should be consistent across different instances, providing stability and reliability in the interpretations. If two similar data points are fed into the model, the explanations for their predictions should also be similar: e.g. Su et al. (2019) demonstrates that deep neural networks can be fooled by changing just one pixel in an image, highlighting the vulnerability of these models to adversarial attacks.

- Causal: whenever possible, explanations should aim to capture causal relationships rather than just correlational ones. Identifying and explaining the direct causes of a prediction, rather than just features that are associated with the outcome. Here, our physical knowledge should come in hand.

- Trustworthy: the explanations should accurately reflect the model's decision-making process, building trust in the model's predictions. Ensuring that the explanations are derived in a way that truly represents how the model operates, even if the model is a complex, black box model.

A key aspect: if we decide to use black boxes, we should at least explain these models. Ideally, the next step is to shift the focus from trying to explain black box models at post hoc to developing models that are inherently interpretable. Post hoc explanations are approximations and may not accurately reflect the model's decision-making process (Rudin and Ertekin, 2018). If a model is a black box, even if we generate post hoc explanations for its predictions, there is a fundamental lack of trust because we do not understand *fully* how the model is making its decisions internally. Post hoc explanations can sometimes be misleading or incorrect, especially if the black box model is complex. If the explanations themselves are not trustworthy, it becomes challenging to justify the use of the black box model.

Let's imagine that our post hoc model can explain 90% of the predictions of the black box, can we trust the explanations? Ideally, we want explanations for 100% of the predictions. But, if we are able to generate explanations that capture all the behavior of a black box model, it raises the question of why we need the black box model in the first place. In such cases, it might be more beneficial to use an interpretable model from the start, as it would provide transparency, trust, and ease of validation.

At the same time context is fundamental: when working with complex domain data (high dimensionality, non-linearity, noisy) like images and videos, the state of the art algorithm are always

black boxes: in the images domain we have for classification problems (He et al., 2016, Tan and Le, 2019, Dosovitskiy et al., 2020 among others), for object detection (Ren et al., 2015, Bochkovskiy et al., 2020, Tan et al., 2020 among others) and for image segmentation (Ronneberger et al., 2015, He et al., 2017 and Chen et al., 2018c among others). In Fig. 1.3 we show the flexibility/interpretablity trade-offs for different kind of models applied in astronomy. The flexibility of these powerful DL models can not easily replaced with natively interpretable models. Although, there are some approaches that aim to address this challenge: for example prototype-based models (Chen et al., 2018b, Nauta et al., 2021) learn a set of prototype images or image patches that are representative of different classes. The classification decision is based on the similarity between the input image and the learned prototypes, providing a more interpretable decision-making process. There have been efforts to apply White box models like decision trees and rule-based models to complex domain like image data. For example, the Oblivious Decision Trees (Kontschieder et al., 2015) and prototype system rule-based (if-then) models (Angelov and Gu, 2018).

I believe that even if interpretable models might not perform as well as black boxes, their transparency and ease of understanding are crucial. It's worth pointing out that with a good amount of effort put into creating the right features, we can actually build interpretable models that can hold their own and even win competitions against black boxes (Chen et al., 2018a). Even more, we can ask ourselves a philosophical question: there is always an accurate and interpretable models across various domains? Is it plausible to believe that nature conceals simple truths, ready to be unveiled through machine learning? While this notion, reminiscent of Occam's Razor, is fascinating, we can discuss it more technically using the notion of function classes and Rashomon Sets, elaborated upon in Semenova et al. (2022). This concept is somewhat related to flat minima, exemplified by Hand (2006).

Let's expand on this Rashomon Set concept: imagine a scenario where the data allows for a plethora of accurate predictive models. Given the abundance of these models, it is likely that at least one of them is interpretable. Hence, we have a model that is both accurate and interpretable.

To elaborate, for a specific dataset, we define the Rashomon set as the collection of reasonably accurate predictive models, maintaining a certain level of accuracy comparable to the best model accuracy. Given the finite nature of data, it is possible to have numerous near-optimal models that predict differently, resulting in a substantial Rashomon set. This is a common occurrence in practice, as various ML algorithms, despite their distinct functional forms (e.g., random forests, neural networks, support vector machines), often exhibit similar performance on the same dataset. If the Rashomon set is sufficiently large and diverse in its predictions, it is likely to include functions that can be well-approximated by simpler functions, thereby also encompassing these simpler models.

Practically, data uncertainty leads to a Rashomon set, and a larger Rashomon set is likely to contain interpretable models, suggesting that interpretable and accurate models are frequently attainable. If this theory is valid, we should anticipate the existence of interpretable models across different domains. Although finding these models through optimization might be challenging, there is a theoretical basis to believe in their existence.

The presence of numerous diverse yet effective models implies that algorithms may lack stability; a minor alteration to the algorithm or dataset could result in a completely different, yet still accurate, model. This variability is not necessarily detrimental; in fact, the availability of diverse high quality models provides domain experts with greater flexibility in selecting a model that they consider

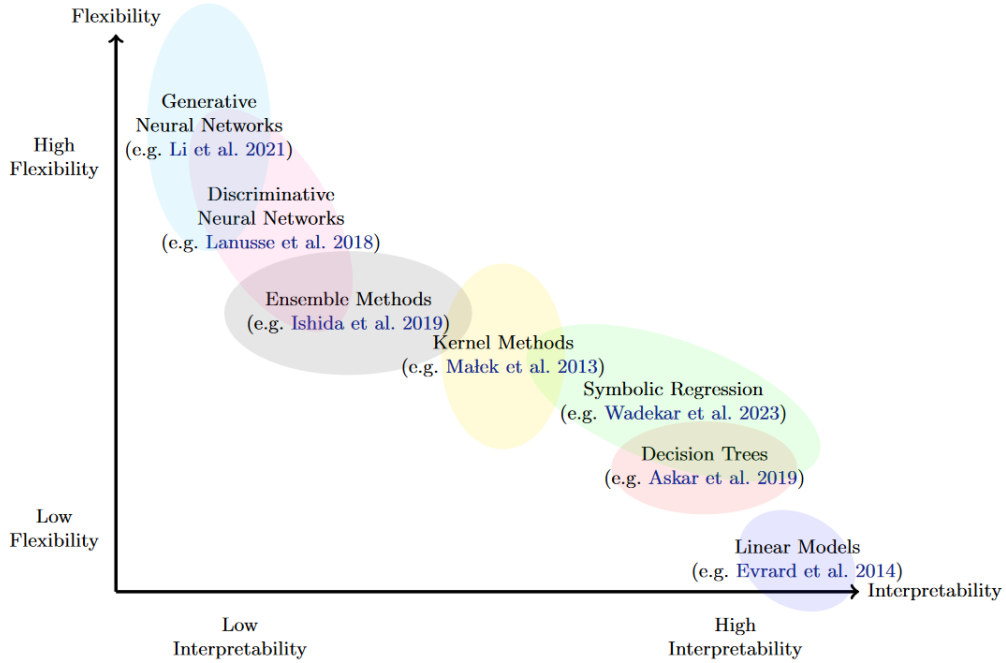


Figure 1.3: Trade-off between Model Flexibility and Interpretability in Astronomy: Highly flexible models can achieve robust fits to complex domain training datasets, though they may be prone to overfitting, require extensive computation time, and exhibit poor generalization, all while lacking interpretability. Conversely, models with lower flexibility may not capture the complexity of certain tasks but are often more suitable for many astronomical applications, particularly when simpler baselines are established. Note that the axes in this figure are qualitative and the optimal balance depends on the specific astronomical problem. The references in order from top left to bottom right are Li et al. (2021), Lanusse et al. (2018), Ishida et al. (2019), Małek et al. (2013), Wadekar et al. (2023), Askar et al. (2019) and Evrard et al. (2014). Credits Huppenkothen et al. (2023).

interpretable. If the reader is interested in the theory of Rashomon sets we refer to Semenova et al. (2022). I discover this fascinating concept really lately in this PhD work but at least it was gratifying to see that intuition on certain matters was underpinned by a much more solid mathematical foundation.

1.2 Machine Learning approach

The fundamental idea behind machine learning is the building of algorithms that "learn" a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task. In order to do so, we have a training set of data, in which we know the relationship between input and output. Using this data we build a prediction model (such as a neural network), which will enable us to predict the outcome for new unseen objects (test set). This approach is the supervised learning one: we have the "labels" in the training set to drive the learning process. More formally, from Mitchell (1997): "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

1.2.1 The Task

ML tasks are usually described in terms of how the algorithm should process an example. The latter, in ML context, is a collection of features that have been measured or are known (pixels of an image, properties of a Globular Cluster). Since ML improved significantly with the number of examples, we want a lot of them. We represent the i-th example as a vector $x^{(i)}$ where each entry x_j is the $j-th$ feature. The most common tasks that can be solved with supervised machine learning are regression and classification (binary or multiclass).

Regression: The task T in a regression problem is the prediction of a numerical value given some input. Therefore, the ML algorithm is asked to output a function f from $\mathbb{R}^n$ to some real value.

Classification: In this type of task, the algorithm has to find which of k categories some input belongs to. In order to do so, the ML algorithm is asked to produce a function f from $\mathbb{R}^n$ to some set $1, \ldots, k$. The input defined by vector x is mapped to a category defined by some value y.

1.2.2 The Performance measure

To measure the ability of our algorithm to perform said task T, we must use some quantitative measure of its performance. This performance measure P will be specific to the task T. For classification tasks, usually we measure the accuracy of the model simply calculating the proportion of examples for which the model produces the correct output. Vice versa, we can obtain the proportion of examples for which the model produces an incorrect output. In general, it is not straightforward to choose a performance measure that corresponds well to the desired behaviour of the system. When performing a classification task, in a dataset where the classes are not equally represented (unbalanced), a model could trivially predict the majority class for all inputs and still achieve high accuracy, despite not having learned to correctly identify instances of the minority class. In this case, we should focus more on the precision rather than the accuracy. Another example is when the cost of false negatives is high we should prioritize recall rather than the accuracy (the proportion of actual positives that were identified correctly). For the proper definition of these terms see Sec. 1.2.8.

1.2.3 The experience

In the formal definition by Mitchell the idea is that the machine learns from some experience E. Based on what kind of experience the machine is allowed to have during the learning process, we can categorize learning as unsupervised or supervised. In almost all cases the learning algorithms are allowed to experience an entire dataset containing examples with many features. The difference is that in supervised learning the machine experiences a dataset containing features, but for each example there is an associated label. For example, the variable stars in the first work have a label corresponding to the class type (see Sec. 2.4).

1.2.4 Training set

The simplest way to describe the dataset $\mathcal{D}$ is with a matrix $X \in \mathbb{R}^{n \times m}$ where m is the number of examples and n the number of features. Thus, X^i_j is the j-th feature of the i-th example. Since

not all examples have the same dimensions (e.g., images with different widths and heights), we describe the dataset as a set with m examples: $x^{(1)}, \ldots, x^{(m)}$. In supervised learning, a dataset is accompanied by its vector of labels y, with $y^{(i)}$ providing the label for example i.

1.2.5 Test set

Usually, we are interested in how well the machine learning algorithm performs on data that it has not seen before, since this determines how well it will work when applied to the "real world". We therefore evaluate these performance measures using a test set of data that is separate from the data used for training the ML system. It is crucial that the ML algorithm *never* sees this dataset during training.

1.2.6 Validation set

A fraction of the dataset composed of examples coming from the same distribution as the training set can be used to estimate the generalization error of the ML algorithm after the learning process has completed. It is vital that the test examples are not used in any way to make choices about the model, including its hyperparameters. For this reason, no example from the test set can be used in the validation set. Therefore, we construct the validation set from the training set: we split the training set into two sets and hold out one for testing the performance of our algorithm with a specific choice of hyperparameters during the training itself. Commonly, one uses about 80% of the training data for training and 20% for validation. One can also divide the dataset into k distinct subsets, training with $k-1$ subsets and leaving one as the validation set. This process is then repeated $k-1$ times, rotating the validation set (see Fig. 1.4).

1.2.7 Deployment Dataset

After a model has been trained, validated, and tested, it is deployed to handle real-world data in a production environment, constituting the deployment dataset. This dataset comprises actual, practical examples the model encounters post-deployment. Contrary to the static nature of training, validation, and test sets, the deployment dataset is dynamic, continually evolving as more real-world data is processed. While this dataset is not used for model evaluation or training, it plays a crucial role in monitoring the model's performance in real-life scenarios. Insights from this dataset can be crucial for iterative model improvements, potential re-training, or fine-tuning to ensure the model's robustness and adaptability over time.

1.2.8 Standard metrics

Important findings from the field of statistics confirm the existence of a statistical cost of interpretability (Dziugaite et al., 2020). This "tradeoff" between performances and interpretability of a model it is well studied (Sarkar et al., 2016, van der Veer et al., 2021, Bell et al., 2022 among others). At the same time, as there is no unambiguous definition for interpretability, there is no canonical and agreed quantification for this trade-off (or even the existence of this tradeoff, see also the Rashomon Set argument in the previous section). In order to understand what is the performance of a model, I present here some of the standard metric used in ML for classification tasks:

Precision Precision, also known as the positive predictive value, is the ratio of correctly predicted positive observations to the total predicted positives.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

It essentially asks: "Of all the samples we labeled as positive, how many were actually positive?"

Recall (or Sensitivity or True Positive Rate) Recall is the ratio of correctly predicted positive observations to all the observations in the actual class.

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

It essentially asks: "Of all the actual positive samples, how many did we correctly label?"

F1 Score The F1 Score is the harmonic mean of precision and recall. It provides a single score that balances the trade-off between precision and recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Specificity (or True Negative Rate) It is the ratio of correctly predicted negative observations to all the observations in the actual negative class.

$$\text{Specificity} = \frac{\text{True Negatives (TN)}}{\text{True Negatives (TN)} + \text{False Positives (FP)}}$$

False Positive Rate It is the ratio of incorrectly predicted positive observations to all the observations in the actual negative class.

$$\text{False Positive Rate (FPR)} = 1 - \text{Specificity}$$

1.2.9 k-Fold

In order to ensure robustness and generalizability in ML rigorous validation schemes are necessary. In k-fold, the dataset is partitioned into k equally-sized folds. Training occurs on k-1 of these, while the remaining fold serves as a validation set. This process is iterated k times, with each fold playing the validation role once. The final model performance is the averaged result across all iterations. Mathematically, for a dataset $\mathcal{D}$ divided into k folds $\mathcal{D}_1$, $\mathcal{D}_2$,..., $\mathcal{D}_k$:

$$\text{Performance} = \frac{1}{k} \sum_{i=1}^{k} \text{Eval}(\mathcal{D} - \mathcal{D}_i, \mathcal{D}_i) \tag{1.1}$$

where Eval is the evaluation function, gauging model performance on each validation fold.

In fig. 1.4, we report a schematics of the k-fold validation.

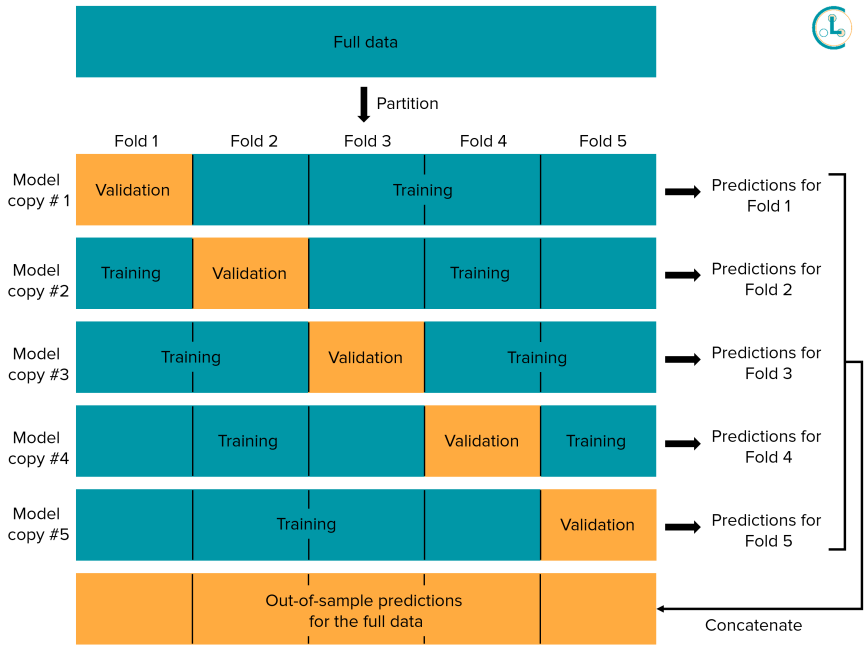


Figure 1.4: Simple explanation on the k-fold procedure. *Credits*: `https://docs.cleanlab.ai/v2.0.0/tutorials/pred_probs_cross_val.html`

1.2.10 Stratification

In the work presented in Chapter 3, the population of Intermediate-Mass Black Hole (IMBH) host and non host are not in the same proportion. We need to address this during the training. Stratified sampling is a way to ensure that samples or subsets, mirror the characteristics of the larger population or dataset from which they are drawn. In essence, stratification divides a population into non-overlapping, homogeneous subgroups (or strata) based on a key attribute, ensuring that each subgroup is adequately represented in the final sample. The overarching goal is to enhance the precision of estimates while minimizing potential biases. When we cross-validate implementing the use of stratified sampling, we are sure that the proportion of the feature of interest is the same across the original data, training set and the test set. This ensures that no value is over/under-represented in the training and test sets.

For a population P, n strata S_1,S_2,...,S_n, the representative samples from each stratum, are denoted s_1,s_2,...,s_n. The goal is to ensure:

$$\frac{|s_i|}{|S_i|} = \frac{|s|}{|P|} \tag{1.2}$$

for all i=1,...,n, where $|s|$ represents the size of the combined stratified sample and $|P|$ the size of the entire population .

In the context of ML, stratified sampling, especially in the case of stratified k-fold cross-validation, ensures that the train-test splits retain the same distribution of classes as the original dataset. In Fig. 1.5, we compare 2 strata in a k-fold ($k = 5$) cross validation. We can see that in the unstratified scenario the proportion of IMBH host per fold vary from 0.23 to 0.33.

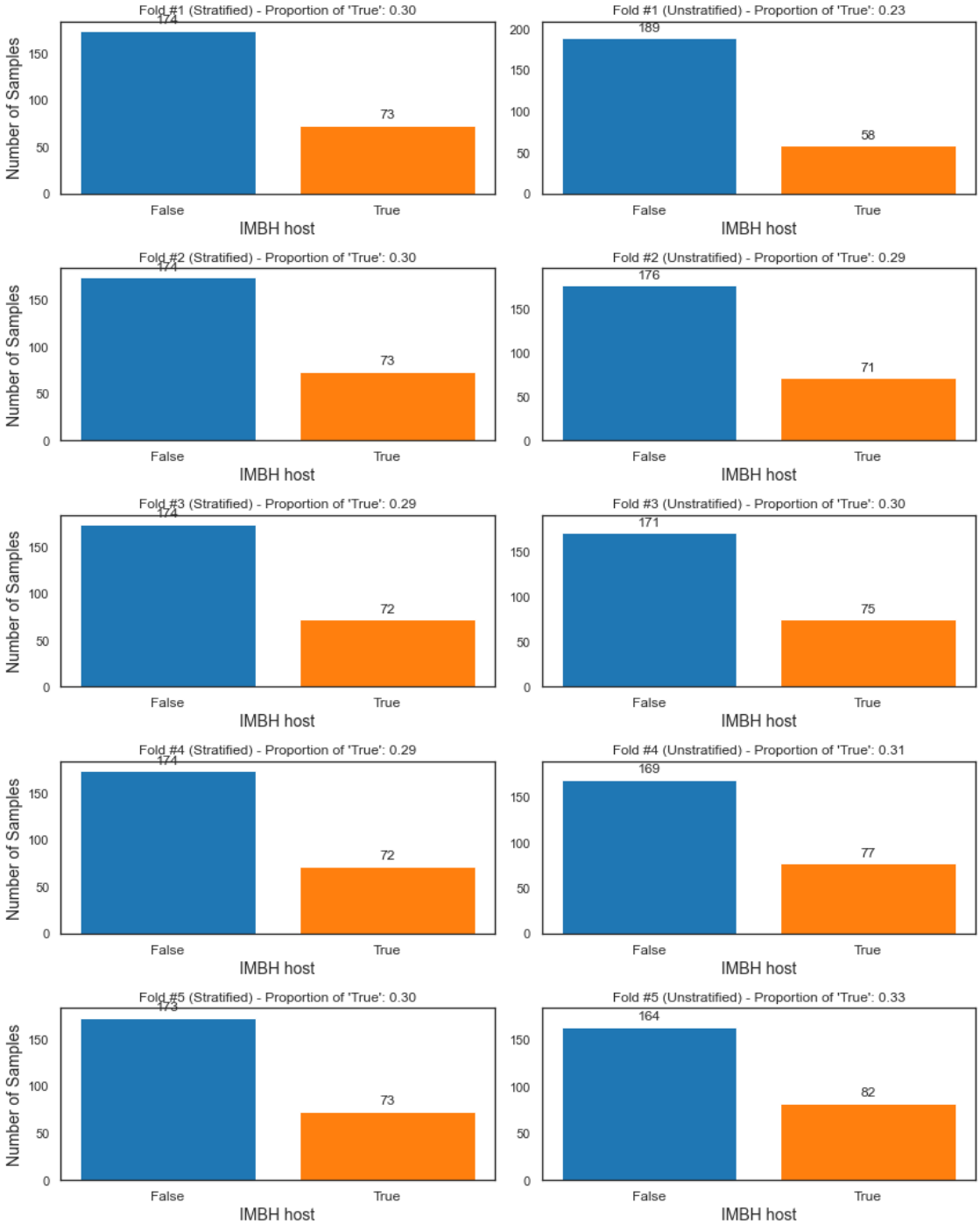


Figure 1.5: Example of stratification in our training dataset. In the *left panel* it can be seen that compared to the unstratified split (*right panel*) the ratio between IMBH host and non-host is kept costant in all the folds.

Thesis outline

The thesis is divided in two main parts: the first part on variable stars and second part on intermediate black holes detection. The first part, after an astrophysical introduction to variable stars, a description of the data and the methods used, is divided in three main sections:

- Sparse logistic regression for RR Lyrae vs binaries classification: Application of gaussian processes to variable stars lightcurve and the implementation of a native intepretable classifier, where interpretability comes from the use of a transparent model with the help of sparsity in order to make it more human-understandable. The model is compared then to black box baseline predictions.
- Sparse Identification of Variable Star Dynamics: we use Sparse identification of nonlinear dynamics (SINDy) algorithm that can find the governing equations of a non-linear system with a focus on sparsity.
- Variable Star Light Curves in Koopman space: Finding the physical meaningful dynamic mode of RR Lyrae stars in a data-driven approach.

The second part of this work instead is about the search for Intermediate Black Holes in Globular Clusters. We trained both an interpretable model, namely rule-list applied via Certifiably Optimal RulE ListS (CORELS) algorithm and a black box model providing local interpretable model-agnostic explanations (Anchors). We compared them and applied to real Globular Cluster (GC).

Chapter 2

Variable stars

I will now present the first case study that resulted in 3 papers (Trevisan et al., 2023, Pasquato et al., 2022 and a paper under second review from ApJ). Firstly, I will present the astrophysical context of variable stars, with a particular focus on RR Lyrae and what are the biggest open problems for these stars. Secondly, I will present the different data used in the three works. Afterward, I will present all the methods applied in the three works. Next, the main course: I present the three papers resulted from the tackling of these problems from a data-driven interpretable perspective. Every section here will retrace the work of the paper (introduction, data cleaning, results and discussion and conclusion).

2.1 The astrophysical context

Astronomical objects exhibit variability across a wide range of timescales and wavelengths, driven by various physical processes. This variability can manifest as changes in brightness, spectrum, or position, providing crucial insights into the underlying astrophysical mechanisms. Studying these variations allows us to probe the internal structures, evolutionary states, and immediate environments of astronomical objects. Variable stars are a subset of astronomical objects whose brightness fluctuates over time due to internal or external factors. Intrinsic variables undergo changes in brightness due to physical processes occurring within the star itself, such as pulsations, eruptions, or evolutionary changes. Examples of pulsating stars include δ Scuti variables and RR Lyrae stars. Extrinsic variables, on the other hand, exhibit variability due to external factors, such as eclipses in binary systems or the presence of dark spots on the stellar surface. Examples of eclipsing systems are W UMa and Algol binaries. This work will focus only intrinsic stars undergoing pulsations (RRL and δ Scuti stars) and extrinsic eclipsing systems. If the reason behind the magnitude variation due to an eclipse is self-evident for pulsating stars the theory is more complex.

In the simplest terms, the foundational framework of the theory is established through the linear perturbation of the hydrodynamic equations of stellar structure, encompassing the conservation of mass, momentum, and energy, coupled with the equation of state and opacity laws. From the perturbation we can derive the stellar pulsation equations. The solutions to these equations, characterized by discrete eigenmodes, reveal the nature and properties of the star's oscillatory behavior. The adiabatic approximation, useful only to get an idea of the physics in a pulsating star, assumes that the timescale of the pulsations is much shorter than the thermal timescale, allowing

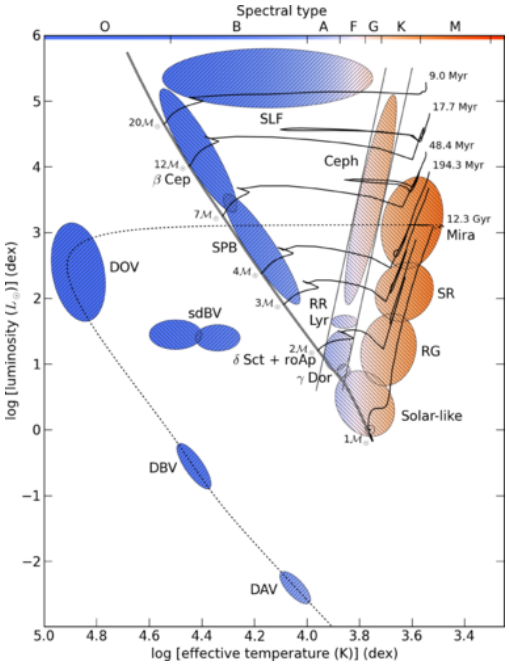


Figure 2.1: Hertzsprung-Russel diagram (HRD) showing the position of different classes of pulsating stars. The borders of the classical instability strip are plotted with gray lines, while the double line represents the zero-age main sequence (ZAMS).

for the decoupling of thermal and mechanical processes. However, to fully understand the driving and damping mechanisms of pulsations, non-adiabatic analyses are necessary, incorporating the effects of radiative transfer and convective motions. The κ mechanism, one of the primary driving forces behind pulsations in the instability strip (see Fig. 2.1), arises due to the dependence of opacity on temperature and density, leading to a phase lag between the luminosity and radius variations and resulting in a net driving of the oscillations. We can extend the theory also to non-radial pulsations, where spherical harmonics are used to describe the angular dependence of the oscillations, providing a richer spectrum of modes and enabling the probing of different stellar layers through asteroseismology. Stellar pulsation theory allows us to decipher the properties of stars, such as their mass, radius, age, and composition. For a good review on the general theory of pulsating stars, the reader is referred to Catelan and Smith (2015).

2.1.1 RR Lyrae

RR Lyrae (RRL) are old ($t > 10$ Gyr), low-mass ($M = 0.6 - 0.8 M_\odot$) variable stars. They typically pulsate in two main radial modes: fundamental and first overtone. A small fraction (5-10%) pulsates simultaneously in both modes.

RRL obey to tight Period-Luminosity (PL) relations for wavelengths longer than the end of the visible spectrum (Catelan et al., 2004). Moreover, the slope of the PL relation becomes steadily steeper and the standard dispersion steadily decreases when moving from shorter to longer wavelenghts. Near infrared PL relations provide individual distances with an outstanding accuracy of $\approx 3\%$ (Beaton et al., 2018, Bono et al., 2003, Braga et al., 2015). Hence, they are a fundamental stepping stone in the calibration of secondary distance indicators, in turn constraining possible systematics in the cosmic distance scale (Beaton et al., 2016). Compared to other distance indicators and stellar tracers, RRL have several advantages:

- RRL are ubiquitous, having been identified both in gas-poor and in gas-rich stellar systems; their lifetimes of the order of 10^8 yrs ensure that they outnumber other evolved radial variables such as type II cepheids, anomalous cepheids, and classical cepheids (Bono, Braga, Pietrinferni, ARA&A submitted).

- RRL cover a very broad metallicity distribution ranging from [Fe/H]$\sim$-3.5 to super-solar iron abundance (Crestani et al., 2021a,b, Fabrizio et al., 2021), so they can be easily identified not only in the MW halo, but also in metal-rich stellar systems such as the MW Disk.

- Relationship with metallicity: The relation between RRL properties and metal content has long been known (Arp, 1955, Baade, 1956, Oosterhoff, 1939), with both period and amplitude decreasing linearly with metallicity, and the shape of the light curve (parametrized by using Fourier coefficients) also correlating with metal content (Jurcsik and Kovacs, 1996b). Recently, (Mullen et al., 2021, Mullen et al., 2022) provided new empirical and homogeneous calibrations for fundamental and first overtone RRL by using spectroscopic calibrators covering a broad range of metal abundances. RRL's color across the phases of the near-minimun light was found to be relatively constant, allowing us to estimate individual reddenings (Blanco, 1992, Sturch, 1966); the distance, the reddening and the metallicity of individual RRL can be estimated with the simultaneous use of both optical and near infrared mean magnitudes (Bono et al., 2019).

- RRL as physics laboratories: Along the pulsation cycle RRL experience several nonlinear phenomena associated with the formation and propagation of several non-linear phenomena (shocks). Accurate and homogeneous spectroscopic investigations covering the rising branch of the light curve found that these phenomena cause the appearance of hydrogen and helium emission lines (Gillet and Fokin, 2014, Preston et al., 2022). These diagnostics are very promising for quantifying the radial displacements of the outermost regions moving in opposite directions (Preston et al., 2022).

RRL are objects fundamental for our understanding of the universe, but same of the open problems in the field are:

- Overlap with eclipsing binaries: The main issue with relying on RRL to accurately measure distances is due to their periods and amplitudes partially overlapping with eclipsing binaries, leading to spurious contamination in RRL samples. In the period-amplitude plane, also called Bailey diagram , the location of first overtone RRL overlaps with δ Scuti variables and eclipsing contact binaries (see Fig. 2.13 to see this clear overlap). Furthermore the so-called W Ursae Majoris (W UMa) variables have sinusoidal light curves similar to RRc variables (an example of these lightcurves is shown in Fig. 2.10), leading to overlap in diagnostics based on Fourier parameters, especially in cases where the sampling of the light curve is either poor or the photometry is noisy so that the typical secondary eclipse of W UMa light curve can be barely detected. While eclipsing binaries can have similar periods and similar light curve shapes to RR Lyrae, they obviously do not follow the same period-luminosity relation, thus degrading the quality of the measured distances. This is the main reason why RRc variables are only partially used in distance and in metallicity estimates (Contreras Ramos et al., 2018)

- Mode identification: the light curves of fundamental and first overtone RRL have different shapes. The former ones when located to the blue edge (larger amplitudes) have a saw-tooth

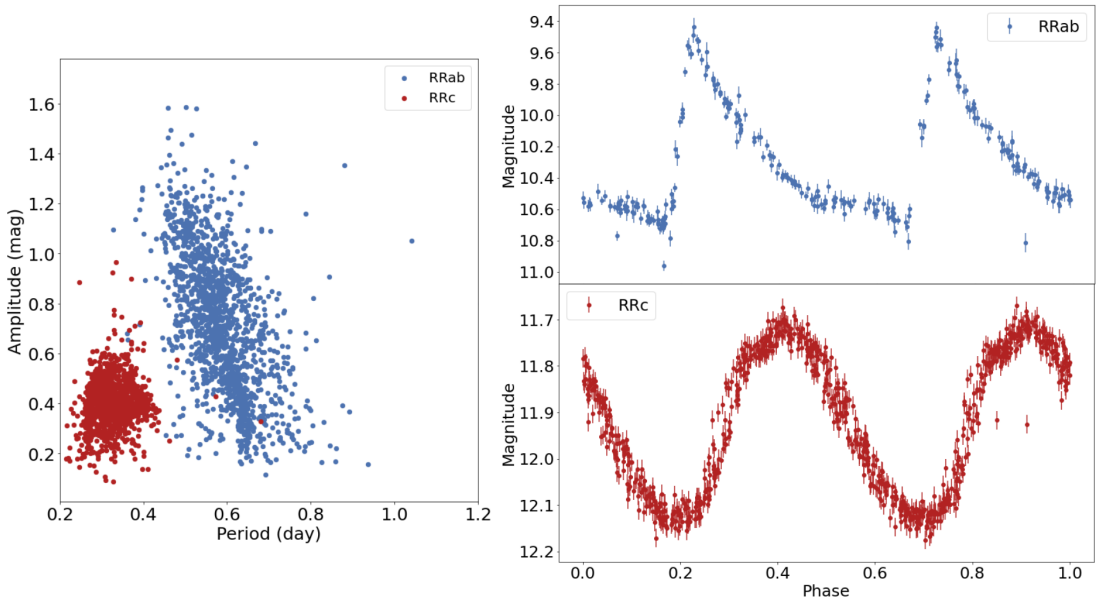


Figure 2.2: On the *left panel* an example of bailey diagram (sample of RR Lyrae from Catalina Sky Survey. We can see that RRc are clustered in the region of period 0.2-0.4 days amplitudes of 0.2-0.6 mag, meanwhile the RRab span bigger amplitude and have longer periods. On the *top right panel* we can see clearly the saw-tooth shape of RRab lightcurve compared to the more symmetric sinusoidal behaviour of a RRc lightcurve *bottom right panel*.

shape, while the latter ones when located close to the blue edge are sinusoidal (see Fig. 2.2). However, there is a region of the instability strip, the so-called "OR" region in which they can pulsate either in the fundamental or in the first overtone or simultaneously in the first two radial modes (RRd, mixed mode). The reader interested in a more detailed discussion concerning the possible occurrence of non-radial modes in RRL is referred to Smolec (2005)

- RRL secondary modulations: The current empirical evidence indicates that roughly the 40% of RRL are affected by secondary modulations (Prudil and Skarka, 2017). The period is relatively constant, but the amplitude can change by several tenths of a magnitude. This phenomenon is called Blazkho effect, and we still lack a firm theoretical understanding of the physical mechanism(s) driving its occurrence.

- RRL phasing: RRL have periods ranging from a few hours to roughly one day. This means that they form thousands of cycles in a few years. Photometric or spectroscopic data collected over a time interval of years or decades do require very accurate pulsation periods to be properly phased.

Some of the methods applied in this work address these issues: the overlap with the eclipsing binaries is addressed in Sec. 2.4, the mode identification and the characterization of the light curves both in Sec. 2.5 and Sec. 2.6 with also hints to secondary modulations in the latter section.

δ Scuti: If we move to shorter period and low amplitudes we find δ Scutis stars. These stars did not become prevalent in my PhD work but we applied the method presented in Pasquato et al. (2022) also to them and a short introduction is mandatory. δ Scutis are short period (P $\approx$ 0.3 d) variable stars and have amplitudes A $<$ 1.0 mag in the V band; they are divided into low-amplitude

(LADS) and high-amplitude (HADS) δ Scutis (Alcock et al., 2000). In addition, metal-poor variables with similar periods and old ages are usually referred to as SX Phoenicis stars (SX Phe, Nemec and Mateo 1990). Both δ Scuti and SX Phe are located where the main sequence crosses the instability strip in the HR diagram (see Fig. 2.1). δ Scutis pulsate in radial and non-radial modes and the driving mechanism for their pulsations is similar to that of RRL stars. This mechanism is relatively well understood, depending on the opacity and equation of state variations (the so-called kappa and gamma mechanisms) in the partial ionization region of both Hydrogen and Helium.

2.1.2 Eclipsing Binaries

We can divide Eclipsing Binaries (EBs) based on the detachment of the 2 stars: EW if the two stars share a common envelope and EA if they are detached:

EW (W UMa type): EW consist of two stars that share a common envelope. Their outer layers are in physical contact, so their surfaces essentially overlap. As such, they do not show a pronounced, separate dip in brightness for each star being eclipsed (bottom panel of Fig. 2.3). Their light curves typically exhibits a continuous, sinusoidal variation in brightness with no flat baseline. This is because there's no period during which one star is completely visible without any overlap from the other. The depth of the eclipse is influenced by the ratio of the luminosities and radii of the two stars:

$$\frac{\Delta F}{F} = \frac{\pi R_2^2 L_2}{\pi R_1^2 L_1}$$

However, since both stars in a contact binary often have similar temperatures and are in contact, the eclipse depth for primary and secondary eclipses is relatively similar. Such systems involve mass exchange between stars, generally marked by similar temperatures and characteristics. The depth of the eclipse in these systems is predominantly dictated by the geometry of the system. Yet, variations in eclipse depth can emerge due to discrepancies in the temperatures of the component stars.

Algol Binaries (EA): EA binaries are detached, meaning the two stars are distinct and separated, not touching each other. This leads to clear and distinct eclipses in the light curve when one star passes in front of the other.

As top panel of Fig. 2.3 depicts, the light curve of an EA binaries typically shows distinct and sharp dips (eclipses) separated by periods of constant brightness when the stars are not eclipsing each other. The depth of the eclipse follows the same equation above. The primary eclipse (when the brighter star is eclipsed) is deeper if the second star is cooler and smaller. Conversely, the secondary eclipse (when the cooler, smaller star is eclipsed) is shallower. This differential is often pronounced in EA systems because the stars can have different temperatures and sizes. In detached systems, the stars can be closer or further apart, affecting the relative duration of the eclipses. However, compared to contact binaries, the flat portions of the light curve (when neither star is eclipsed) are generally more extended.

2.2 Data

I present here the different data utilized in the three studies.

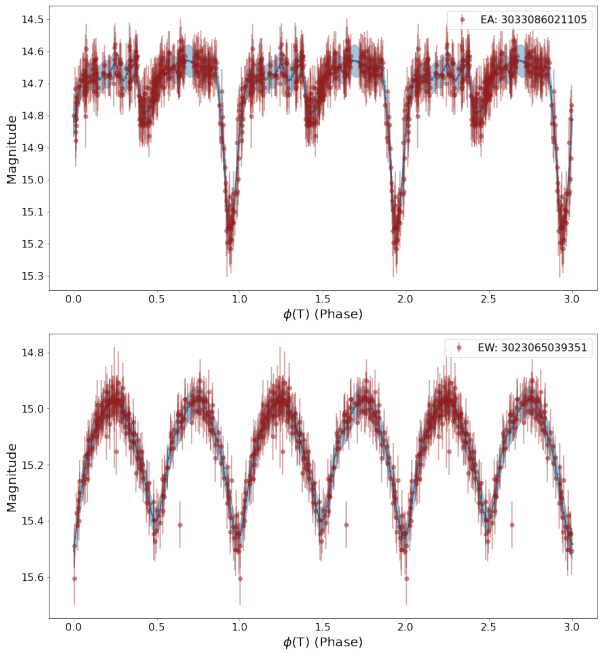


Figure 2.3: Light curves of eclipsing binaries from the Catalina Schmidt Survey (CSS): in the *top panel* a sample lightcurve of a Algol Type binaries. In the *bottom panel* a sample lightcurve of a W UMa type. Note the difference in the depth of the eclipses for the two types.

2.2.1 Catalina Sky Survey

We use time-series data from the Catalina Sky Survey (CRTS) (Drake et al., 2013a). The CRTS consists of three different telescopes: the Catalina Schmidt Survey (CSS), the Mt. Lemmon Survey (MLS), and the Siding Spring Survey (SSS). The CSS and MLS are located in Tucson, Arizona while the SSS in Australia. Overall, these telescopes scanned $\sim$ 33000 deg^2 of the sky in the -75° to $+65^\circ$ declination range.

The CSS is equipped with a 0.7-m Schmidt telescope and an unfiltered 4k $\times$ 4k CCD with an 8 deg^2 field of view (FOV). The CSS' exposure time is $\sim$ 30s and it observes objects in the $\sim$ 11.5–19.5 magnitude range. The MLS survey uses a 1.5-m Cassegrain reflector telescope and an unfiltered CCD (1.2 deg^2 FOV). The MLS bright and faint magnitude cut-offs are 13.0 and 21.5 mag, respectively. Finally, the SSS is equipped with the smallest 0.5-m Schmidt telescope, an unfiltered 4k $\times$ 4k CCD, and has a 4.2 deg^2 FOV. It observes objects in the $\sim$ 11.0–19.0 magnitude range.

The CRTS was originally designed to discover, track, and catalog near-Earth objects, including potentially hazardous asteroids. At the same time, the CRTS data was used to study variable sources and a full catalog of CRTS variable stars was published in a series of papers (e.g., Drake et al. 2013a,b, 2014b, 2017, Torrealba et al. 2015).

To search for variable sources, the authors of the latter studies followed a systematic approach. At first, they distinguished variable from non-variable sources using the Welch-Stetson variability index (I_{WS}; Welch and Stetson 1993). Then, they used the Lomb-Scargle periodogram analysis (LS; Lomb 1976, Scargle 1982) to look for periodicity signals and they used the M-Test (Kinemuchi et al., 2006) to measure how much time was spent by each star above or below the mean magnitudes. They then calculated the periods of variability using different period finding algorithms including the Analysis of Variance (AoV; Schwarzenberg-Czerny 1989) technique.

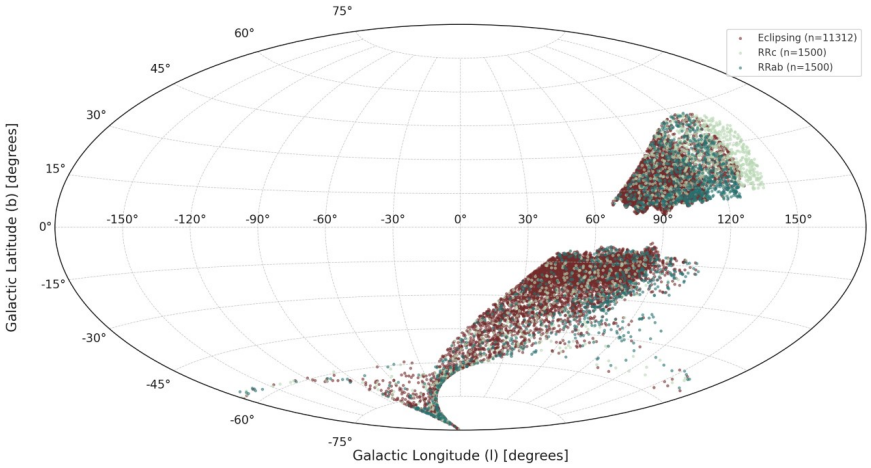


Figure 2.4: Positions of RRab/RRc and eclipsing stars in galactic coordinates from the Catalina sample. We can see that this sample does not include regions around the Galactic plane.

The combined catalogs of variable stars results in $\sim$ 110,000 variable sources. Each object was observed between $\sim$ 25 and 600 times with an average of $\sim$ 200 epochs. For objects brighter than $V = 17$ mag, the completeness and purity levels for RRab stars reach $\sim$ 90% and $\sim$ 80%, respectively. These values decrease for fainter objects and for RRc and δ Scuti variables. Finally, the identification numbers (ID_{CRTS}), equatorial J2000.0 right ascension (R.A.) and declination (Dec.) coordinates, average magnitudes ($\langle V \rangle$), periods in days (P), amplitudes (A), number of observations (N_{CRTS}), and other properties are included in the published catalogs (in Fig.2.5 I report some relevant properties for the studies divided by variable type). On our adopted sample the median time elapsed between the first and the last observation was 2794 days for δ Scuti variables, 2787, and 2703 days respectively for RRc and RRab variables.

An issue worth noting is that Catalina data mostly do not include dense regions around the Galactic plane (see Fig. 2.4), which might affect the generality of our results given the expected dependency of variable star properties on metallicity. Our choice of CSS is, however, justified by its large number of relatively homogeneous light curves, which is a crucial prerequisite for training most machine-learning models.

We point out that CSS images were collected without filters. The conversion to the V-band was performed by setting up a sample of comparison stars (G0-G8 dwarfs) from the 2MASS Point Source Catalog (Skrutskie et al., 2006), converting to the $JHKs$ bands (first step) and then to the V band (second step) (see Larson et al., 2003, Skrutskie et al., 2006, for details).

Different folded-lightcurves of the adopted datasets are shown in Fig. 2.10.

2.2.2 ASAS/ASAS-SN

For the test set of the classifier paper, we utilized 614 stars from the All Sky Automated Survey data (ASAS; Pojmanski, 1997, 2002) and All-Sky Automated Survey for Supernovae (ASAS-SN; Jayasinghe et al., 2019, Shappee et al., 2014). ASAS-SN is an extensive optical survey designed to monitor the entire sky with a cadence of approximately 2-3 days to a depth of V 17 mag. Since its inception in 2014, ASAS-SN has been collecting roughly 100-500 epochs of observations per field. The survey's methodology involves the classification of V-band light curves for candidate variables identified during the search for supernovae. The catalogue has 66179 new bright variable stars, which

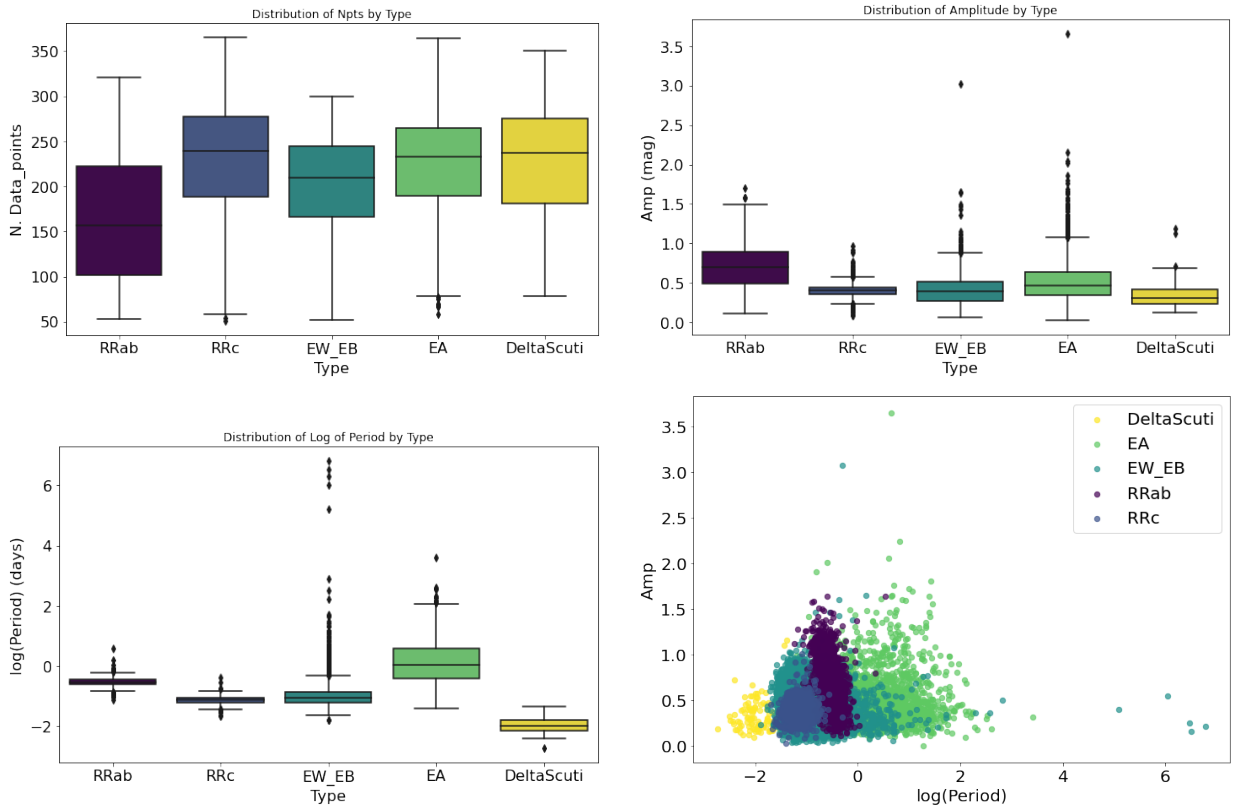


Figure 2.5: Initial adopted sample of CSS. *Top-left panel:* the boxplot of the number of data points. *Top-right panel:* boxplot of the amplitudes. *Bottom-left panel:* boxplot for the period and finally in the *bottom-right panel* the bailey diagram of the initial adopted sample.

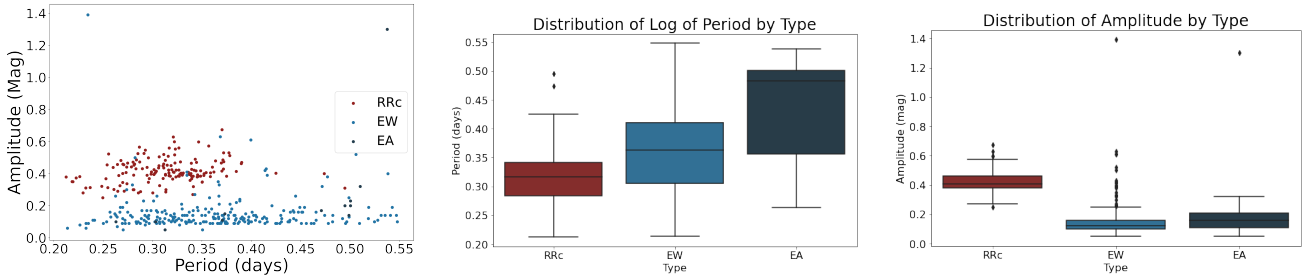


Figure 2.6: Bailey diagram on the left for the ASAS-ASAS-SN sample, some statistics on the period (*center panel*) and amplitude *right panel*).

Table 2.1: Number of RRc and RRab light curves included in the ω Cen initial and final dataset for each filter. Details concerning the selection are specified below.

Filter	$RRc_{initial}$	RRc_{final}	$RRab_{initial}$	$RRab_{final}$
V	46	45	36	36
B	30	27	29	28
I	3	3	4	3
R	10	8	1	1
Total	89	83	70	68

includes 27479 periodic variables and 38700 irregular variables.

Some of the properties of sample adopted are shown in Fig. 2.6 This dataset differs from our CSS dataset because it was taken by a different instrument and also displays a different class balance: RRL are 12% of our CSS sample and 35% of our ASAS/ASAS-SN test sample. The V-band and g-band magnitudes obtained in the ASAS/ASAS-SN catalogue are calculated in real-time using aperture photometry, with zero-points calibrated using the APASS catalog. ASAS-SN images have 8" pixels (∼15" FWHM PSF) so blending/crowding might be significant in some cases.

2.2.3 ω Cen data

For the Dynamical Mode Decomposition work (see Sec.2.6), we adopted the optical (BVI-band) photometry of the RRLs in the Globular cluster ω Cen published in Braga et al. (2016, 2019a). This data set is homogeneous in terms of photometric data reduction, even though it includes images collected during more than 20 years from seven different 1-to-8 m class telescopes, in a sky area of 57'×56' (see Sec.2 in Braga et al. 2016 for more details). Still, to further improve the consistency of our analysis, we chose to adopt only the photometric data collected at the same telescope (Danish 1.5m, at La Silla), during five years (1995 to 1999), which constitutes the largest part (almost 60%) of the original data set. The resulting Danish subset is a truly homogeneous data set, with the exception of sporadic positional offsets in the 1998 images. As a result, the analysis is based on light curves with 100 to over 2000 phase points in at least two of the three bands (V and B, since I-band phase points are one order of magnitude less) for 125 variables.

This subset of the data is comprised of light curves taken from 89 RRc and 70 RRab variables. Of these, 6 RRc and 2 RRab light curves were discarded because we failed to obtain an acceptable interpolation, due either to outliers or phase gaps. This leaves us with 83 RRc and 68 RRab light curves. Tab. 2.1 summarizes the number of available light curves in the various photometric bands.

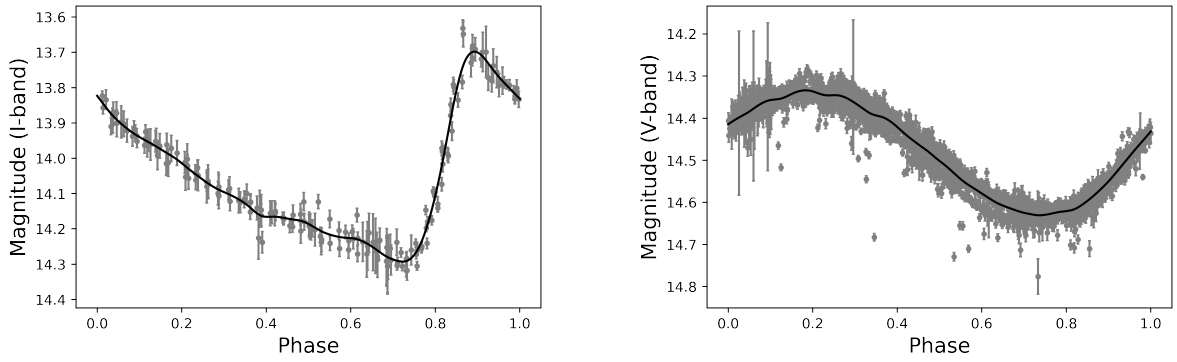


Figure 2.7: Left panel: Light curve for star V59, a typical RRab variable from ω Cen sample. Observational data points with error bars are shown in gray, our spline fit is shown as a black solid line. Right: Light curve for star V103, a typical RRc variable.

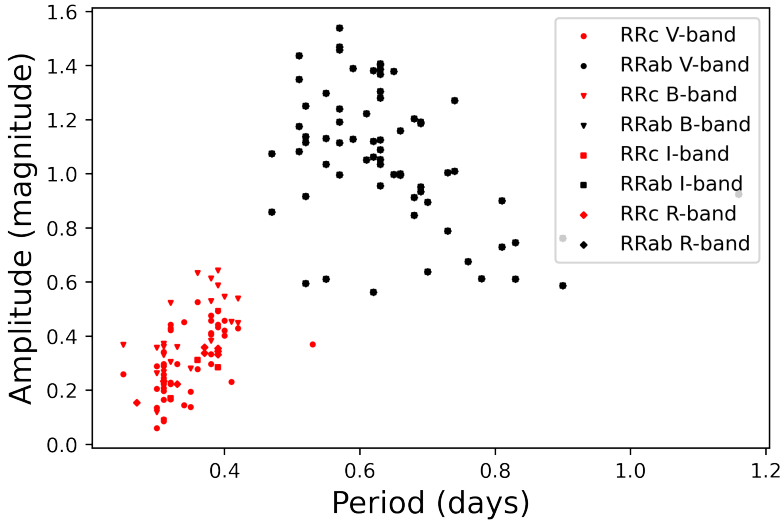


Figure 2.8: Adopted sample from the *omega* Cen sample in the period-amplitude plane, RRab and RRc stars are respectively represented by black and red points.

Typical light curves for RRab and RRc are shown in Fig. 2.7, in the left and right panels respectively. The Bailey diagram is shown in Fig. 2.8 instead.

2.3 Methods

I present in this section the methods applied in the different works for variable stars light curve characterization.

2.3.1 Gaussian Processes

Given a set of points there are infinite functions that can fit these points. Gaussian Processes (GPs) offers an elegant solution to this problem[1]. I will define now what is a gaussian process and some

[1]GPs can be used for a lot of tasks: regression, classification, optimization, active learning, reinforcement learning and in general every branch of ML. As we can see in Sec. 2.4 I used to fit the variable stars lightcurves so the

properties following (Kanagawa et al., 2018) with a focus on the kernels applied to our work.

Gaussian processes: Let X be a nonempty set, $k : X \times X \to \mathbb{R}$ be a positive definite kernel, and $m : X \to \mathbb{R}$ be any real-valued function. Then a random function $f : X \to \mathbb{R}$ is said to be a Gaussian process (GP) with mean function m and covariance kernel k, denoted by $\mathcal{GP}(m, k)$, if the following holds:

For any finite set $X = (x_1, \ldots, x_n) \subset X$ of any size $n \in \mathbb{N}$, the random vector

$$f_X = \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix}$$

follows the multivariate normal distribution $\mathcal{N}(m_X, k_{XX})$ with covariance matrix

$$k_{XX} = \begin{pmatrix} k(x_1, x_1) & \ldots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \ldots & k(x_n, x_n) \end{pmatrix}$$

and mean vector

$$m_X = \begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix}$$

.

This definition implies that if f is a Gaussian process then there exists a mean function $m : X \to \mathbb{R}$ and a covariance kernel $k : X \times X \to \mathbb{R}$. At the same time, it is also true that for any positive definite kernel k and mean function m, there exists a corresponding Gaussian process $f \sim \mathcal{GP}(m, k)$ (chapter 12 in Dudley, 2018). Therefore, there is a one-to-one correspondence between Gaussian processes $f \sim \mathcal{GP}(m, k)$ and pairs (m, k) of mean function m and positive definite kernel k. In practice, the use of a specific kernel k and a mean function m implicitly leads to the use of the corresponding $\mathcal{GP}(k, m)$. Therefore it is practically important to understand the properties of $\mathcal{GP}(k, m)$ (such as smoothness) that are induced by the specification of k and m. For example, if k is a square-exponential kernel on an open set $X \subset \mathbb{R}^d$, then a sample path $f \sim \mathcal{GP}(0, m)$ is infinitely continuously differentiable.

I will present now different types of kernels utilized in the work presented in Sec. 2.4 The specification of the kernel k force properties of the distribution over functions $f(x)$, i.e. by choosing a specific kernel function, it is possible to set prior information on this distribution.

Exponential Sine Squared Kernel: Also known as the periodic kernel (in `scikit-learn` is called `ExpSineSquared`) is used for modeling functions that repeat themselves exactly. It is defined as:

$$k(x_i, x_j) = \exp\left(-\frac{2\sin^2(\pi|x_i - x_j|/p)}{l^2}\right)$$

explanation will focus on this but GPs are not restricted to fitting

where l is the length scale and p the period of the kernel. This kernel can be derived as a stationary solution to the following Stochastic Partial Differential Equation (SPDE):

$$\left(\frac{2\sin(\pi x/p)}{l^2} - \Delta\right)^2 f(x) = w(x), \quad x \in \mathbb{R}$$

where Δ is the Laplacian operator and $w(x)$ is a Gaussian white noise process.

The Matérn kernel is defined as:

$$k(x_i, x_j) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{|x_i - x_j|}{\rho}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{|x_i - x_j|}{\rho}\right)$$

where ν is a parameter that controls the smoothness of the resulting function, ρ is the length scale and K_ν is a modified Bessel function of order ν. This corresponds to the stationary solution to the following SPDE

$$\left(\frac{2\alpha}{h^2} - \Delta\right)^{(\alpha+\frac{h}{2})/2} f(x) = w(x), \quad x \in \mathbb{R}^d,$$

where $\Delta := \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ is the Laplacian and α corresponds to the smoothness parameter ν and h in the SPDE can sometimes be related to the length scale ρ from the Matérn kernel, but the exact relationship might be context-dependent.

Usually Matern kernels forces smoothness on the resulting function allowing for better interpolation. In general the length scale both in the Matern and Exponential Sine Squared Kernel controls the smoothness of the function modeled by the Gaussian process. A smaller value of l means that the function values can change rapidly, indicating that the function is wiggly or has high variance over short distances. Conversely, a larger l implies that the function varies more slowly, indicating a smoother function with lower variance over short distances.

Finally we have:

Gaussian noise and constant kernel: The `WhiteKernel` in `scikit-learn` generates a component of Gaussian noise. It is defined as:

$$k(x_i, x_j) = \sigma^2 \delta_{ij}$$

where σ^2 is the noise level and δ_{ij} is the Kronecker delta function. This corresponds to the SPDE:

$$f(x) = w(x), \quad x \in \mathbb{R}^d$$

where $w(x)$ is a Gaussian white noise process with variance σ^2. The constant kernel instead, as its name suggests, returns a constant value. It is defined as:

$$k(x_i, x_j) = c^2$$

where c is the constant value.

The noise component of the kernel is fundamental when working with noisy data. A larger noise term allows the GP to accept that observed data might be noisy and it doesn't need to fit them perfectly. This is particularly useful for regression problems where there is known measurement

error in the observed outcomes. If the noise term is close to zero or very close to zero, the GP will attempt to interpolate the data, essentially trying to pass through every point resulting in a poor interpolation. The structure and hyperparameters of the kernel function determine the flexibility of the GP. For instance, a very flexible kernel (or a kernel with certain hyperparameters) might try to fit closely to the data, including the noisy fluctuations, while a less flexible kernel will provide a smoother function that doesn't capture the high-frequency noise

We can combine all these kernels meaning we add/multiply their respective covariance matrices together to obtain the final covariance matrix used in Gaussian Process regression (example of this process in Fig. 2.14).

When I fit the Gaussian process to my test data (X_{test}, y_{test}) (here test has a different meaning from the training-validation-test one), the predictions made are conditioned on the data. This has an effect on the covariance matrix. The posterior covariance matrix, after observing the data, is a modified version of the prior covariance matrix. Specifically, it is:

$$K_{\text{posterior}}(X_*, X_*) = K(X_*, X_*) - K(X_*, X)(K(X, X) + \sigma_n^2 I)^{-1} K(X, X_*)$$

Where:

- X_* are the test inputs.
- $K(X_*, X)$ is the cross-covariance matrix between the test inputs and the training inputs.
- $K(X, X)$ is the covariance matrix of the training inputs.
- σ_n^2 is the noise variance (given by the WhiteKernel).
- I is the identity matrix.

This posterior covariance matrix captures the effect of the data on our uncertainty. Where we have data, the uncertainty will be reduced, and this will be reflected in the posterior covariance matrix and the pair (m, k).

2.3.2 (Regularized) Logistic Regression

An interpretable model not only is a native advantage being a white box but also should have aim to have some of the following properties in order to be human-understandable: linearity, monotonicity and not too many cross-interaction between features. A model is considered linear when the link between its features and the target is linearly represented. When a model has monotonicity constraints, it means the relationship between a feature and its target always remains consistent across the feature's full range. Specifically, when the feature value rises, it either consistently leads to the target outcome rising or declining. Monotonicity aids in comprehending a model since it simplifies understanding of a relation. Also we can include interactions in any type of model by manually creating interaction features. Interactions can improve predictive performance, but too many or too complex interactions can hurt interpretability. Logistic regression checks the box of full transparency (not black box), monotonicity and, if not manually created, absence of interaction between features.

Given our dataset $\mathcal{D}$ represented by $X \in \mathcal{R}^{d \times m}$ where d is the number of features and m the number of examples, we can express the feature vector (for example i) $X^i = [x_1^i, x_2^i, ..., x_d^i]$, as a

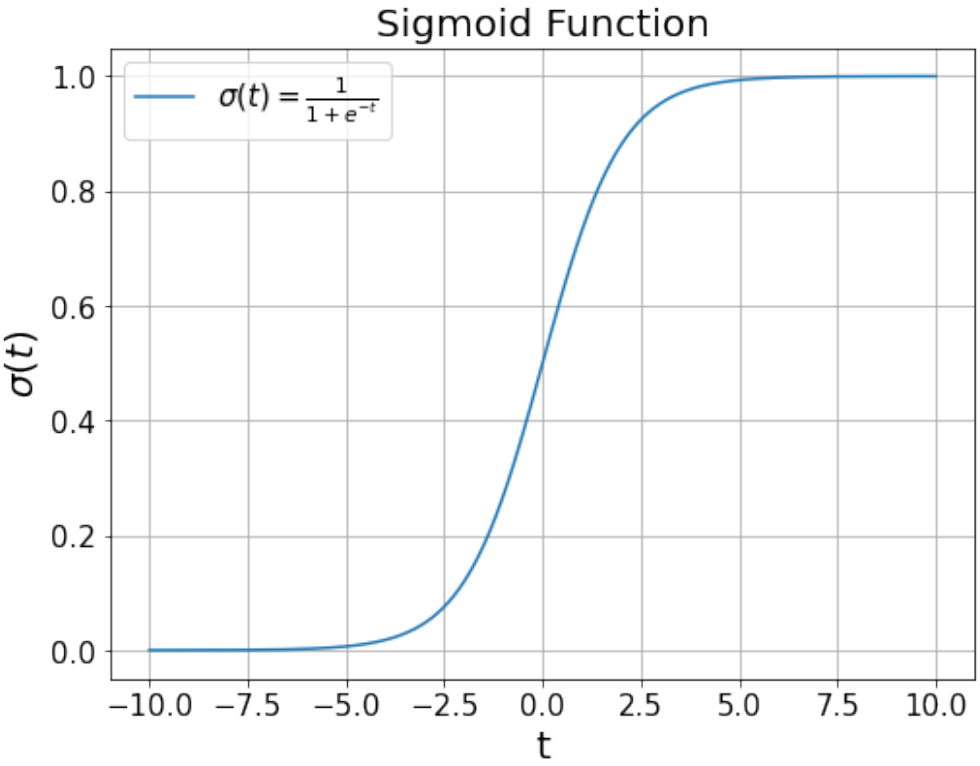


Figure 2.9: Sigmoid function.

linear combination of features with the model's coefficients is given by (let's drop the apex i from here):

$$t = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_d x_d$$

The probability that the dependent variable Y is 1 (or success) given the feature vector X is:

$$p(X) = \sigma(t) = \frac{1}{1 + e^{-t}}$$

$$p(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_d x_d)}}$$

From a probabilistic perspective:

$$P(Y = 1|X) = p(X)$$

$$P(Y = 0|X) = 1 - p(X)$$

We can see from fig. 2.9 that the sigmoid function has a bounded y-range between 0 and 1. The sigmoid is differentiable everywhere. This property is crucial for optimization methods such as gradient descent. Not only, it is also monotone, a property, as I said in the introduction, really important for interpretability. Given this, the likelihood of observing the actual data is:

$$\mathcal{L}(\beta|X, Y) = \prod_{i=1}^{m} p(X^i)^{Y^i} (1 - p(X^i))^{(1-Y^i)}$$

Where m is the total number of examples in the dataset and β represents the vector of coefficients, including the intercept.

In practice, since we are adopting a ML approach, the goal during the training process is to maximize this likelihood, which is done by minimizing the negative log-likelihood:

$$-l(\beta|X,Y) = -\sum_{i=1}^{m} \left[Y^i \log(p(X^i)) + (1-Y^i)\log(1-p(X^i))\right]$$

The negative log-likelihood can be thought of as a loss function (for the single training example) or cost function (the average of the loss functions of the entire training set). For logistic regression, the goal of the training process is to find the coefficients β that minimize this loss function. To minimize the negative log-likelihood, optimization algorithms are employed. One common method is via gradient descent. The general idea is:

1. Initialize the coefficients β with some values, often randomly.
2. Calculate the gradient of the negative log-likelihood with respect to β. This gradient points in the direction of the steepest increase of the function.
3. Update β in the opposite direction of the gradient to reduce the negative log-likelihood. This involves taking a step size, often called a learning rate, in the direction of the negative gradient.
4. Repeat steps 2 and 3 until the coefficients converge, or until a specified number of iterations is reached.

During training, the coefficients are updated iteratively to find the best fit for the data, such that the model's predicted probabilities are as close as possible to the observed outcomes Y in the dataset. Usually, the ML packages adopt some optimizer that are more refined than the process described above. We can regularize our logistic regression to prevent overfitting and introduce sparsity. In the classifier paper we utilize the L1 regularization (or L1 penalty sometimes in the text). Here's how the L1 regularization is incorporated into the logistic regression model:

Given our negative log-likelihood $\mathcal{J}(\beta) = -l(\beta|X,Y)$:

$$\mathcal{J}(\beta) = -\sum_{i=1}^{m} \left[Y^i \log(p(X^i)) + (1-Y^i)\log(1-p(X^i))\right]$$

The L1 penalty is the sum of the absolute values of the coefficients:

$$L1(\beta) = \lambda \sum_{j=1}^{n} |\beta_j|$$

where λ is the regularization strength and β_j represents the j-th coefficient in the vector β.

When we incorporate the L1 regularization into the logistic regression model, our objective function becomes:

$$\mathcal{J}_{L1}(\beta) = -\sum_{i=1}^{m} \left[Y^i \log(p(X^i)) + (1-Y^i)\log(1-p(X^i))\right] + \lambda \sum_{j=1}^{n} |\beta_j|$$

The goal during training is to minimize $\mathcal{J}_{L1}(\beta)$. A larger value of λ will result in more regularization:

- When $\lambda = 0$, the objective becomes the original negative log-likelihood, meaning no regularization.
- As λ increases, the influence of the regularization term grows, and the model will tend to have smaller coefficients, leading to a simpler and potentially more interpretable model.
- With L1 regularization, as λ becomes very large, some coefficients might become exactly zero, leading to a sparser model.

In `scikit-learn`'s `LogisticRegression` implementation, the inverse of regularization strength, `C`, is used. `C` is defined as $\texttt{C} = \frac{1}{\lambda}$. Thus, smaller values of `C` imply more regularization. Sparsity of the coefficients given by the L1 regularization has different benefits:

- Interpretability: A sparse model has fewer non-zero coefficients. This means that only a subset of features significantly influence the prediction. This makes the model more interpretable, as it is easier to explain the effect of a small number of features rather than a large set. This was our drive for our work: we wanted a white box, but if this one has too many features (like the 100 of our dataset) it is not actually human interpretable.
- Feature Selection: Sparsity acts as an embedded feature selection mechanism. Meaning that, by driving irrelevant feature weights to zero, it effectively selects a subset of the most informative features for the prediction task.
- Efficiency: Finally, sparse models are computationally more efficient; they require less memory storage
- Generalization: By reducing the model's complexity, sparsity can help in preventing overfitting. A model that relies on fewer features is less likely to fit the noise in the training data and is more likely to generalize well to new, unseen data.

2.3.3 SINDY

Understanding complex dynamical systems from data is a fundamental challenge in many areas of science and engineering. Traditional modeling techniques, like differential equations, can be hard to determine for high-dimensional, non-linear systems from empirical data alone. The SINDy algorithm (Brunton et al. 2016) is designed to determine sparse dynamical systems directly from data without the need for a detailed system knowledge or guesswork.

Given the state measurements of a dynamical system, the goal is to identify the most parsimonious dynamical system that describes the data. Parsimonious here implies a system that is described using the fewest possible terms or functions, eliminating unnecessary complexities. Remember, the aim is to capture the underlying dynamics with as few terms as necessary to ensure that the model remains interpretable and avoids overfitting. Given a dynamical system, the evolution of its state can be represented as:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}(t))$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ is the state vector of the system at time t.

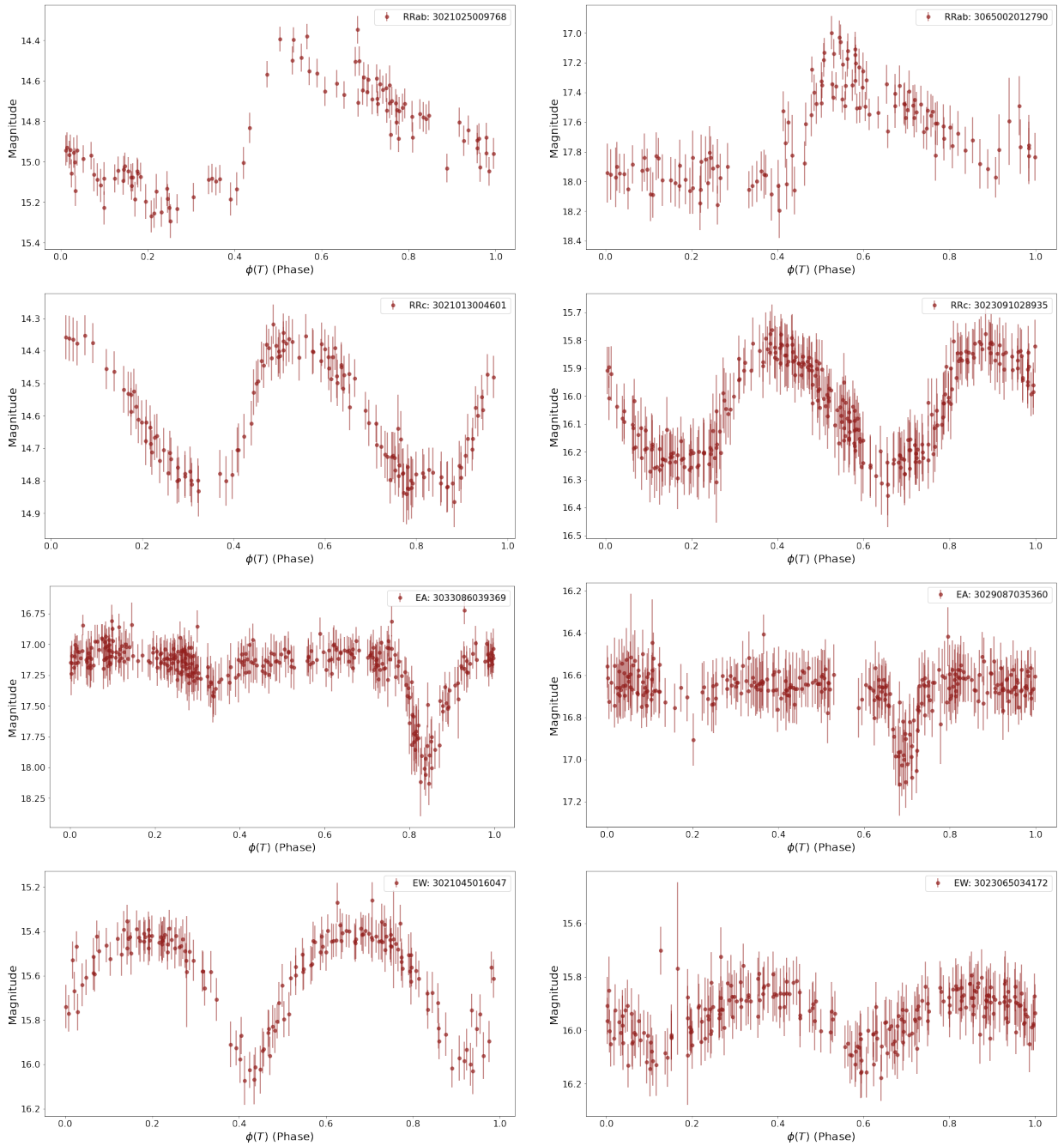


Figure 2.10: Examples of phase-folded lightcurves from CSS of RRab (*first panel*), RRc stars(*second panel*), EA binaries (*third panel*) and EW binaries (*fourth panel*).

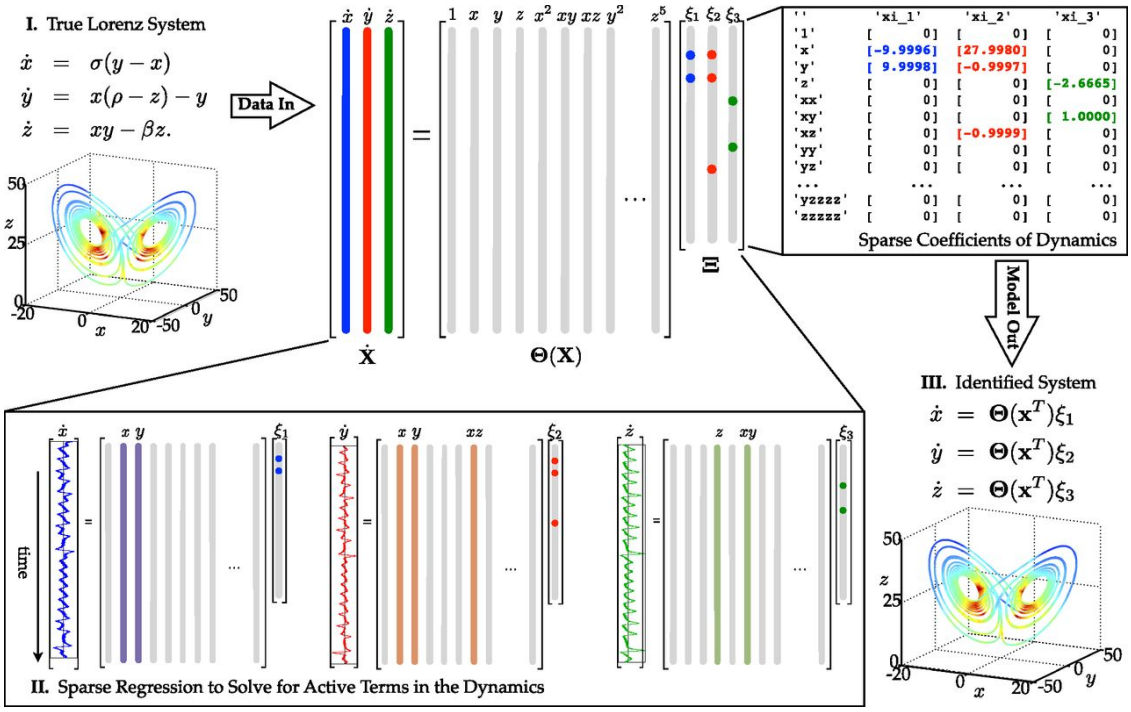


Figure 2.11: Schematic of the SINDy algorithm, demonstrated on the Lorenz equations. Credits to Brunton et al. (2016)

The primary idea behind SINDy is to approximate the function f from data. In order to do this we need to first collect data of the system state $\mathbf{x}(t)$ and its time derivative $\frac{d\mathbf{x}}{dt}$ over a period of time. Then build a library of candidate functions $\Theta(\mathbf{x})$, which includes polynomial terms, trigonometric functions, or any other nonlinear functions of the state variable. For example:

$$\Theta(\mathbf{x}) = \begin{bmatrix} 1 & \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_1^2 & \mathbf{x}_1\mathbf{x}_2 & \mathbf{x}_2^2 & \sin(\mathbf{x}_1) & \dots \end{bmatrix}$$

Afterwards, use sparse regression techniques to solve:

$$\frac{d\mathbf{x}}{dt} \approx \Theta(\mathbf{x})\boldsymbol{\xi}$$

where $\boldsymbol{\xi}$ is a coefficient vector. Sparse regression ensures that many of the coefficients in $\boldsymbol{\xi}$ are zero, leading to a parsimonious representation of f.

Finally, The nonzero terms in the vector $\boldsymbol{\xi}$ correspond to the active terms in the governing equations. For instance, if only the coefficients associated with 1, $\mathbf{x}_1$, and $\mathbf{x}_1^2$ are nonzero, then the dynamics are approximated as:

$$\frac{d\mathbf{x}}{dt} \approx \xi_0 + \xi_1\mathbf{x}_1 + \xi_2\mathbf{x}_1^2$$

In Fig. 2.11, an application on a True Lorenz System is shown. Remembering the basic least squares problem (with no sparsity) where $\dot{\boldsymbol{x}}$ is used instead of $\frac{d\mathbf{x}}{dt}$:

$$\min_{\boldsymbol{\xi}} \|\boldsymbol{\Theta}(x)\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2$$

different optimization techniques are available in `PySINDy` (de Silva et al., 2020) +`scikit-learn`:

Sequentially Thresholded Least Squares (STLSQ): Iteratively applies a thresholding step after least squares.

$$\boldsymbol{\xi}^{(k+1)} = \arg\min_{\boldsymbol{\xi}} \|\boldsymbol{\Theta}(x)\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2$$

followed by:

$$\xi_i^{(k+1)} = 0 \quad \text{if} \quad |\xi_i^{(k+1)}| < \lambda$$

where λ is a threshold value. The algorithm iterates between the two steps until convergence.

LASSO (L1 Regularized Least Squares): LASSO adds a penalty to the absolute values of the coefficients.

$$\min_{\boldsymbol{\xi}} \|\boldsymbol{\Theta}(x)\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2 + \lambda\|\boldsymbol{\xi}\|_1$$

where: λ is a regularization parameter (`alpha` in `scikit-learn`).

Elastic Net: Combines both L1 and L2 regularization.

$$\min_{\boldsymbol{\xi}} \|\boldsymbol{\Theta}(x)\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2 + \lambda_1\|\boldsymbol{\xi}\|_1 + \lambda_2\|\boldsymbol{\xi}\|_2^2$$

where λ_1 and λ_2 are regularization parameters for the L1 and L2 penalties, respectively.

Group LASSO: Applies the L1 penalty to groups of coefficients.

$$\min_{\boldsymbol{\xi}} \|\boldsymbol{\Theta}(x)\boldsymbol{\xi} - \dot{\boldsymbol{x}}\|_2^2 + \lambda \sum_j \sqrt{p_j}\|\boldsymbol{\xi}_{\text{group j}}\|_2$$

where p_j is the number of coefficients in the j-th group, $\boldsymbol{\xi}_{\text{group j}}$ are the coefficients in the j-th group andλ is a regularization parameter.

For an overview on SINDy applications, see table 1 of Dong et al. (2023) and the references therein. For the proof of convergence of SINDy, see Zhang and Schaeffer (2019). Finally, as the same Steve Brunton and collaborators noted in Brunton (2021), the three main challenges of SINDy are: which variables should we use, what library of functions should we use to describe the dynamics and finally what algorithm is better to carry out the optimization. Initially, they thought that the optimizer would be the biggest problem during SINDy applications. However, as we discovered during the work described in Sec. 2.5, the choice of the library is not so trivial and is the main reason why certains stars were not modelled correctly.

2.3.4 Koopman Analysis and DMD

Dynamic Mode Decomposition (DMD) algorithm is a dimensionality reduction algorithm guaranteed in the framework of Koopman analysis for the study of nonlinear dynamical systems. The core principle of Koopman analysis is a non linear coordinate transformation that embeds a nonlinear dynamical system in an equivalent linear dynamical system of observables. Mathematically, if ψ is an observable function of the system's state $\mathbf{x}$, the action of the Koopman operator, U, can be expressed as:

$$U\psi(\mathbf{x}_0) = \psi(f(\mathbf{x}_0)) \tag{2.1}$$

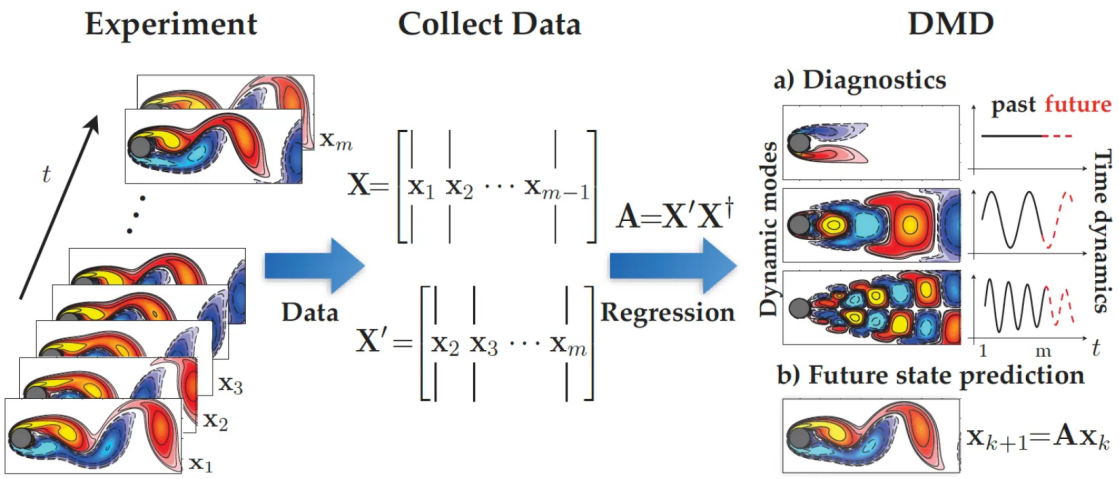


Figure 2.12: Schematic of the Dynamic Mode Decomposition on a fluid flow dataset. *Credits:* Kutz et al. (2016)

where f denotes the nonlinear evolution function of the system. The existence of such an embedding is due to an intrinsic trade-off between linearity and dimension. In this framework, it is necessary to figure out what the coordinate transformation is and what the dynamical system is turned into. For our purpose, we are interested to do this from data in one shot. This is possible thanks to tangent linear approximation of the system and a few arguments used in iterative methods for computing solutions to linear eigenvalue problems, that DMD algorithm naturally encodes. More precisely, in our case, given a discrete time series $V_1^N = \{v_1, \dots, v_N\}$, where $v_i \in \mathbb{R}^m \ \ \forall i \in [m]$, it is assumed the existence of a linear mapping $A \in GL_n$ that associates two successive points of the time series, i.e. $v_{i+1} = Av_i$. This last relation defines a linear dynamical system that stays approximately the same over the sampling period, formally using the language of matrices $V_2^N = AV_1^{N-1} + re_{N-1}^T$, where r is the vector of residuals that encodes the behaviour not predicted by A, e_{N-1} is the $(N-1)$th vector of the canonical base and $V_1^{N-1} = \{v_1, \dots, v_{N-1}\}$ and $V_2^N = \{v_1, \dots, v_N\}$. Now comes the power of DMD: roughly speaking, DMD computes the eigenvalues and the eigenvectors of A. Computing eigenvalues and eigenvectors is an elementary task for which several subroutines can be used; commonly, DMD uses Singular Value Decomposition (SVD) algorithm. Instead of directly dealing with the high-dimensional operator A, DMD applies SVD to the data matrices to obtain a reduced-order representation.

Given the snapshot matrix V_1^{N-1}, the SVD decomposes it as:

$$V_1^{N-1} = U\Sigma W^*$$

where U is an $m \times r$ orthogonal matrix containing the left singular vectors of V_1^{N-1}, is an $r \times r$ diagonal matrix with the non-negative singular values in descending order and W^* is the conjugate transpose of an $(N-1) \times r$ orthogonal matrix, the right singular vectors of V_1^{N-1}.

With this decomposition, DMD aims to find a reduced-order representation of A in the span of the columns of U, which provides a basis capturing the dominant spatial structures in the data. Specifically, the reduced-order approximation $\tilde{A}$ is computed as:

$$\tilde{A} = U^* AU$$

where U^* is the conjugate transpose of U. This $\tilde{A}$ matrix is of much smaller dimension $r \times r$ compared to the original A and captures its main dynamics.

Eigenvalues and eigenvectors of $\tilde{A}$ are then computed. The eigenvalues provide information about the temporal behavior of the modes (like frequency and growth/decay rates), while the eigenvectors, when projected back onto the original space using U, provide spatial structures associated with each mode. This entire process exploits the efficiency of SVD to extract meaningful dynamic information from the data without ever explicitly computing or storing the large matrix A.

Characteristics of the eigenvector: Not only, the eigenvalues and the eigenvectors of A completely define a set of modes each of which is associated with a fixed oscillation frequency and decay or growth rate. If the eigenvalues lie within the unit circle (for discrete systems) or have a negative real part (for continuous systems), the system is stable. The farther away they are from the unit circle, the faster the system will diverge or become unstable. These modes are a natural generalization of the global stability modes. Let's remember that stability modes in dynamical systems refer to the intrinsic patterns in the system or behaviors that determine how the system responds over time, especially in relation to perturbations or disturbances. Therefore global stability modes indicates how the *entire* system reacts and returns to equilibrium.

In DMD, we can say that there is a "hardness" regarding the complexity in transforming a nonlinear system into an approximately linear one. The number of distinct eigenvectors A needed to represent a system accurately gives an indication of this hardness. A system that requires a large number of eigenmodes to be linearized or approximated indicates that it has complex, intertwined behaviors that don't lend themselves easily to linear representation. When the system is close to a "sinusoidal silhouette" (like variable stars for example), it suggests that the dynamics can be captured using oscillatory (sinusoidal) modes. At the same time, if transforming the system requires a high number of eigenmodes, it indicates the system's behavior is far from simple oscillations, and it may involve interactions of multiple frequencies, chaotic behaviors, or other intricate patterns.

Higher order DMD: In the application of DMD in the work presented in Sec. 2.6, we deal with a low dimensional dynamical system, for this reason, we employ the extension of DMD known as *Higher Order Dynamical Mode Decomposition* (HODMD; Le Clainche and Vega, 2017). The modification introduced by HODMD is the following: instead of assuming $v_{i+1} = Av_i$, we consider an higher order Koopman assumption, i.e. that $v_{i+d} = A_1 v_i + A_2 v_{i+1} + \cdots + A_d v_{i+d-1}$ where $d \geq 1$ is a tunable parameter. It is not difficult to notice that we are simply superimposing DMD on a sliding window over the time series. The main advantage of such an approach lies in enabling the computation of *higher order modes* enriching the space returned by classical DMD when considering dimensional systems presenting low *spatial complexity* but a large number of hidden frequencies.

For example, in our experiments described in Sec. 2.6, the time series $V_1^N = \{v_1, \ldots, v_n\}$ is given by the uniformly sampled value of magnitude measured at regular intervals over the interpolated light curve described in Sec. 2.2.3 and the output of the algorithm can be interpreted as the higher order modes that encodes the generalization of global stability modes and their underling frequencies.

Regarding the implementation, we used the library `PyDMD` (Demo et al., 2018), that presents several versions of DMD already efficiently implemented in Python. We take 30 snapshots per phase (for a total of 150 snapshots) so that enough snapshot are fed to DMD to produce a model that

fits with the original data but not too much as to model certain parts of the spline that may be due to noise. For example, a snapshot density that is low enough reduces the likelihood of having a snapshot on parts of the spline where a non-representative spike or drop is caused by a few outliers. In other words, it acts as prevention of overfitting and eliminates the need of outlier deletion. By the same token, the eigenvalues found are fewer and more representative of the physics of the system.

2.4 Sparse Logistic Regression for RR Lyrae versus Binaries Classification

Based on Trevisan et al. 2023 paper.

2.4.1 Introdution

As described in the open problems in Sec. 2.1.1, accurate classification separation of RRL versus eclipsing binary stars has strong implications for our ability to avoid biases in distance determination. This has long been a priority in catalog building efforts, as in the work by Drake et al. (2014a), who used a robust version of the skewness of the magnitude distribution introduced by Kinemuchi et al. (2006) to separate the two classes in Catalina Sky Survey (CSS) data. It is thus natural to employ machine learning methods to address this task. Since the large-scale diffusion of deep learning, more and more classification tasks in astrophysics have been entrusted to various neural network architectures (see e.g. Hinners et al., 2018, Lanusse et al., 2018, Gillet et al., 2019b, Bialopetravičius et al., 2019, Trevisan et al., 2020, Chardin and Bianchini, 2021, Lalande and Trani, 2022). Unfortunately, the excellent performance of such tools is often accompanied by insufficient theoretical guarantees as well as a lack of human interpretability. Consequently, making use of deep networks for extracting high-reliability patterns to be interpreted as physical models is challenging. While physics-informed deep learning (see e.g. Raissi et al., 2019, Raissi et al., 2020, Haghighat and Juanes, 2021) holds the promise to mitigate this issue, in this work we choose to favor a statistical foundation of the classifier that ensures its reliability. Moreover, in the context of scientific research and astronomy in particular, the ability to interpret our classifier is crucial. Not only, as stated in the introduction, it has recently been convincingly argued that natively interpretable models are thus preferable to black boxes even when the latter are explained a posteriori using relevant techniques because explanations cannot possibly be faithful everywhere or they could simply be used in place of the underlying black box (Rudin, 2019).

Several works have already appeared in the literature attempting to address the problem of automatic classification of variable stars, but in general the main priorities of these were different from interpretability. In a recent investigation, Pantoja et al. (2022) computed a wide range of features (including the Feature Analysis for Time Series (FATS features) introduced by Nun et al., 2015) on light curve data and apply dimensionality reduction and clustering. They also leverage the results to perform semi-supervised learning achieving a precision of 90% for main variable classes with a support-vector machine, thus reducing the need for human assigned labels. On the other hand,Castro et al. (2018) focussed their attention on modelling errors, so they apply Gaussian process regression to light curves -as we do in the following- and use it as a basis for bootstrapping and calculate FATS features, using them to train a random forest. The goal of a variety of works

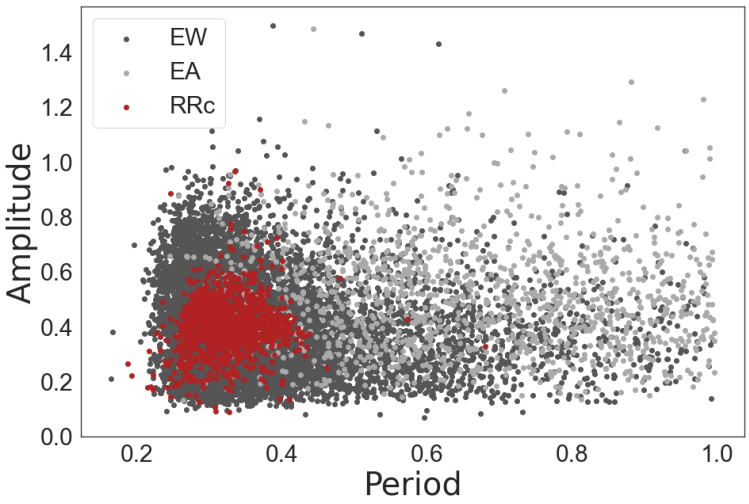


Figure 2.13: Period–amplitude plot of our CSS data set. Only the region up to 1.5 mag in amplitude and 1 day in period is shown. The stars labeled as RRc are shown in red; the eclipsing variables are shown in different shades of gray. Substantial overlap between the two groups is readily evident.

employing a range of neural network architectures is arguably speed and accuracy of classification, e.g. Debosscher et al. (2007), Aguirre et al. (2019), Becker et al. (2020), Becker et al. (2020), Zhang and Bloom (2021) and Szklenár et al. (2022). Brink et al. (2013) also generated features and trained several supervised models, including logistic regression. While the latter is certainly more interpretable than a neural network, the usual L^2 loss employed does not result in sparse coefficients and the general goal of that work was to separate real variability from bogus variability in the Palomar Transient Factory survey. Other application of DL and ML method in variables classification, can be found in Dubath et al. (2011) where a random forest is applied on hipparcos variables. Debosscher et al. (2007) perform multiclass classification using neural nets and support vector machines. These approaches often reach high classification performance but tend to be opaque at least when compared to the method we introduce in the following, regularized logistic regression . Other works, e.g. Richards et al. (2011) classified all variable stars, including RR Lyrae and eclipsing binaries, using tree-based, hierarchical and pairwise classifiers. In that case more than 32 periodic features and 20 non-periodic features were utilized. Showing that interpretability wasn't a priority. Moreover, as we can see by the effort of Drake et al. (2014b) on separating RRc stars from binaries from the CRTS in the ultra-short period (periods below the 0.22 day) shows us the difficulty of this task.

As we discussed, it is more challenging to distinguish RRc from other eclipsing binary stars because some binary stars have a more sinusoidal light curves that are similar to those for RRc stars and with similar period range (as we can see from Fig. 2.13).

2.4.2 Data processing

The original sample from CSS consisted in 12812 stars and is described in depth in Sec. 2.2.1. From that dataset we selected the RRc, EW and EA stars. As remarked before, one of the flaw of this dataset is the absence of stars in the proximity of the Galactic plane (see Fig. 2.4 to see the effect on our sample), which in principle may introduce a bias in our results. However, the CSS selection is justified due to its vast collection of relatively homogeneous light curves, which are essential for

training numerous machine learning models. Furthermore, to improve the quality of our variable star sample and remove any potential false positives, we undertook several steps: one of these steps involved utilizing the The Gaia Renormalized Unit Weight Error (RUWE), the square root of the normalized χ^2 of the astrometric fit to the along-scan observation, values obtained from Gaia (Gaia Collaboration et al., 2016). To do this, we selected a circle with a radius of 3", centered at the positions provided by the CSS, and crossmatched it with the Gaia Data Release 3 database (Gaia Collaboration et al., 2022). We then removed all stars with RUWE values greater than 1.4. This value was chosen as a threshold as RUWE values larger than 1.4 are indicative of potential eclipsing binaries, which are not single stars, and therefore not suitable for our RRc sample.

By implementing this methodology, we were able to eliminate potential false positives and increase the efficiency of our variable star sample.

The final phase will be: $\phi(T) = \frac{T - \lfloor \frac{T}{kP} \rfloor \times kP}{kP}$ where T is the Modified Julian Date (MjD), P : Period of the pulsating signal given in the dataset (nominal period found by the authors through analysis of variance (Schwarzenberg-Czerny, 1989) and k an adjustment factor to the period ($k = 1$ for the binaries $k = 2$ for the RRc). This allows us to compare the two on an equal footing because a miscalculated period may result in misclassification. In particular, an eclipsing binary folded to half its actual period will be much more likely to be classified as an intrinsic variable and vice versa We repeated the resulting phased light curve three times to enforce periodicity. Afterward we fitted our data with a GPs (see sec. 2.3.1 for insights on GPs) with the following kernel:

$$k(x_i, x_j) = \exp\left(-\frac{2\sin^2(\pi|x_i - x_j|/p)}{l^2}\right) + \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\frac{|x_i - x_j|}{\rho}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{|x_i - x_j|}{\rho}\right) + \sigma^2\delta_{ij} + c^2$$

where we fixed the length scale $l = 2$ and periodicity$p = 3$ for the periodic kernel, $\rho = 5$ and $\nu = 3/2$ for the Matern kernel and $\sigma = 5000$ for the white kernel.

It is worth noting that the kernel parameters were kept constant throughout the study, and no optimization on a single variable type was performed. This means that only one kernel function with fixed parameters was used for all stars. We used the Python library scikit-learn (Pedregosa et al., 2011) based on Rasmussen (2003) to fit our Gaussian-process model. The resulting interpolated light curve is sampled at 300 equally spaced points in phase, normalized between 0 and 1. Finally, we phase-shifted to make the maximum coincide with the beginning of the period, and then, we selected 100 consequent points. By operating in this way, we avoided any possible edge effect that could affect our final data set. A quick visual recap of this process on a random EW star is shown in Fig. 2.15. Since, we aimed to utilize the same kernel for all variables, this approach meant that the Gaussian Processes (GPs) could not perfectly fit every star. After this pre-selection, we retained 1,298 out of the original 1,500 RRc stars. For the EW type, the count reduced to 7994 from 8803. Additionally, our process resulted in a reduction from 2509 to 2446 for the EA binaries, leading to a loss of 63 stars. In table 2.2, we recap the number of stars at every step of the data-cleaning process.

For purposes of visualization, we calculated the first three principal components (F.R.S. 1901) of our data set in feature space. These are shown in Fig. 2.16, where the eclipsing binaries (including both EA and EW) and the RR Lyrae are shown in different colors. Principal components have been calculated on normalized light curves, so they depend only on the shape of the light curve and not on amplitude and period. Fig. 2.16 thus shows that the light curves of RRc and eclipsing

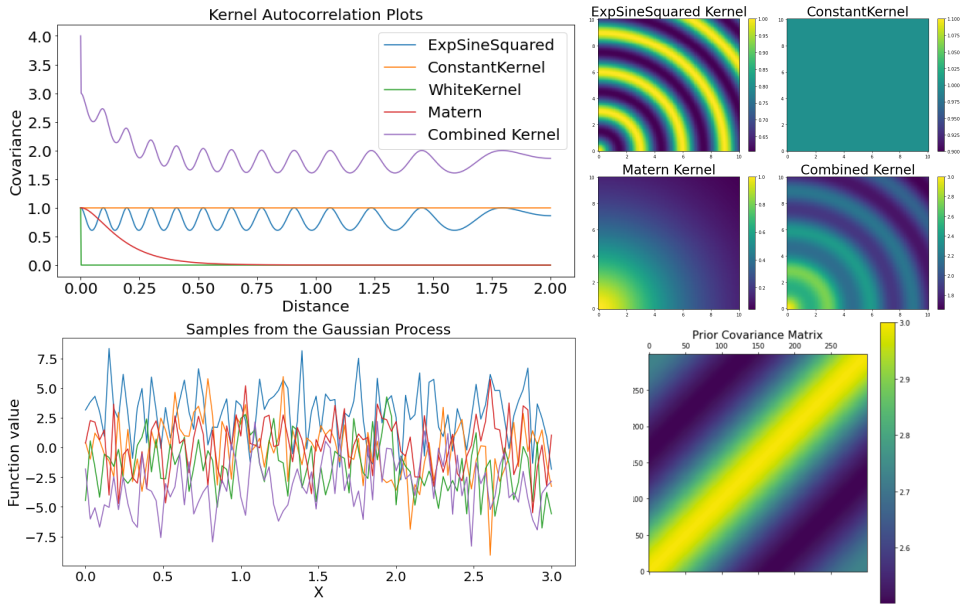


Figure 2.14: *Top-left panel:* Autocorrelation plots for the different kernels. I am plotting the distance vs the covariance for a single row of the covariance matrix. This effectively shows how the covariance between points decreases as the distance between them increases, which is a way to visualize 1D kernel functions. *Top-right panel:* we can see the spatial autocorrelation in two dimensions with the covariance values color coded. *Bottom-right panel:* the covariance matrix is computed by evaluating the kernel function on pairs of points. The range for the colormap is set to be between the 10th percentile and the 90th percentile of the kernel values, to increase the contrast for better visualization. *Bottom-left panel:* 5 samples from the $\mathcal{GP}$ with the combined kernel. Each sample is a possible realization of the $\mathcal{GP}$ over the given grid of x values.

Table 2.2: Adopted data set summary. The initial number of stars (second column) is shown broken down by variable type (first column). The third column shows the number of stars with a successful Gaussian process fit, and the fourth column the stars with a RUWE<1.4; only these stars were included in the final adopted data set.

Variable Type	N	N_{gp}	N_{RUWE}
All	12812	11809	11738
RRc	1500	1369	1298
EW	8803	7994	7994
EA	2509	2446	2446

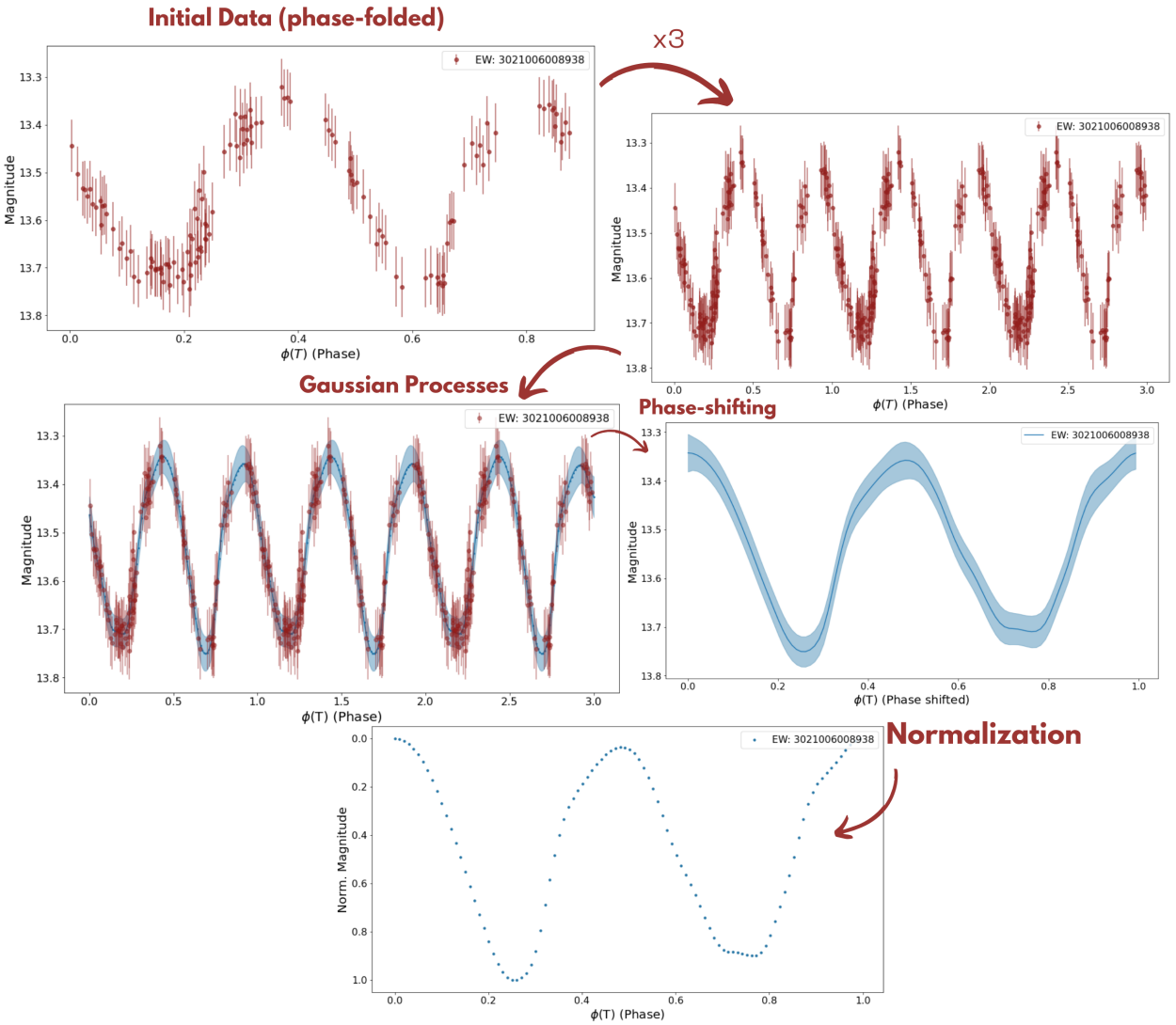


Figure 2.15: Recap of our data processing with a random EW star. Step 1: folding of the data with the nominal period. Step 2: replicate three times our data to enforce periodicity and remove contour effects. Step 3: gaussian process with the kernel described in the text. Step 4: phase-shift our data in order to have the maximum at the start and selection of only one "phase". Step 5: normalization in magnitude: now we have removed information about the amplitude and period and our classifier will use only the light curve shape. The light blue shaded area is the 95%confidence interval based on the standard deviation.

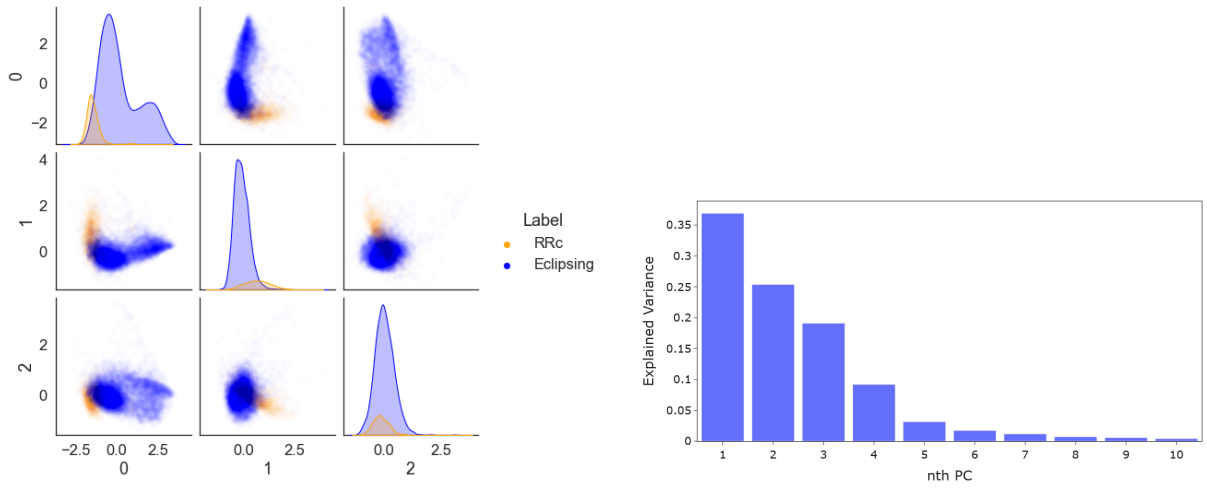


Figure 2.16: *Left Panel*: Distribution of our data in principal component space. The first three components are shown. The RRc variables are shown in orange, the eclipsing binaries in light blue. *Right Panel:* Percentage of variance explained by the first 10 principal components. The first four principal components explain 90% of the variance of the dataset.

binaries differ in shape enough that the two groups appear as clusters in principal component space. Even though these clusters are not easily separated by eye, their degree of overlap is less than that witnessed in the Bailey diagram in Fig. 2.13. In principle one could build a classifier by placing a cut along one of the principal components shown in Fig. 2.16; this would result in a separating hyperplane in feature space. These procedure is shown in table A.2 in appendix.

2.4.3 Results

We applied the logistic regression described in Sec. 2.3.2. The effect on sparsity on the validation set is shown in Fig. 2.23.

The results obtained in the experimental analysis may seem at a first glance to be unsatisfactory when compared to results of other machine learning methods: in fact, as a baseline comparison we trained two black box model: a Random Forest (Ho, 1995) and the simplest neural network in the form of a Deep Neural Multilayer Perceptron (Rosenblatt, 1958, MLP) with the default parameters of sklearn. Finally, for a baseline comparison, we add another white-box in the form of decision tree. A Random Forest is a black box due to the ensemble nature, randomness and depth of trees. Predictions are made by aggregating (e.g., averaging) the predictions of each individual tree. While a single decision tree can be visualized and interpreted, understanding the collective decision of hundreds or thousands of trees is less intuitive. If this tree are also deep an extra-layer of difficulty of interpretability is added. Moreover, as the name imply, Random Forest are random meaning that two random forests trained on the same data might have different individual trees, making interpretation less stable. MLPs have different problems in interpretability: they suffer from complexity and non-linearity. MLPs usually have millions of parameters (even our simple "mock" MLP with one deep layer has 10,202 parameters). Understanding how all these parameters interact to produce a decision is challenging, moreover MLPs utilize activation functions that introduce non-linear transformations, making it difficult to trace how the input variables are transformed into the output. Not only, while we know the mathematical operations occurring within each neuron,

Table 2.3: Different metrics for different models. The L1 Log Regression is the model that have a 90% of the coefficients equals to zero

Model	Precision (RRc)	Recall (RRc)	Specificity	AUC	F1 Score
Log Regression	0.86	0.75	0.98	0.87	0.80
L1 Logistic Regression (90% sparse)	0.86	0.66	0.99	0.82	0.75
RandomForest	0.88	0.79	0.99	0.89	0.83
MLP	0.88	0.81	0.99	0.90	0.84
Decision Tree	0.73	0.75	0.97	0.86	0.74

understanding the "why" behind each neuron's contribution to the final decision is elusive. If we remember the introduction on the criteria to have an interpretable model (monotonicity, few interactions between features, human-understandability...) we can see that both Random Forest and MLPs are perfect examples of black boxes.

The results of these different models are shown in table 2.3. Immediately, it is clear that the performances (measured by standard criteria of accuracy, precision, recall, F1 score and so on) of the black boxes are superior. However, it should be noted that there are trade-offs between the accuracy and interpretability of a model (Sarkar et al., 2016, van der Veer et al., 2021 and Bell et al., 2022). As there is no unambiguous definition for interpretability, there is no canonical quantification for this trade-off; however, important findings from the field of statistics confirm the existence of a statistical cost of interpretability Dziugaite et al. (2020). In our investigation, we strengthen the interpretability side, in fact we do not simply choose a classical statistical model such as regression, but increase its sparsity. Augmenting sparsity has an interest per se, for a survey around sparsity see Hoefler et al. (2021). Adding a sparsity-inducing penalty to the usual loss function optimized by logistic regression allows us to increase the number of coefficients that equal zero. This makes the classifier easier to understand and reduces the risk of overfitting. We chose to use a L^1 penalty, and we varied the weight assigned to the penalty term in the resulting loss function, leading to different regularization strengths. The effect on sparsity and on the performance of our classifier is shown in Fig. 2.23; this figure was used to select a regularization strength around the knee of the performance curve. Our choice resulted in a model with 90% of the coefficients equal to zero. Funnily enough, the number of false positive is lower in the penalized logistic regression even compared to the black boxes[2].

Remember: our goal was not to construct the best classifier for this problem, but rather to determine the simplest, fully transparent classifier achievable using only the light curve shape.

Misclassifications: Comparing misclassifications across multiple models can help identify which data points are consistently difficult to classify and which models perform better for certain types of data.

For each model, I identified which samples from the validation set were misclassified. Then we can analyze misclassifications shared between the models. This can give insights into potential "hard" samples. An important remark: building a model it is an iterative process (data processing, data exploration, training and validate different models, maybe hyperparameter optimization). This part of the ML process often is usually never stated in ML works. The core idea of validation itself,

[2] *Note: The hyperparameters have not been fine-tuned.*

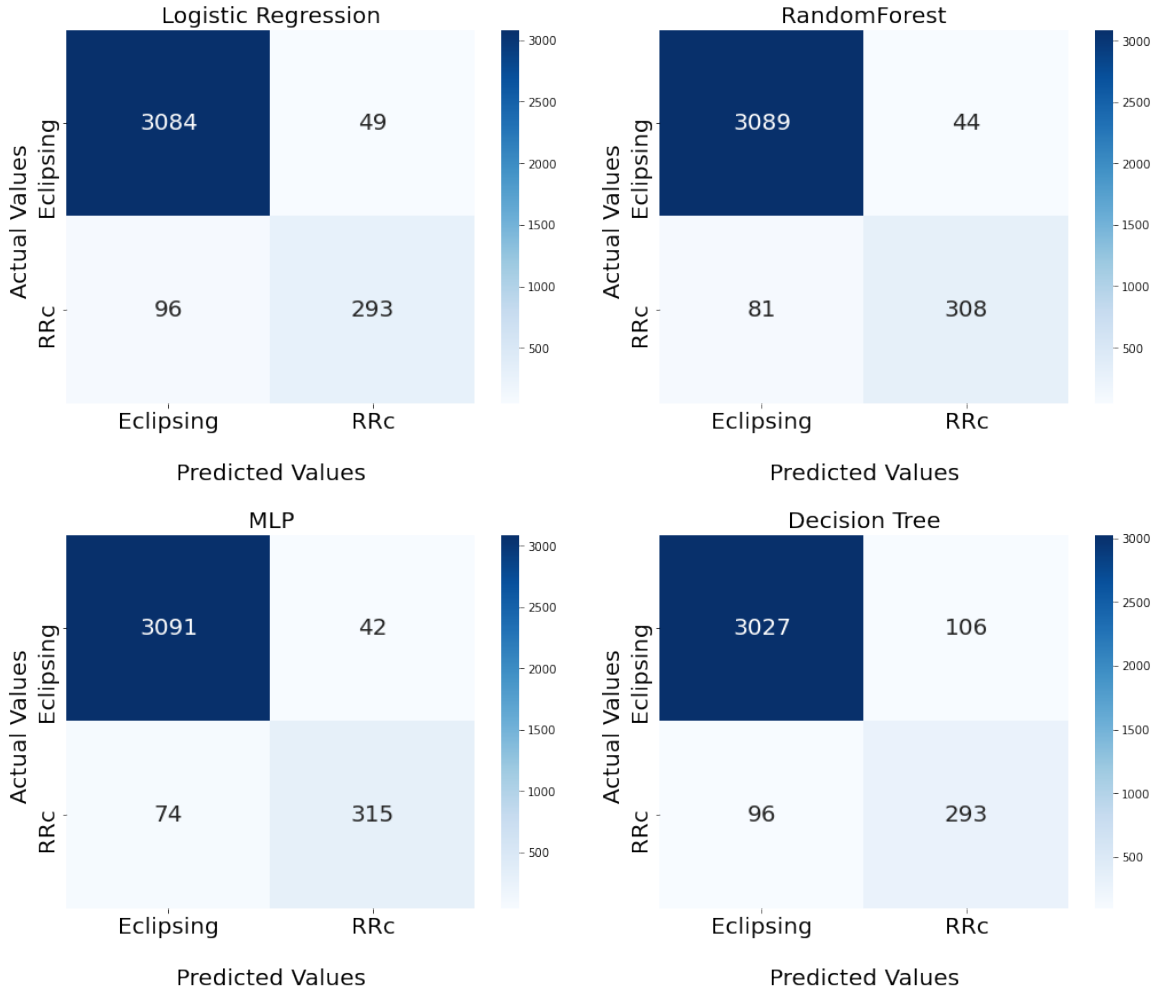


Figure 2.17: Confusion matrix for the different models. The results are referring to the validation dataset

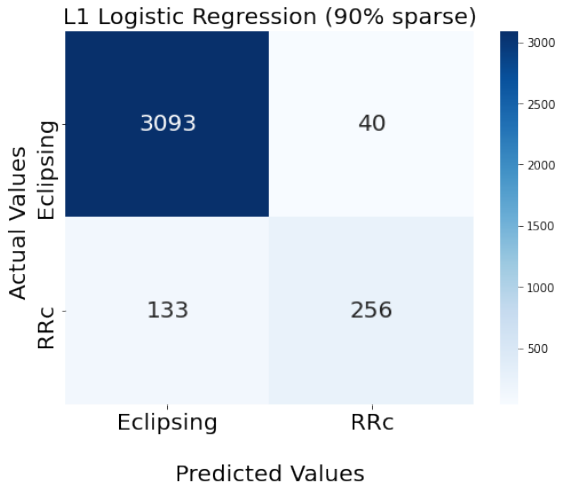


Figure 2.18: Confusion matrix for the penalized logistic regression. The results are referring to the validation dataset

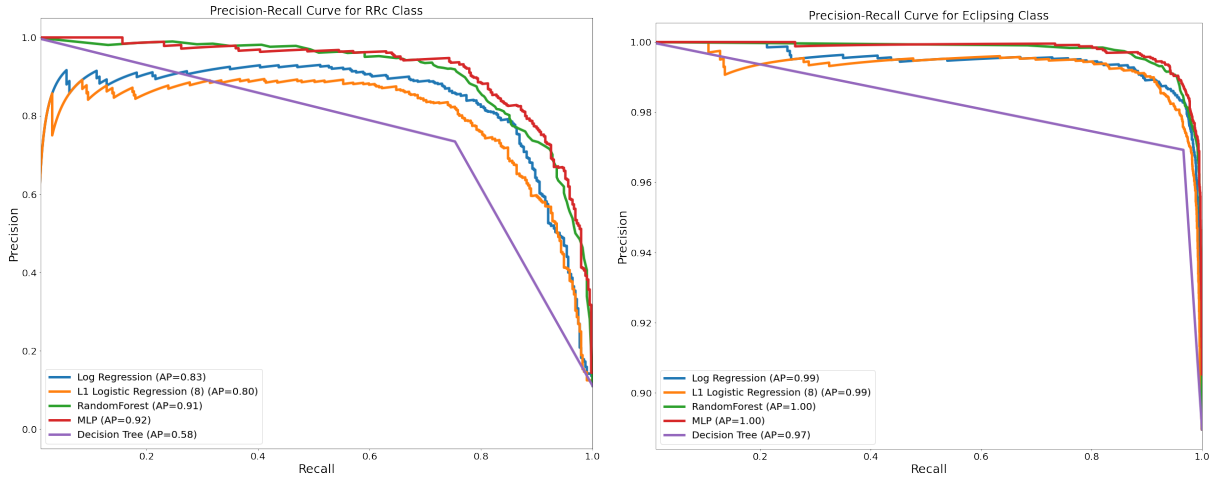


Figure 2.19: Precision-recall curve colorcoded for the different models for both the RRc and EB classes respectively on the left and right panel. When working with unbalanced dataset a precision-recall curve is preferred to a ROC curve.

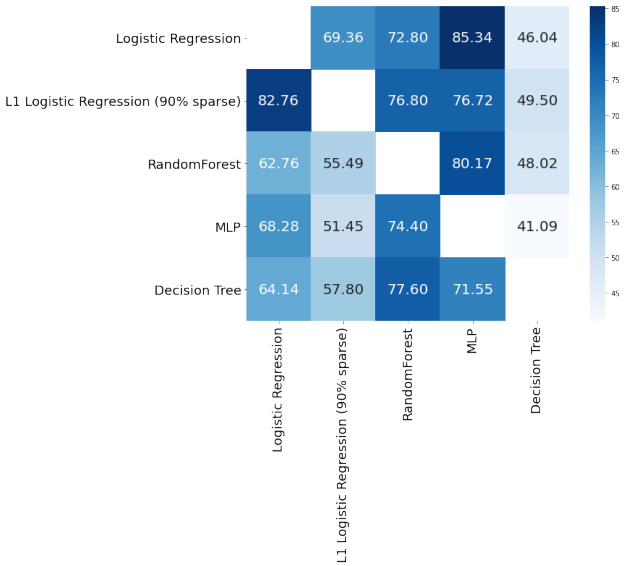


Figure 2.20: Percentage of shared misclassifications between the models.

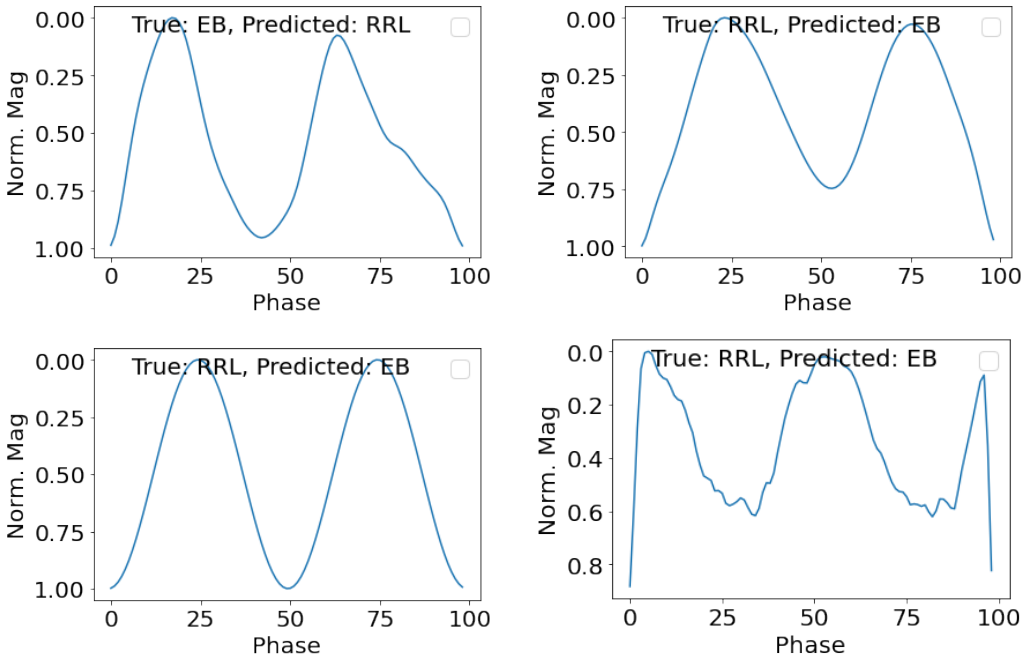


Figure 2.21: Example of misclassified light curve by all the models. Everytime, I went back to fix the gaussian processes or the parameters of models in order to address these misclassification, I had obviously a better classifier but at the same time I was slowly overfitting my validation set.

helps identify if the model is overfitting or underfitting to the training data. Not only, looking at misclassification during validation offers insights into potential adjustments to the model architecture or training process and provides a way to understand patterns or specific instances the model struggles with before final evaluation. Also, it can reveal potential wrong labels and "dirty" data that leaked during the data-cleaning process. At the same time, due to the iterative nature of model development, examining misclassifications during validation itself can lead to overfitting to the validation set itself. Also hyperparameter optimization can have the same effect.

Interpretation: The qualitative interpretation of the coefficients is as follows: increasing the value of the corresponding feature (moving one of the points in the light curve downwards) increases the probability of being classified as a RRc if the coefficient is positive, decreases it otherwise. Most coefficients are nonzero in the nonpenalized model, making interpretation harder. The penalized model drastically reduces the number of nonzero coefficients, showing which parts of the curve are actually important for the classifier. These coincide roughly with the approach toward the minima of the light curve. This is showing that the classifier relies on the quicker rise (or fall) of the RRc light curve. So through our sparsification approach, we are able to identify how the classifier is using the light-curve shape to cast its prediction. The features with nonzero coefficients correspond to the phase approaching the minima, suggesting that the steepness of the fall is used by the classifier. In terms of physical explanation, noncontact eclipsing binaries are in fact characterized by much flatter maxima (both the components are visible for an extended period of time) broken by steep minima (when one of the two components is occluded). The work was not focused on decision trees but we can see from Fig.2.24 that also the decision tree find feature 37 as the important one.

In order to see how well our model predict on an unseen data from a different catalogue, we

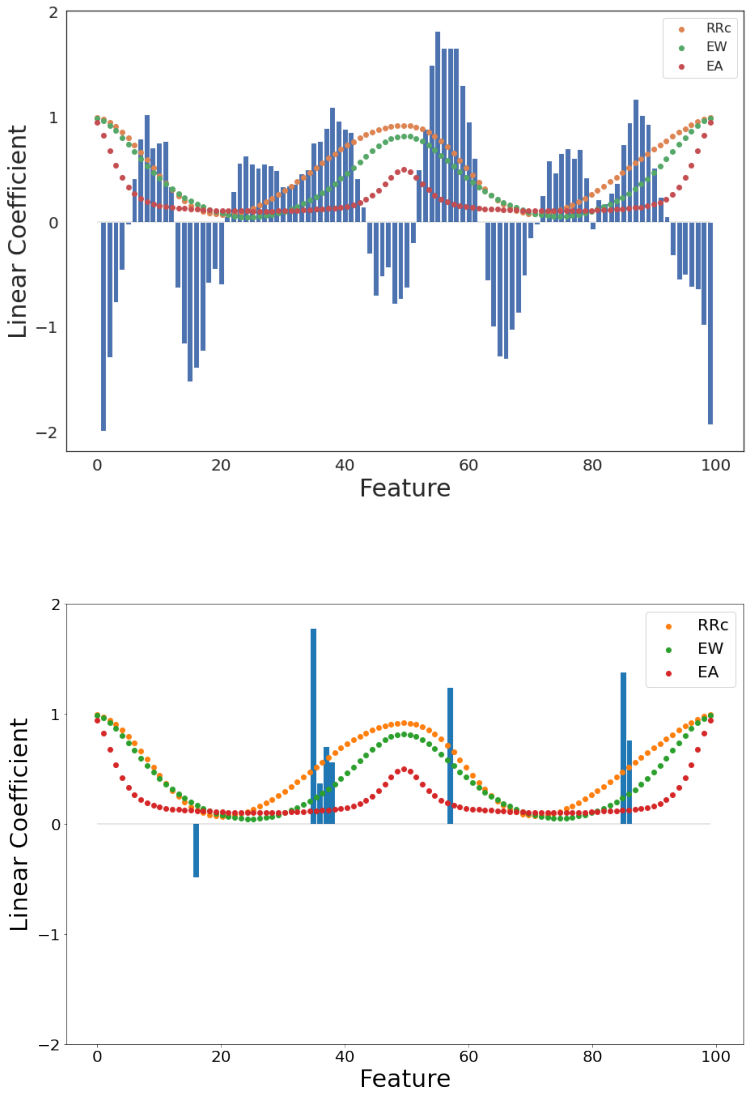


Figure 2.22: Top: values of the linear coefficients of the logistic regression for the non-penalized model (blue bars). A positive coefficient means that an increased value of the magnitude at the corresponding phase would increase the confidence of classification as RRc. Meanwhile, a negative coefficient increase the confidence of classification as EB. Mean light curve of all RRc in the whole data set (orange), mean light curve of all EW (green) and EA (red) are shown as dots. Bottom: same for the L1 regularized model: the regularization strength was chosen so that 92% of the coefficients became zero.

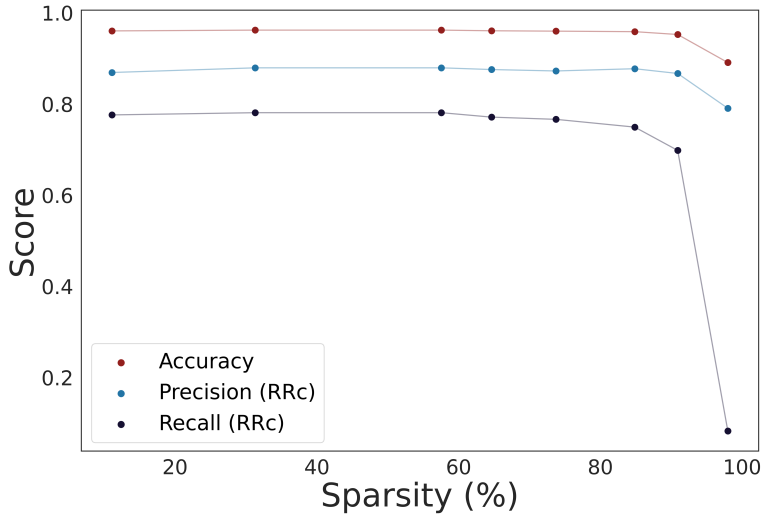


Figure 2.23: Performance metrics of our classifier (y axis) as a function of the fraction of coefficients that equal exactly zero (sparsity, x-axis). At over 90% we reach a knee where the recall for the RRc class drops precipitously but for lower sparsity the three metrics are approximately constant.

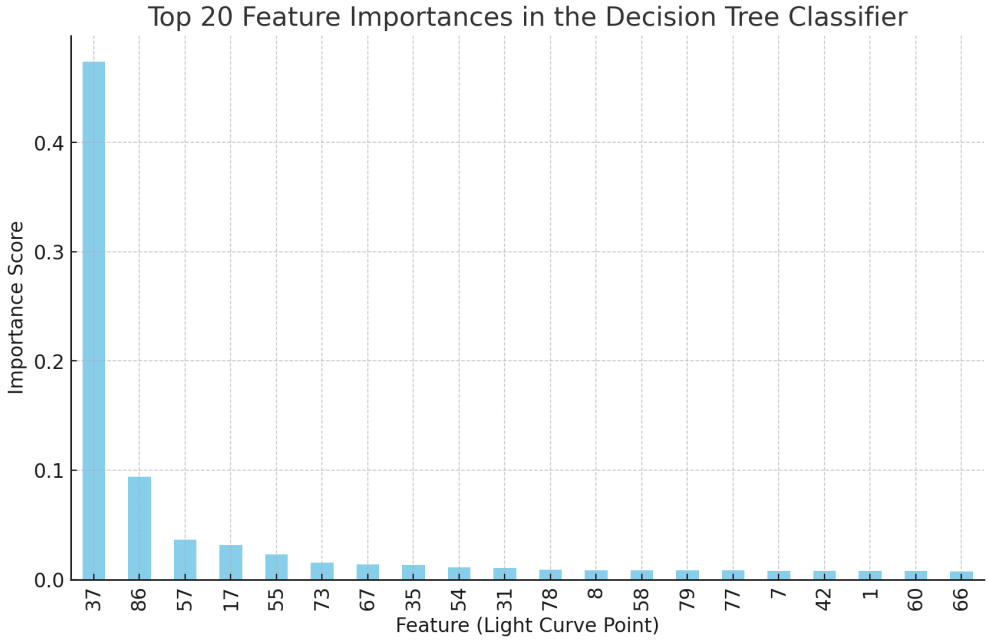


Figure 2.24: Feature importance scores for the different features in the decision tree model: feature importance scores are calculated based on the amount that each attribute split point improves the purity of the node. The idea is to measure the effectiveness of each feature in separating the classes.

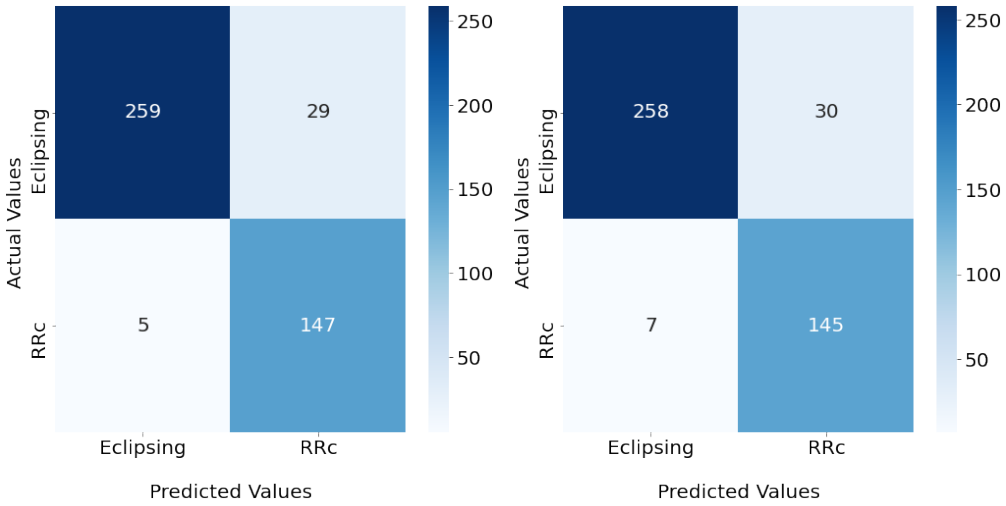


Figure 2.25: Left: confusion matrix of our non-penalized classifier on the test dataset; right: confusion matrix of our L1 regularized classifier on the test dataset.

Table 2.4: Performance metrics broken down by class (first column) of our non-penalized (second column) and L1 regularized model (third column) on the test set.

Metrics	Score	Score_{92}
Accuracy	0.92	0.92
Recall (RRc)	0.96	0.95
Recall (EB)	0.91	0.97
Precision (RRc)	0.85	0.83
Precision (EB)	0.98	0.90
F-score (RRc)	0.90	0.89
F-score (EB)	0.94	0.93

tested both the logistic classifiers on ASAS/ASAS-SN stars. The relevant metrics are summarized in Tab. 2.4 and the confusion matrices are shown in Fig. 2.25.

2.4.4 Conclusion

Distance determination relying on RR Lyrae stars as standard candles is negatively affected by contamination due to eclipsing binaries. In this work, we have shown that a simple logistic regression classifier trained on CSS light curves can separate RRc variables from eclipsing binaries reaching 0.87 precision at 0.78 recall on a validation sample of unseen CSS light curves. On ASAS/ASAS-SN light curves recall becomes 0.96 and precision becomes 0.85, showing very good generalization even to data taken by a different instrument. It has to be noted that the fraction of RRcs in the ASAS/ASAS-SN sample is different from the CSS sample (12% VS 35%). Thus good generalization is due to our deliberate choice of a simple model which is unlikely to overfit the data. Since our features are normalized magnitudes at different phases, our classifier is making use only of the shape of the light curve, without receiving any information about its amplitude, absolute mean magnitude, or period. We are also able to visualize the logistic regression coefficients corresponding to the normalized magnitude at each phase point. Through our sparsification approach, we are able to

identify how the classifier is using the light curve shape to cast its prediction. The features with non-zero coefficients correspond to the phase approaching the minima, suggesting that the steepness of the fall is used by the classifier. In terms of physical explanation, non-contact eclipsing binaries are in fact characterized by much flatter maxima (both the components are visible for an extended period of time) broken by steep minima (when one of the two components is occluded). We provide a precision-recall curve for the full classifier as well as for the penalized one, which in principle allows for a selection of the classification threshold based on the relative cost of false positives (eclipsing binaries classified as RR Lyrae) and false negatives (vice versa). We conclude with a few caveats and perspective improvements for the future. Our classifiers rely on turning the observed light curve, which is sampled at irregular intervals, into an equally spaced time series by means of interpolation. We carried out this interpolation based on Gaussian processes, which depend on several parameters to fully specify the covariance structure in terms of a kernel function. In the work described here we fixed those parameters once and for all, even though this resulted in a handful of bad fits which were later discarded. There is definitive room for improvement in this preprocessing of the data, especially if it is to be adapted to different data sets. In this context it is worth mentioning that the current Zwicky Transient Facility and near-future Vera C. Rubin Observatory (Masci et al. 2018; Ivezić 2019) long-term variability surveys will provide well-sampled, long-baseline light curves in several optical- and near-infrared bands (see, e.g., LSST Science Collaboration et al. 2017), thus providing the unique opportunity to use physically rooted diagnostics (amplitude in different photometric bands) to properly separate intrinsic from geometrical variables.

2.5 Sparse identification of variable-star dynamics

Based on Pasquato et al. 2022 paper. In this work, we applied the SINDy algorithm (see section 2.3.3) to RRab, RRc, DeltaScuti stars (described in section 2.1.1) from the Catalina Sky Survey (details of the catalogue in section 2.2.1) to automatically learn governing equations from observed light curves. SINDy was applied already to different numerous field (e.g. Arzani and Dawson, 2020, Guan et al., 2020, Horrocks and Bauch, 2020) and more recently by Course and Nair (2023) but never on variable stars data. Unlike three-dimensional hydrodynamic simulations, concise systems of governing equations are easily interpretable, similarly to those derived by a simple physical model. Unlike the latter, though, SINDY is entirely data driven, except for a few user choices (sparse regression hyper-parameters and the basis of functions against which to carry out said regression). We report several cases of striking agreement between a light curve and the solution to the governing equations learned from it. In general, the success rate of SINDY depends on variable star type, being much higher on RRc and δ Scuti variables than on RRab variables. This is likely due to the fact that RRab stars have unique and asymmetric light curves, while RRc stars are known for their symmetric, nearly sinusoidal light curves. The intrinsically more complex dynamics of RRab stars could likely be described with the use of higher order equations than those we adopted here. The learned governing equations have a common form, which reduces to a single second order differential equation (ODE) known as a generalized Lienard equation (see e.g. Moehlis et al., 2006, Abdullah, 2017). A typical, well studied Lienard equation is the van der Pol oscillator (Van der Pol, 1934). Equations of this class have previously been applied to stellar pulsation (Tanaka et al., 1990, Addo-Asah et al., 1995) and to the Solar cycle (Nagy and Petrovay, 2013).

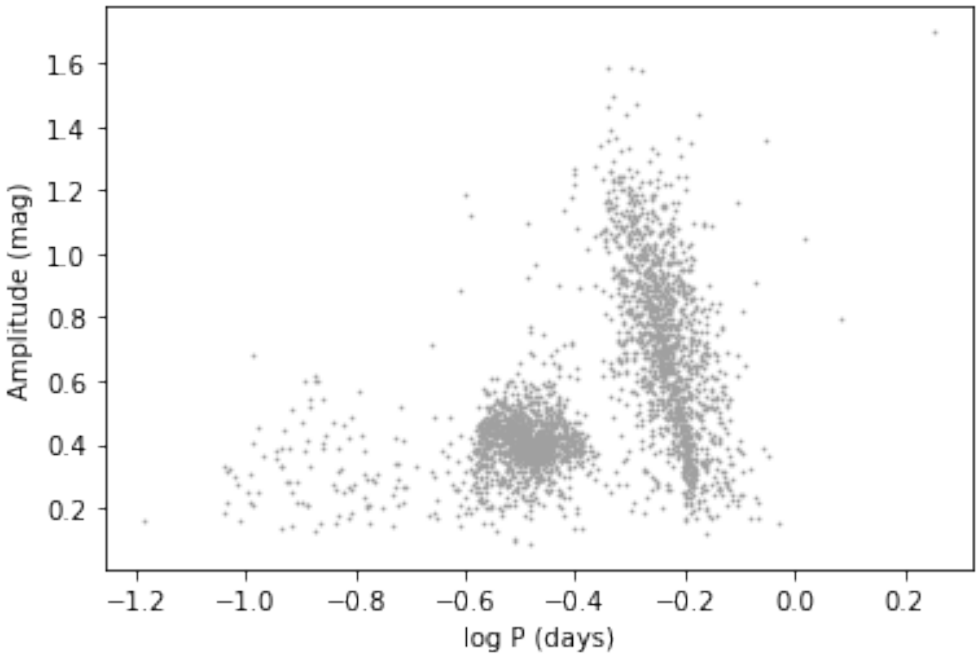


Figure 2.26: Adopted sample for this study in the period-amplitude plane.

2.5.1 Data

In the following, we adopted periods from the CSS catalogs descried in Sec. 2.2.1 and phased light curves shifted to mean zero magnitude. The light curves were fitted with a smoothing spline[3]. This approach results in a smooth curve and tolerates a few outliers in the raw data, as shown e.g. in Fig. 2.27, top left panel. Furthermore, the time derivative was easily computed by the spline fit. The phased light curves represent the stars' light intensity variation folded over specific periods. Periodicity has been enforced by adding copies of the light curve at the beginning and at the end of the interval used to fit the spline. The resulting smooth curve and its derivative were then sampled regularly at 200 points per period. The resulting time series, identically repeated 10 times, were provided as input to SINDY, treating the light curve and its derivative as two separate variables.

This procedure was carried out on a subsample of light curves containing 100 δ Scuti variables, 1500 RRab variables, and 1500 RRc variables. The Bailey diagram for full sample is shown in Fig. 2.26. Six light curves from this sample are shown in Fig. 2.27 with our smoothing spline superimposed. The relevant parameters for each star in Fig. 2.27 are reported in Tab. 2.5.

In our case, the state variables are the magnitude of a given star and its time derivative. In the following, we denote these quantities by x and y, respectively. After some experimentation, for simplicity and ease of interpretation we chose a polynomial basis truncated to the second power, including interaction terms. Thus the most general system of governing equations that SINDy can learn is in the form

$$\begin{aligned} \dot{x} &= a_{00} + a_{01}x + a_{02}y + a_{03}x^2 + a_{04}xy + a_{05}y^2 \\ \dot{y} &= a_{10} + a_{11}x + a_{12}y + a_{13}x^2 + a_{14}xy + a_{15}y^2. \end{aligned} \tag{2.2}$$

[3]Obtained by applying the smooth.spline function from the stats package in the R programming language (Bates et al., 1992, R Core Team, 2019) with all the points being treated as knots.

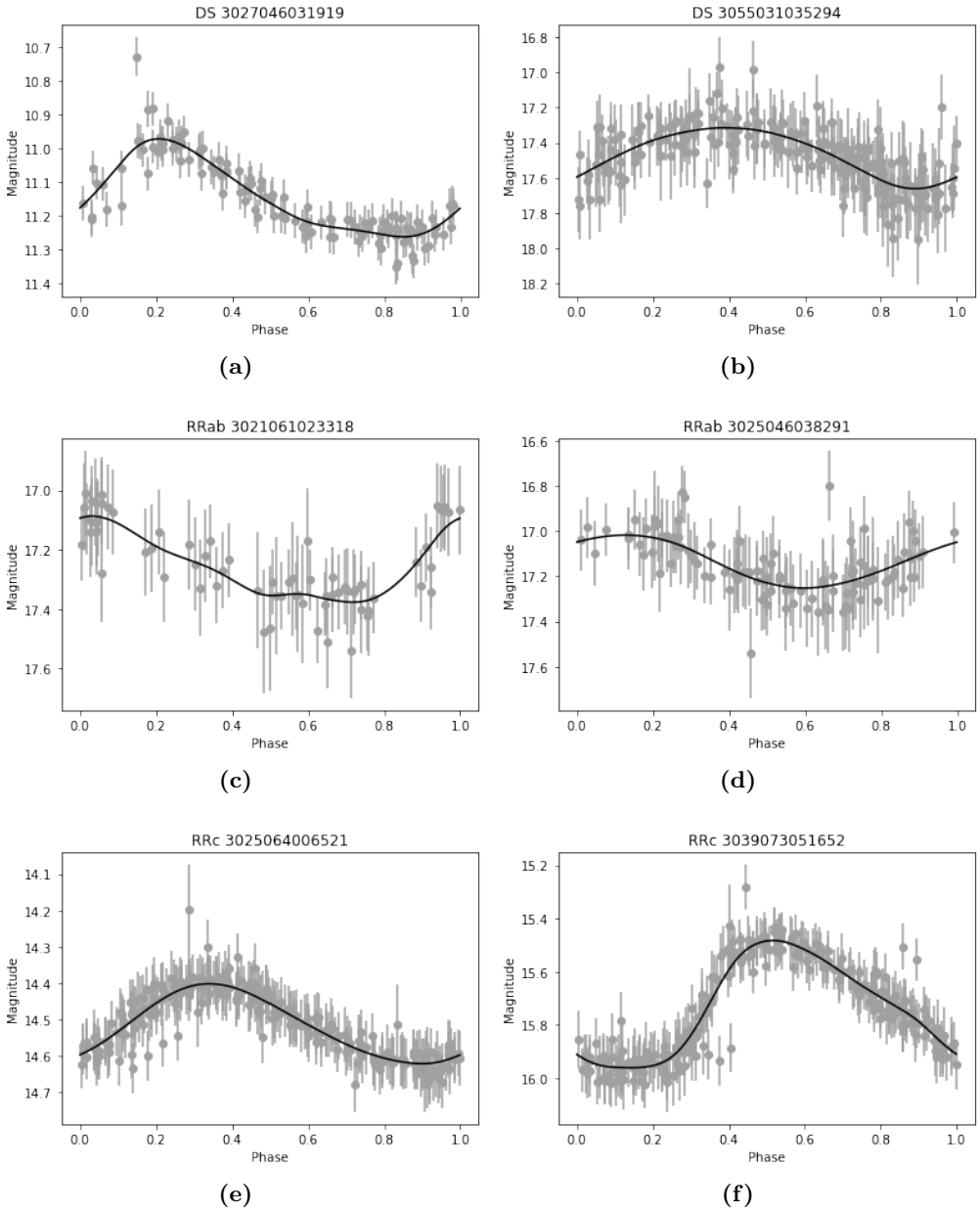


Figure 2.27: Raw light curves for six stars from our original catalog (gray points with error bars) with our best fit spline superimposed (black solid curve).

Numerical ID	RA	Dec	Period	V	Npts	Amp
3027046031919	102.84	−27.48	0.17	11.14	116	0.29
3055031035294	106.15	−55.24	0.19	17.46	222	0.34
3021061023318	129.66	−21.69	0.63	17.24	69	0.33
3025046038291	99.42	−25.28	0.58	17.12	107	0.25
3025064006521	140.68	−26.04	0.27	14.52	251	0.22
3039073051652	186.27	−38.83	0.29	15.74	253	0.48

Table 2.5: Relevant parameters for the stars whose light curves are shown in Fig. 2.27. We report the numerical ID (col. 1), right ascension (col. 2) and declination (col. 3), period in days (col. 4), V-band magnitude (col. 5), number of measurements (col. 6), and amplitude (col. 7).

We focused on autonomous governing equations, so time does not appear on the right hand side of Eqs. 2.2. Sparse regression was implemented for each light curve using the *Lasso* function from `scikit-learn` (see section 2.3.3), with regularization parameter $\alpha = 0.01$. Increasing α forces more terms in Eq. 2.2 to be exactly zero, resulting in more parsimonious equations, but at the extreme of too high α the resulting dynamics may be too simple, predicting e.g. a constant magnitude. The adopted value of α was chosen by interactive experimentation on a handful of curves, with no access to the full data-set later used for the main analysis. The learned systems were integrated numerically using the *simulate* method of `PySINDy` for ten periods. We computed the mean squared error (MSE) between the solution to the governing equations and the actual light curve during the last of these ten periods. This results in a more conservative MSE, where solutions that initially appear to follow the data but do not do so in the long term (such as a periodic solution whose period is slightly off with respect to the original), are penalized. In the following we are using the normalized MSE, that is the ratio between the calculated MSE and the MSE of a constant, taken to be the median magnitude of the light curve.

In Fig. 2.28, Fig. 2.29, and Fig. 2.30 we show a small selection of our results, obtained by applying SINDy to the light curve of four RRab variables, eight RRc variables and six δ Scuti variables respectively. In each plot we compare the interpolated observational light curve (thick black dotted line) with the solution to the differential equations learned by SINDy (solid red curve). Note that this is not a fit to the curve, but the result of solving the system of differential equations in Eq. 2.2 with the coefficients learned by SINDy. Solutions start from initial conditions corresponding to the first point (a couple of values representing magnitude and its derivative) on each light curve and are calculated up to ten periods of the observational light curve. In each plot we show only the last such period.

The stars presented in Fig. 2.28, Fig. 2.29, and Fig. 2.30 have been selected randomly among those for which normalized MSE is under 0.16, which we define as an acceptable level of matching between the original curve and the solution to the learned equation, as can be visually ascertained. In general, in most cases in which such a result is obtained, the coefficients a_{00}, a_{01}, a_{03}, a_{04} and a_{05} in Eq. 2.2 are all exactly zero, and a_{02} is approximately one. This is expected, because we are feeding SINDy the first derivative of magnitude together with magnitude. While we treat the two variables as independent, the algorithm rediscovers their relationship. Typically, when a good solution is found, also a_{13} is found to equal zero (for 187 out of 197 stars for which normalized MSE is under 0.16). The resulting equation can thus be rewritten as

$$\begin{aligned} \dot{x} &= Ay \\ \dot{y} &= B - Cx + Dy + Exy + Fy^2 \end{aligned} \tag{2.3}$$

which reduces to

$$\ddot{u} = -u + (a + bu + c\dot{u})\dot{u} \tag{2.4}$$

with the substitutions

$$\begin{aligned} u &= (Cx - B)A \\ \Omega &= \sqrt{AC} \\ t' &= \Omega t \\ a &= \frac{1}{\Omega}\left[D + (\frac{EB}{C})\right] \\ b &= \frac{E}{\Omega^3} \\ c &= \frac{F}{A\Omega^2} \end{aligned} \tag{2.5}$$

where the dots now indicate derivation with respect to t'. However, we will drop the prime in the following.

In the following we will indicate the family of equations 2.4 with the symbol ٥ (Eastern Arabic numeral for five, pronounced *khamsa*) due to the resemblance it bears to the phase curves of its periodic solutions. We will call the case where all coefficients a, b, and c differ from 0 the general ٥ form or g-٥, and the particular case where $a = b = 0$ the simplified 0 form or s-٥. For Eq. 2.3 to reduce to the s-٥ form it is necessary and sufficient that $D = E = 0$. If also $F = 0$ then the equation reduces to a harmonic oscillator of frequency Ω.

Eq. 2.4 immediately reveals some interesting properties: neither g-٥ nor s-٥ are conservative, as can be seen by multiplying both sides by $\dot{u}$, obtaining

$$\frac{1}{2}\frac{d}{dt}\left[\dot{u}^2 + u^2\right] = (a + bu + c\dot{u})\dot{u}^2 \tag{2.6}$$

with the right-hand side in general being nonzero, since otherwise the phase trajectory would be confined to the line $a + bu + c\dot{u} = 0$ or to the line $\dot{u} = 0$. Consequently, if Eq. 2.4 is capable of producing oscillatory behaviour at all, it must be a non-conservative oscillator, with the term $a + bu + c\dot{u}$ changing sign at least twice during a period.

The phase plane for s-٥ is shown in Fig. 2.32a and for g-٥ in Fig. 2.32b, for choices of the coefficients in the range appropriate for reproducing observational light curves. In the following we show that s-٥ can be fully solved analytically and find a condition for g-٥ to have a stable limit cycle [4].

2.5.2 Analytical solutions for s-0

The coefficient c, the only one to be non-zero in the s-٥ form, sets the strength of the $\dot{u}^2$ perturbing term in Eq. 2.4. While this term is quadratic in $\dot{u}^2$, it is not equivalent to a damping with quadratic viscosity because it does not change sign with $\dot{u}$. The fact that $\dot{u}^2 \geq 0$, endows the s-٥ form with a symmetry whereby mapping $u \to -u$ is equivalent to changing the sign of c. Therefore in the following we will restrict our analysis to $c > 0$, keeping in mind that the case $c < 0$ will yield mirrored phase trajectories in the $u, \dot{u}$ plane.

Another interesting property derives from the fact that the right-hand side of s-٥ depends only on the square of the magnitude derivative, $\dot{u}$. Therefore, the phase trajectory is symmetric with respect to reflection about the u axis, a property that is not shared by the g-٥ form.

We solve s-0 by introducing the new variable $v = \dot{u}^2$ and observing that $\dot{v} = 2\dot{u}\ddot{u}$ hence by the

[4] A limit cycle is a concept from the field of nonlinear dynamics and refers to a closed trajectory in phase space that an oscillating system will converge to regardless of its initial conditions.

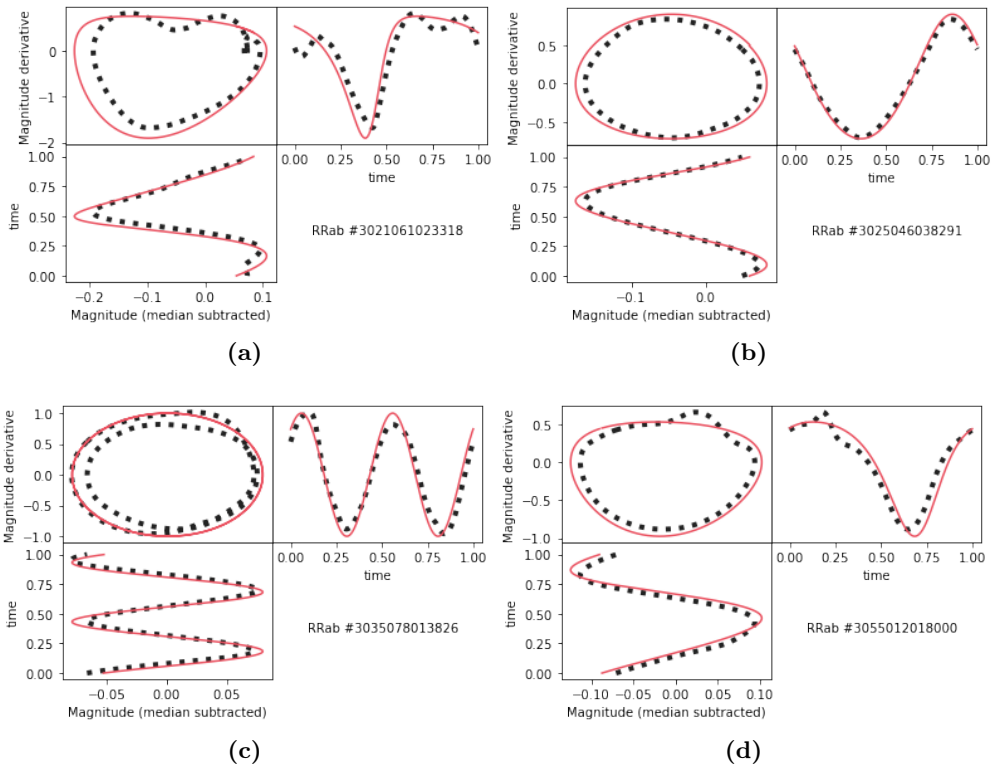


Figure 2.28: Four light curves for stars classified as RRab (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panel show the magnitude derivative versus time and time versus the magnitude, respectively. The name and type of the star are shown in the bottom right panel.

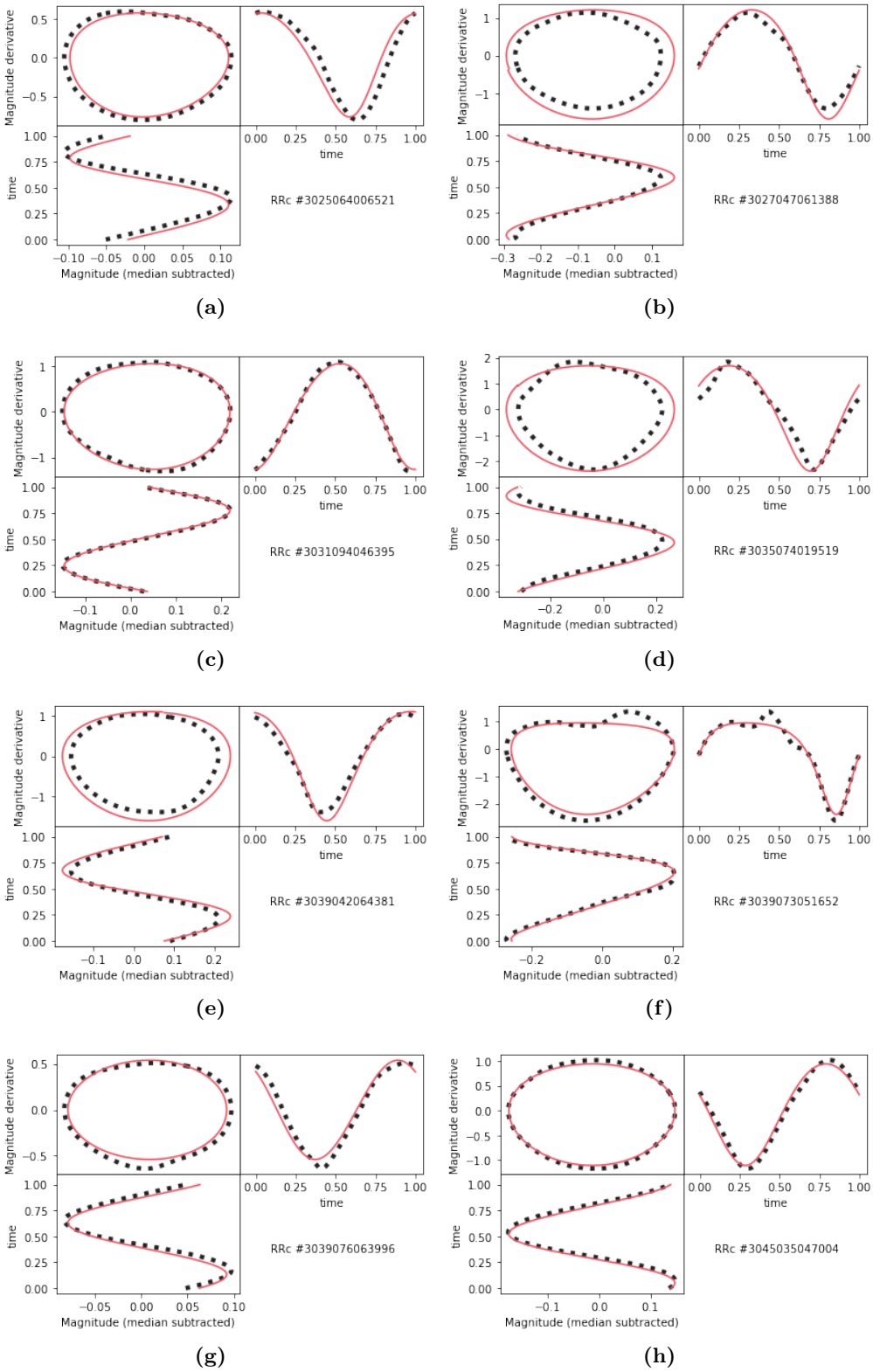


Figure 2.29: Eight light curves for stars classified as RRc (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panel show the magnitude derivative versus time and time versus the magnitude, respectively. The name and type of the star are shown in the bottom right panel.

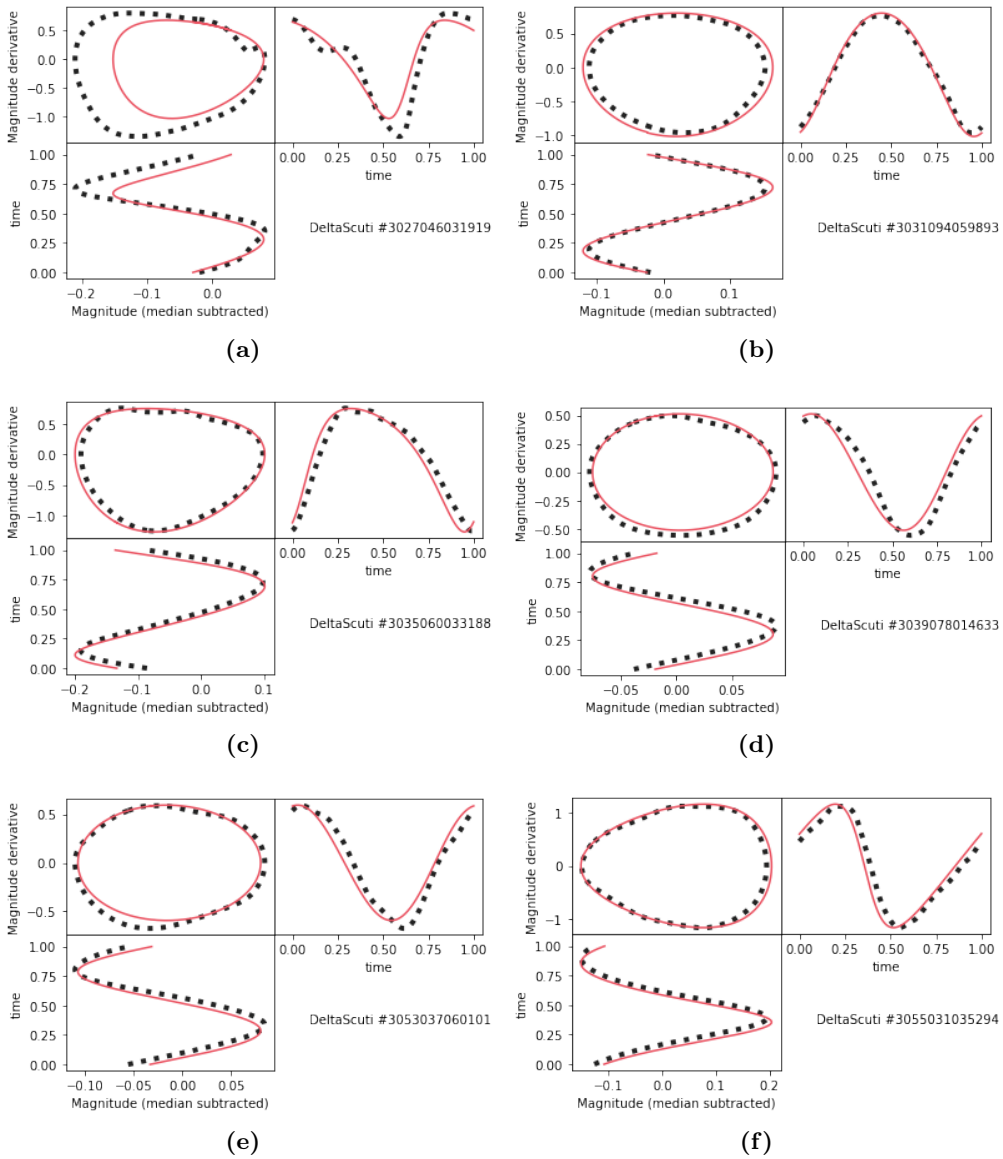
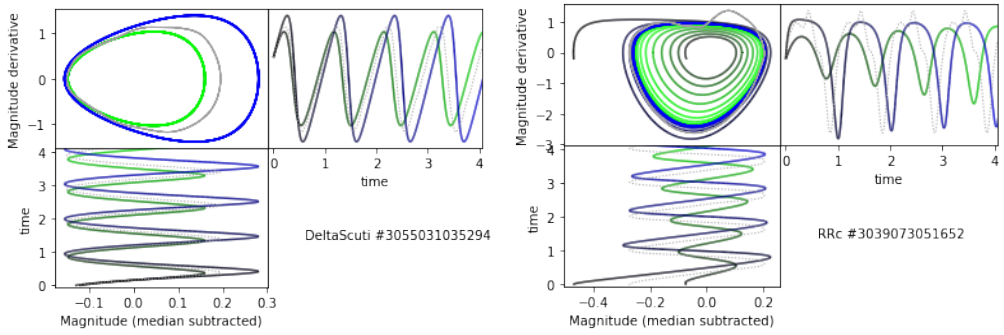
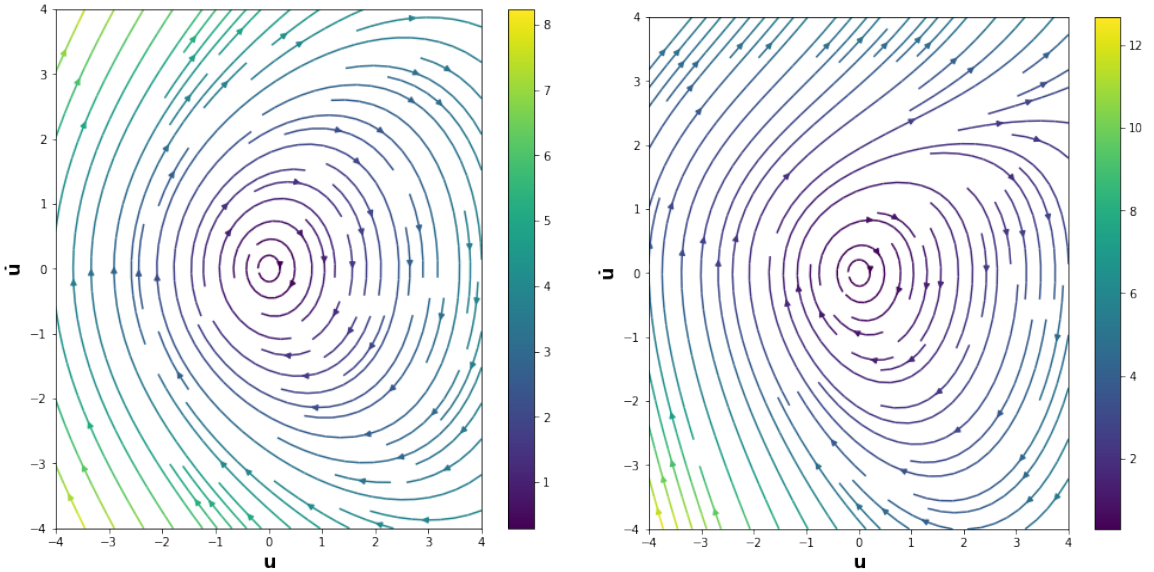


Figure 2.30: Six light curves for stars classified as δ Scuti (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panel show the magnitude derivative versus time and time versus the magnitude, respectively. The name and type of the star are shown in the bottom right panel.



(a) **(b)**

Figure 2.31: Effects of perturbing the initial conditions of the equation learned by SINDy in two different cases. The left panel shows star 3055031035294, which is a δ Scuti variable, whose learned equation is in the simplified ৩ form, with $a = 0$ and $b = 0$, the right panel star 3039073051652, which is an RRc variable described by an equation in the full ৩ form, with $a \neq 0$ and $b \neq 0$. The dotted gray line is the solution with unperturbed initial conditions, and the solid lines fading from gray to green (blue) correspond respectively to perturbed initial conditions where the magnitude has been increased (decreased). In the left panel these correspond to different, but still closed, phase curves. In the right panel they are not closed, but they evolve slowly towards the limit cycle corresponding to the unperturbed solution. Unlike in the previous figures, here the top right and the bottom left panels show three periods rather than one.



(a) **(b)**

Figure 2.32: Phase planes for s-৩ with $c = 0.2$ (left) and for g-৩ with $a = -0.01$, $b = c = 0.25$ (right). Color encodes the modulus of the vector tangent to the phase trajectory in each point.

chain rule $\ddot{u} = dv/2du$. So s-ɔ becomes

$$\dot{u} = \pm\sqrt{v} = \pm\sqrt{\frac{1}{2c^2} + ke^{2cu} + \frac{u}{c}}, \tag{2.7}$$

where k is a constant of integration. We thus reduced the problem to quadratures. Phase trajectories are symmetric with respect to the u axis, as argued above. Eq. 2.7 yields two kinds of phase curves: ones exponentially escaping to infinity ('supernova solution') with $k \geq 0$ and closed periodic ones with $k < 0$.

The value of k is set by the initial conditions u_0 and $\dot{u}_0$. If these obey the constraint $c\dot{u}_0^2 - \frac{1}{2k} < u_0$, then $k < 0$ and periodic evolution follows. The coefficients found by SINDy always place the initial conditions well within this range. This is demonstrated in the left panel of Fig. 2.31, where we alter the initial conditions slightly for a star whose learned governing equation is in the s-ɔ form, showing that the resulting solutions are still periodic.

We could not find closed-form solutions for the g-ɔ form, as opposed to the s-ɔ case. The right panel of Fig. 2.31 shows that, in general, solutions are not periodic. Numerical integration from slightly perturbed initial conditions reveals, however, that the solution found by `PySINDy` (red line in the figure) is close to what appears to be a limit cycle, and perturbed solutions (green and blue lines in the figure) apparently converge towards it. This is by no means a rigorous proof that the trajectories are located within the basin of attraction of a limit cycle, nor that the limit cycle itself exists; still, it suggests that this may be the case. A more rigorous mathematical discussion follows.

2.5.3 Limit cycle behavior of the g-0 form

The g-ɔ form has a single critical point at the origin $(u, \dot{u}) = (0, 0)$ with linearization in the local neighborhood producing the Jacobian J with corresponding stability eigenvalues $\lambda_{1,2}$

$$J = \begin{bmatrix} 0 & 1 \\ -1 & a \end{bmatrix}, \qquad \lambda_{1,2} = \frac{a}{2} \pm \frac{\mathrm{i}\sqrt{4 - a^2}}{2}. \tag{2.8}$$

The results from Eq. 2.8 suggest that the phase space dynamics locally around the origin resemble a stable spiral for $a \in (-2, 0)$ and an unstable spiral for $a \in (0, 2)$. Indeed, we note that there is a bifurcation at $a = 0$, whereby the origin transitions from attracting locally to repelling. To establish the existence of limit cycles near the origin, we begin by showing that g-ɔ can be written in the Cartesian Hopf normal form

$$\begin{aligned} \dot{y} &= \alpha y - \omega v + f(y, v), \\ \dot{v} &= \omega y + \alpha v + g(y, v), \end{aligned} \tag{2.9}$$

where $\lambda_{1,2} = \alpha \pm i\omega$ and $f(y, v)$ and $g(y, v)$ are arbitrary functions to be determined by transformation from the original g-ɔ equation.

From the linearization eigenvalues, we construct a transformation matrix

$$\Phi = \begin{bmatrix} \frac{a}{2} & \frac{\sqrt{4-a^2}}{2} \\ 1 & 0 \end{bmatrix}, \tag{2.10}$$

so that the matrix product $\Phi^{-1} J \Phi$ produces the correct coefficients in Eq. **??** to first order in (y, v),

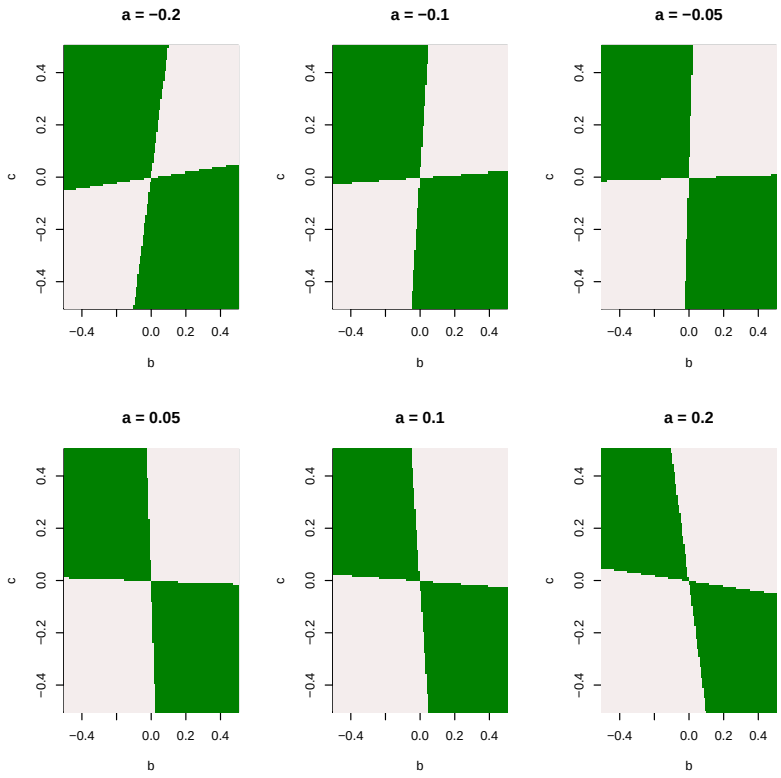


Figure 2.33: Relation between the a, b, and c coefficients for which $\hat{A} > 0$ (green shaded area), resulting in the existence of a limit cycle for g-ʚ.

thereby ensuring the linear relationships necessary between $(u, \dot{u})$ and (y, v) to change system g-ʚ into the normal Hopf form

$$\begin{bmatrix} y \\ v \end{bmatrix} = \Phi^{-1} J \begin{bmatrix} u \\ \dot{u} \end{bmatrix} = \begin{bmatrix} -u + a\dot{u} \\ \frac{2\dot{u}+a(u-a\dot{u})}{\sqrt{4-a^2}} \end{bmatrix}, \qquad \begin{bmatrix} u \\ \dot{u} \end{bmatrix} = J^{-1}\Phi \begin{bmatrix} y \\ v \end{bmatrix} = \begin{bmatrix} \frac{y(a^2-2)+av\sqrt{4-a^2}}{2} \\ \frac{ay+v\sqrt{4-a^2}}{2} \end{bmatrix}. \tag{2.11}$$

In this case, the functions $f(y, v)$ and $g(y, v)$ are explicitly given by

$$f(y,v) = \frac{1}{4}a^2\left[\left(a^2-2\right)b+ac\right]y^2 + 2\sqrt{4-a^2}\left[\left(a^2-1\right)b+ac\right]yv - \left(a^2-4\right)(ab+c)v^2,$$

$$g(y,v) = -\frac{a\left(a^2-2\right)\left[\left(a^2-2\right)b+ac\right]y^2}{4\sqrt{4-a^2}} - \frac{1}{2}\left(a^2-2\right)\left[\left(a^2-1\right)b+ac\right]yv + \frac{\left(a^4-6a^2+8\right)(ab+c)v^2}{4\sqrt{4-a^2}}. \tag{2.12}$$

With these transformations into the Cartesian Hopf normal form, proof of the existence of limit cycles automatically follows. Furthermore, we are able to derive asymptotic approximations for both the amplitude[5] $\hat{A}$ and period $\hat{T}$ of these solutions, including their stability, in (y, v) space provided a parameter regime close to the bifurcation is chosen, $|a| \ll 1$, and the trajectories remain close to the origin $(y, v) = (0, 0)$. The reasoning is that in this regime, the nonlinearities of Eq. **??** are

[5]Note that this is not the amplitude, measured in magnitudes, of the original light curve, nor does it bear any relation with the coefficient A introduced previously. It is the amplitude of the oscillation in the transformed Hopf system.

Variable type	RRab	RRc	δ Scuti	All
norm. MSE < 1.00 solutions (%)	251 (16.7%)	468 (31.2%)	36 (36.0%)	755 (24.3%)
norm. MSE < 0.25 solutions (%)	12 (0.8%)	225 (15.0%)	15 (15.0%)	252 (8.1%)
norm. MSE < 0.16 solutions (%)	10 (0.7%)	179 (11.9%)	8 (8.0%)	197 (6.3%)
norm. MSE < 0.10 solutions (%)	7 (0.53%)	139 (9.3%)	8 (8.0%)	154 (5.0%)
norm. MSE < 0.01 solutions (%)	3 (0.2%)	40 (2.6%)	3 (3.0%)	46 (1.5%)
Sample size	1500	1500	100	3100

Table 2.6: Summary of SINDy results broken down by variable type (columns). The number of governing equation solutions found to match the observed light curve to within a given normalized MSE is shown, with percentage on the sample in parenthesis.

subdominant to the linear terms in y and v so that the results from the prototypical Hopf normal form carry over to our system. For our system, the amplitude and period are given by (see e.g. Glendinning, 1994):

$$\begin{aligned}\hat{A} &\sim \tfrac{-1}{16\omega}\left(\left(f_{yyy}+f_{yvv}+g_{yyv}+g_{vvv}\right)\omega + f_{yv}\left(f_{yy}+f_{vv}\right)\right.\\ &\qquad \left. -g_{yv}\left(g_{yy}+g_{vv}\right) - f_{yy}g_{yy} + f_{vv}g_{vv}\right)|_{(y,v)=(0,0)}\,,\\ &= \tfrac{(ab+2c)(b+ac)}{4(a^2-4)},\\ \hat{T} &\sim \tfrac{2\pi}{\omega} = \tfrac{4\pi}{\sqrt{4-a^2}},\end{aligned}$$

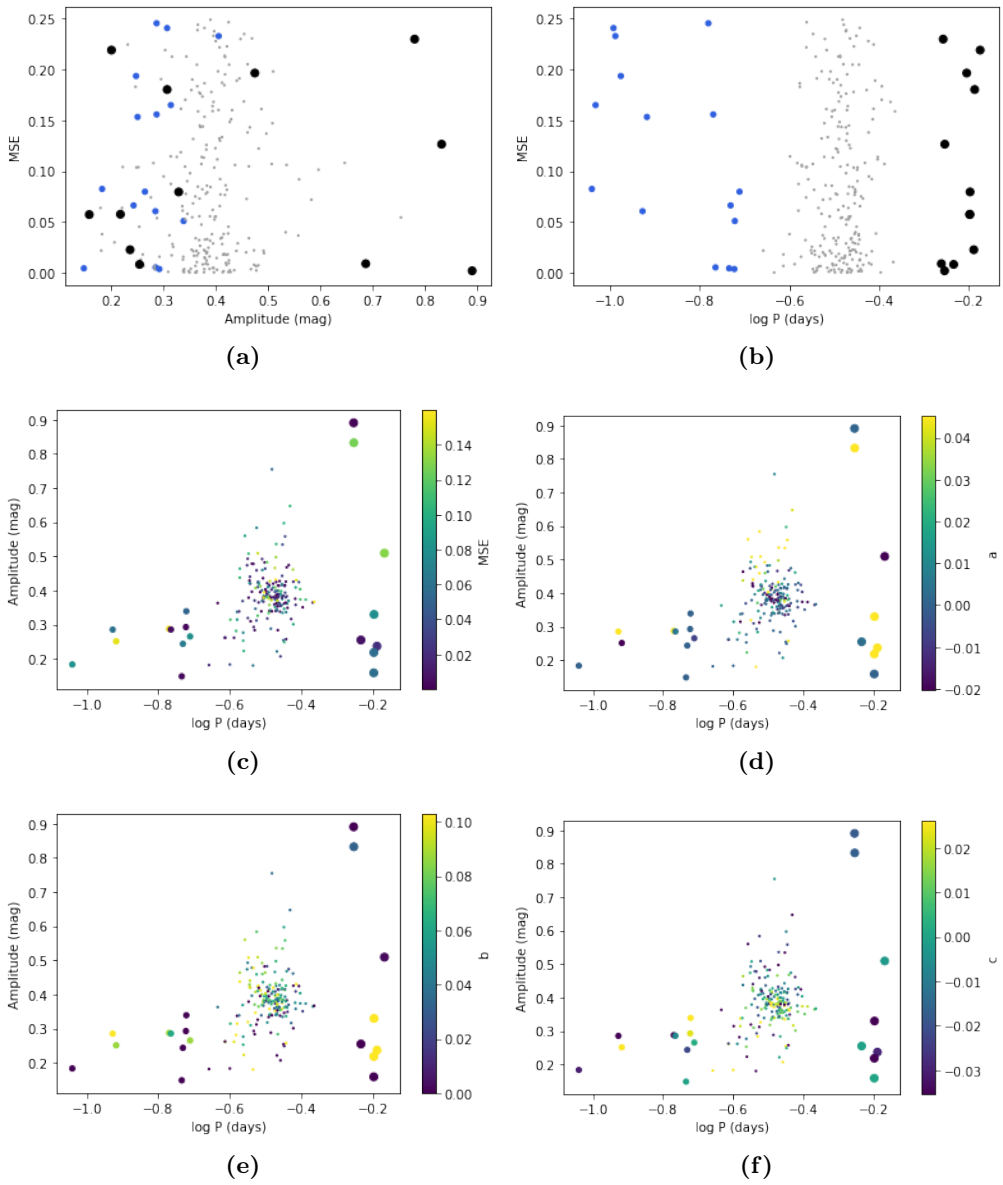
with the limit cycle being stable if $\hat{A} > 0$ and unstable if $\hat{A} < 0$. For a given value of a this corresponds to the area delimited by two intersecting lines through the origin in the (b, c) plane, as shown in Fig. 2.33.

2.5.4 Statistics on the full sample

Out of the 3100 stars in the sample, 755 showed an agreement with the solution to the governing equation discovered by SINDy with a normalized MSE below 1. We also considered more stringent criteria on MSE, as shown in Tab. 2.6. Interestingly, regardless of the threshold in MSE chosen, RRab stars appear quite impervious to SINDy, with a consistently lower success rate than RRc and δ Scutis. The latter two classes of variables perform similarly, with essentially identical success rate. Very good results for RRab, with normalized MSE below 1/100, are essentially absent, and it is not unlikely that they correspond to misclassified RRc stars.

In Fig. 2.34 we show the normalized MSE as a function of the light curve amplitude (top left panel) and period (top right panel). To avoid cluttering the plots, only stars with normalized MSE below 1/4 are shown here. There is no apparent relation with amplitude or period, but the RRab variables that have the lowest MSE appear to have periods more similar to the RRc, suggesting that they may be misclassified RRcs. In the other panels of Fig. 2.34 we show the Bailey diagram, with the values of MSE, a, b, and c color coded. Again no obvious patterns are apparent.

In Fig. 2.35 we plot MSE, a, b, and c against each other. The only emerging regularity is that c appears to anticorrelate with a for RRc variables.



(a) (b) (c) (d) (e) (f)

Figure 2.34: Normalized MSE as a function of the light curve amplitude (panel a) and period (panel b). Big black circles represent RRab variables, small gray circles RRc variables, and blue circles represent δ Scuti variables. Panels c to f show, in order, normalized MSE, a, b, and c color-coded onto the Bailey diagram. Colors scales have been clipped to the first and last decile of each variable. The sizes of circles codify the star type as in the first two panels.

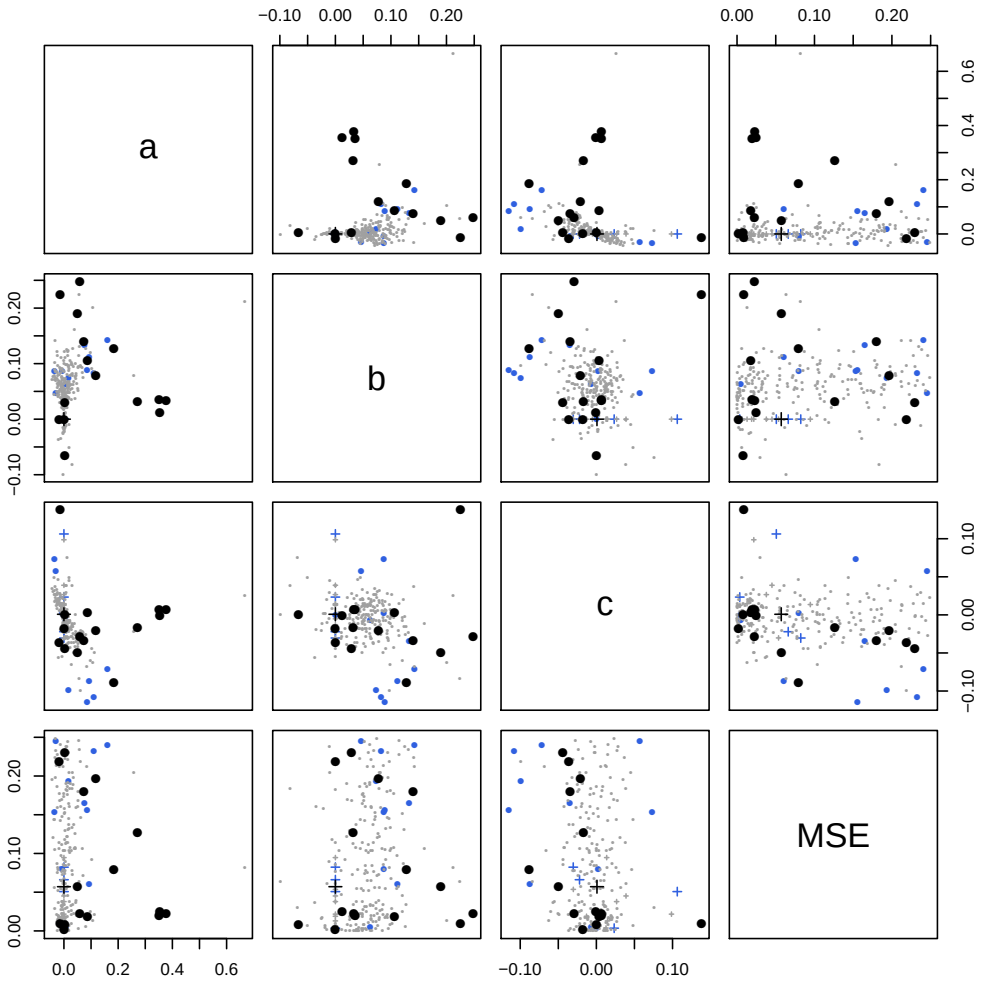


Figure 2.35: Scatter plots of the a, b, and c coefficients versus each other and versus normalized MSE. Big black symbols represent RRab variables, small gray symbols RRc variables, and blue symbols represent δ Scuti variables. Filled circles correspond to g-δ stars and crosses to s-δ stars.

2.5.5 Conclusion

Variable stars play a fundamental role in measuring stellar distances, allowing us to build up the cosmic distance ladder by calibrating secondary distance indicators. They can also be regarded as laboratories to test the accuracy of input physics in stellar models and the general reliability of such models. Detailed models of stellar interiors -including those of variable stars- typically rely on hydrodynamical simulations. However, like many other complex systems, variable stars are also amenable to a synthetic description based on mathematical relations aimed at capturing a simplified version of the underlying physics.

The art of capturing complex dynamical behaviour by means of concise sets of equations is as old as Science itself, but only recently it has become possible to learn such governing equations directly from the data using a machine learning approach.

In this section, we have shown that simple governing equations learned from variable-star light-curve data are able to reproduce the dynamics of RRc and δ Scuti stars, while RRab variables prove impervious to this approach at least with the modeling choices we adopted, with only a handful of stars from our RRab sample being fit successfully. Large amplitude RRab variables are affected by nonlinear phenomena such as shock formation and propagation. There is evidence that across the phases approaching the maximum in luminosity the radial velocities in the outermost envelope regions become sonic (see Gillet et al., 2019a, and references therein). This may be the underlying physical reason for our finding. The presence of outliers in the light curves and general issues with the quality of the data are, on the other hand, likely to affect both RRab and RRc stars similarly, so they should not play a role in the different behaviour of the two types of variables.

At any rate, when modeling is successful, the learned equations take on a universal form which is essentially a perturbed harmonic oscillator. The perturbing nonlinear terms however are non-trivial: even in a simple particular case that we solved exactly, two classes of solutions emerge, only one of which has periodic behavior. The general case, which falls within the generalized Liénard equation class, displays a limit cycle for which we found a simple stability condition.

While variable type has a clear effect on whether effective governing equations can be discovered, the position on the Bailey diagram appears not to. Therefore our result may have implications for variable type classification because it reveals information on the light curve shape that is not immediately tied to period and amplitude. A straightforward approach could be to use the coefficients found by SINDY as features to train a simple ML model, possibly after extending our analysis to higher order equations or higher degree polynomials on the right-hand side, thus possibly gaining the ability to model also RRab stars. This may be valuable in light of the upcoming large scale surveys, such as the Vera C. Rubin Observatory, and could allow us to test whether our approach can be applied to describe more variable star physics such as e.g. the Blazhko effect.

2.6 Variable Star Light Curves in Koopman space

Based on Variable Star Light Curves in Koopman Space (ApJ under second review). In the precedent section (based on Pasquato et al. 2022), we leveraged a data driven approach to find a non-linear, non-conservative oscillator equation capable of describing the light curves of first overtone pulsators. This result was based on the SINDy algorithm. The same data-driven approach failed to obtain an equally concise equation for fundamental pulsators. While this is related to the choices of the

basis function and regularization that we made in that work, it is an indication of the different challenges met in trying to model the two different classes of pulsators. The most popular diagnostic to separate RRc and RRab variables is the Bailey diagram (see Fig. 2.2), since the two groups cluster around two different regions of this plane. However, the distinction between the two is not very sharp, as there is a region of the instability strip, located between the regions in which RRc (hotter) and RRab (cooler) attain a stable limit cycle, in which RRLs pulsate simultaneously in the first two radial modes (mixed mode, RRd).

In this study, we move beyond this restriction using the DMD approach, a mathematical framework presented by Schmid (2010) and explained in Sec. 2.3.4, and to our current understanding, infrequently employed in astronomy thus far (Albidah et al., 2021, 2022, Darling and Widrow, 2019, Hori et al., 2020, Linkmann and Dikpati, 2021, Linkmann et al., 2020, Widrow et al., 2020) but rigorously utilized in several diverse scientific scenarios (such as, notably, fluid dynamics; Taira et al., 2017). The strength of this method is that the derived dynamic modes might be intuitively perceived as an expansion of global stability modes and might be feasibly associated to the foundational physical dynamics inherent in the data series.

In this study we describe the temporal progression of the optical brightness of a group of RRL stars within the globular cluster ω Centauri. We opted to engage with cluster variables as this diminishes most variances attributed to distance and absorption, given all stars reside in the identical locale. Moreover, despite the emergence of metallicity gradients in ω Centauri (Johnson and Pilachowski, 2010, Kraft, 1994, Lee et al., 1999, Norris and Da Costa, 1995), our selection offers a degree of chemical and environmental consistency for our star set. We systematically derive models for nonlinear dynamical mechanisms capturing the time-evolution of the brightness of each variable star, and validate that the predicted change aligns with the gathered data. Such models are articulated in terms of linear frameworks in designated Koopman spaces, an eigenvector base of which is retrieved directly by our approach. The linked eigenvalues fundamentally distill the light curve and might, in context, be utilized as attributes for machine learning endeavors. We ascertain that the Koopman space needed to encapsulate RRc variable dynamics is fewer relative to the one required for representing RRab variable dynamics. This echoes the less intricate and more consistent shapes of RRc curves and renders a direct strategy for RRab vs RRc categorization. Ultimately, we recognize stars impacted by the Blazhko effect by contrasting the eigenvalues related to light curves documented at different junctures in the Blazhko period.

2.6.1 Data processing

Given the high quality of data, presented in Sec. 2.2.3, the data was simply interpolated, using the *gcvspline* Python wrapper (Le Losq and Feng, 2007) of a FORTRAN package for generalized, cross-validatory spline smoothing. We interpolate the remaining light curves repeated 5 times and picked equidistant (in phase) snapshots from this spline with 30 points per phase which makes a smoothed out, periodic set of 150 points from which DMD can build a model.

2.6.2 Results

The DMD reconstruction of the light curve for an example RRab is shown in Fig. 2.36 and for an RRc in Fig. 2.37. In general, the quality of the reconstruction, determined in terms of the mean

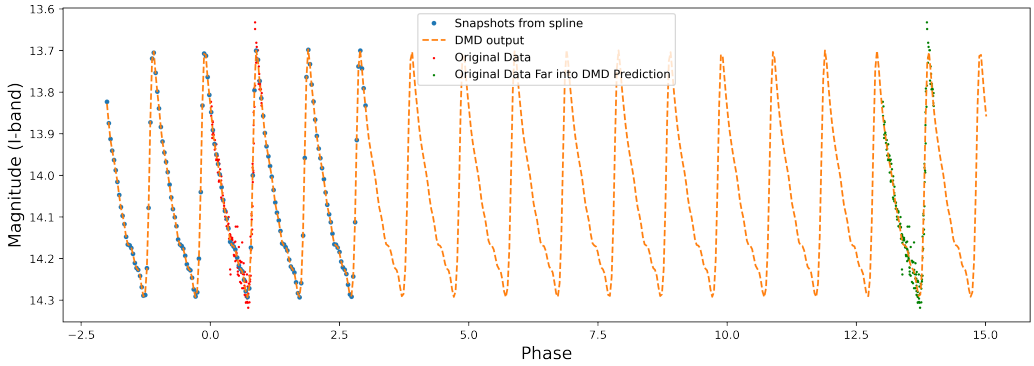


Figure 2.36: DMD reconstruction (orange dashed line) of the interpolated light curve (blue points) compared with the original (red points) and time shifted (green points) raw data for RRab star V59.

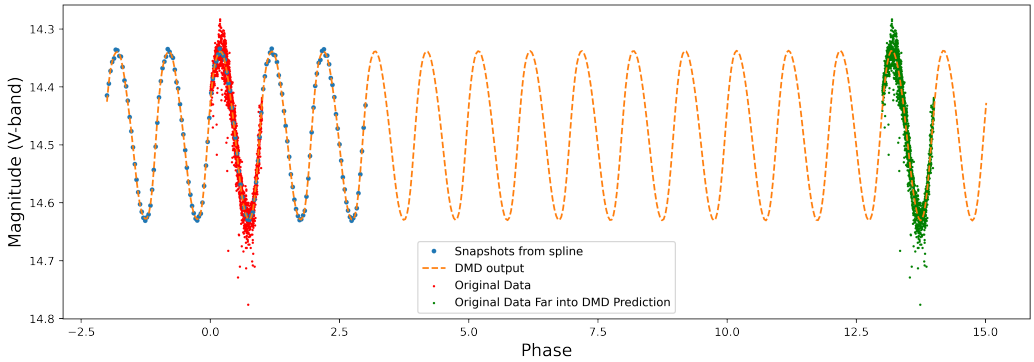


Figure 2.37: DMD reconstruction (orange dashed line) of the interpolated light curve (blue points) compared with the original (red points) and time shifted (green points) raw data for RRc star V103.

squared error (MSE) with respect to the original data, varies with the number of eigenvalues (and of the corresponding eigenvectors) used for the reconstruction (see Figure 2.46).

Separating RRab from RRc variables: As we vary the Koopman dimension d, which corresponds to the maximum allowable number of eigenvectors, we find that DMD models from RRc and RRab variables respond differently to increasing d. In fact, as the spatial dimension increases, the normalized mean squared error (nMSE; the ratio between the mean squared error (MSE) of the reconstructed light curve and the MSE of a constant fit) for models of RRc variables drops off steadily compared to models of RRab variables which have a more or less constant nMSE until about $d = 25$ where we can observe a sharp drop. This is shown in Fig. 2.38 in the left panel and, for the number of eigenvalues, in the right panel. Note how nMSE starts decreasing immediately with d for the RRc and keeps decreasing in a linear fashion, while it suddenly drops at about $d = 25$ for the RRab[6]. From observing this phenomenon, we can choose a nMSE threshold (the nMSE is calculated 10 phases away from the last snapshot to promote stability of the DMD model) such that, between RRc and RRab variables, different numbers of spatial dimensions (and different numbers of

[6]At larger d MSE slightly increases again, which is somewhat puzzling. From a mathematical point of view, increasing d allows for capturing weaker and weaker oscillations within the dynamics of the system. On real data, such oscillations may correspond to noise due to the error intrinsic in the experimental data. When d is high enough such sequences are amplified in the Koopman space and bring in further information affecting the quality of the final result.

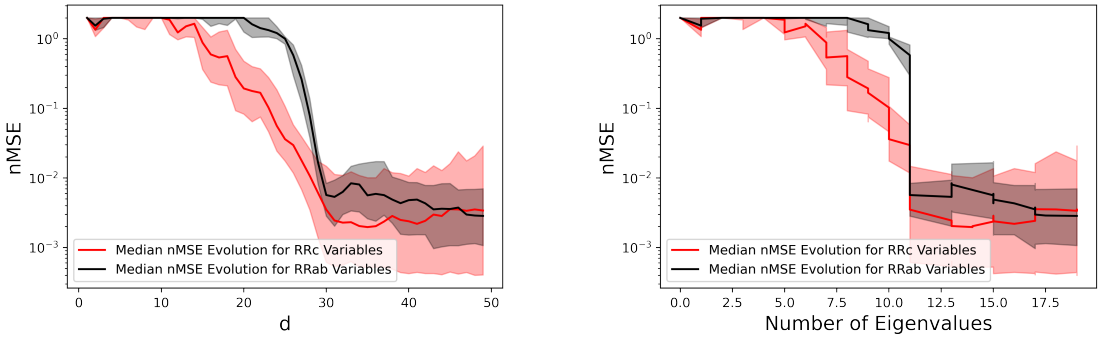


Figure 2.38: Left panel: nMSE achieved by the DMD reconstruction against the original light curve as a function of the spatial dimension of the HODMD snapshot (d parameter in pyDMD). The black solid line represents the median curve for the RRab stars, the red solid line for the RRc stars. The lower and upper boundaries of the black and red shaded areas represent the first and third quartiles respectively. Right panel: Same as the left panel but nMSE is plotted as a function of the number of eigenvalues.

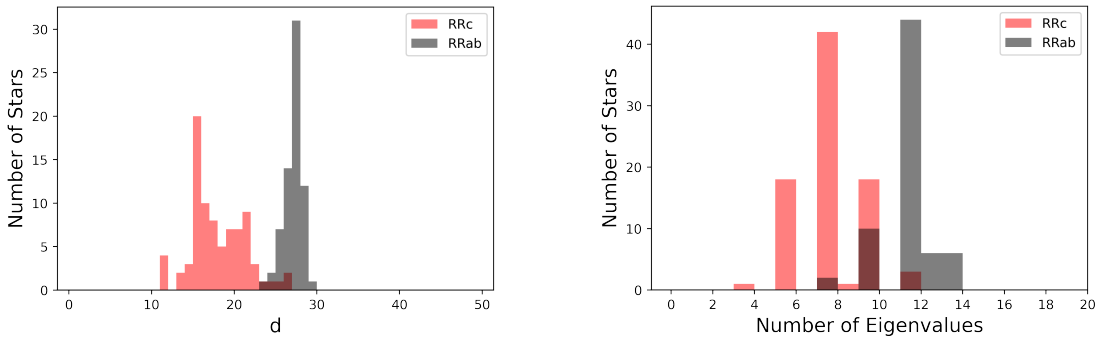


Figure 2.39: Left panel: histogram of the minimum values of d needed to achieve nMSE $= 0.45$ (or below) for the RRab (black) and the RRc variables (red). Right panel: same, but for the number of eigenvalues.

eigenvalues) are sufficient to produce models which yield an error equal or lower than the threshold.

We find that, for the nMSE to be below 0.45, models for RRc variables require between 11 and 26 spatial dimensions (d) and between 3 and 11 eigenvalues. For RRab variables, between 23 and 29 spatial dimensions (d) and between 7 and 13 eigenvalues are needed to yield an nMSE below threshold. We notice that there is some overlap especially in the numbers of eigenvalues. However, this distinction leads to a straightforward classification algorithm. In fact, we can find that, most likely, a variable star is RRc if $d < 24$ is sufficient to yield an nMSE below 0.45 and is RRab otherwise. The histogram of the values of d needed to achieve nMSE$= 0.45$ is shown in Fig. 2.39. The relevant confusion matrix for the associated classifier with $d < 24$ is shown in Tab. 2.7. This is calculated on the same data set on which the threshold $d = 24$ was chosen, so the associated measure of accuracy (97%) is optimistically biased and should not be used to predict performance on unseen data. However this is still an indication that the modes identified by DMD are naturally suited to classification. One more point to note is that classification based on the Koopman dimension d needed to achieve a given level of mean squared error (nMSE=0.45), which depends on the shape of the light curve, overlaps quite well with the distinction between RRab and RRc variables in the Bailey diagram, as shown in Fig. 2.40. Discrepancies between the two could be leveraged to single out potentially interesting stars, among which perhaps mixed-mode oscillators. A different way

	$d < 24$	$d \geq 24$
RRab	1	67
RRc	79	4

Table 2.7: Confusion matrix for a classifier based on the minimum Koopman dimension d needed to achieve nMSE= 0.45.

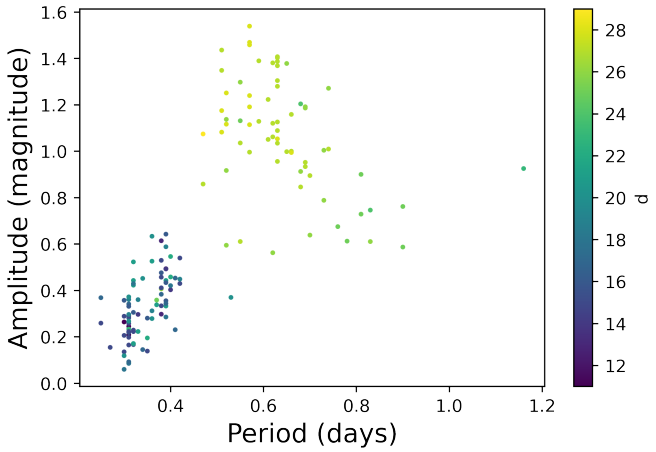


Figure 2.40: Bailey diagram as in the right panel of Fig. 2.7, with the Koopman dimension d needed to attain nMSE = 0.45 color coded.

to look at the data presented in Fig. 2.40 is shown in Fig. 2.41, where we plot d as a function of the amplitude of the magnitude variation of each star (left panel) and as a function of the period (right panel). Additionally, in Fig. 2.42 we show the number of eigenvalues as a function of the same two quantities. Regarding the dependence on amplitude, both d and the number of eigenvalues increase with amplitude. Since both are measures of the complexity of the light curve shape, this is an expected result: higher amplitude oscillators become increasingly anharmonic. However, the increase is largely driven, in both cases, by the fact that large amplitude variables are RRab and small amplitude variables are RRc. Within each class the trend is weaker but still visible for the RRab, and absent for the RRc. Regarding period, the general trend is again increasing for both quantities, but it is decreasing for the RRab. This means that even though RRab stars typically have both a longer period and a higher number of eigenvalues than RRc, within the class of the RRab an increasing period corresponds to a *lower* number of eigenvalues, i.e. a simpler light curve. The relevant correlation coefficients and p-values are reported in Tab. 2.8. This evidence fully supports

Class:		RRab		RRc	
Parameter 1	Parameter 2	R	p value	R	p value
N_{eig}	Amplitude	0.45	1.2×10^{-4}	0.13	0.26
N_{eig}	Period	-0.68	2.3×10^{-10}	-0.08	0.46
d	Amplitude	0.45	1.0×10^{-4}	0.05	0.68
d	Period	-0.71	1.1×10^{-11}	0.09	0.41

Table 2.8: Spearman correlation coefficients (col. 3, col. 5) between the number of eigenvalues and d with amplitude and period over each class of variable stars and the associated p-values (col. 4, col. 6).

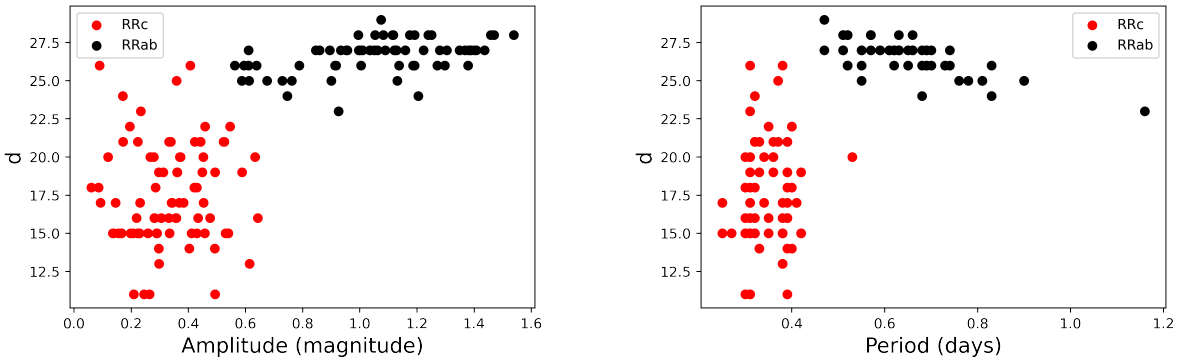


Figure 2.41: Left panel: Koopman space dimension d as a function of amplitude for RRab (black points) and RRc (red points). Right panel: Koopman space dimension d as a function of period, same color coding as the left panel.

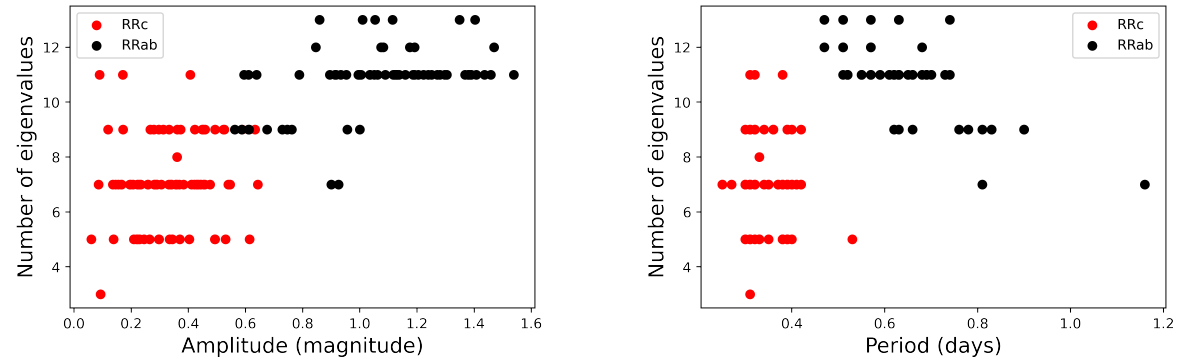


Figure 2.42: Left panel: number of eigenvalues as a function of amplitude for RRab (black points) and RRc (red points). Right panel: number of eigenvalues as a function of period, same color coding as the left panel. It is known that short period, large amplitude RRab are harder to fit with fourier and in our case they also need more eigenvalues.

theoretical and empirical findings concerning the variation of RRab pulsation properties across the instability strip. The RRab display a steady decrease when moving from the blue to the red edge of the instability strip. The variation also applies to the shape of the light curves, indeed RRab close to the blue edge display a sawtooth shape and a more sinusoidal shape approaching the red edge. The main culprit causing this difference is convection, since it becomes more efficient when moving from hotter (bluer) to cooler (redder) effective temperatures and eventually quenches the pulsation activity (Bono and Stellingwerf, 1993).

In terms of physical interpretation, it is particularly interesting that the MSE drops for RRab when d approaches 30, the number of points sampled along a period by our interpolation strategy. Since the data we fed to HODMD repeats exactly after one period, we have that $x_{31} = x_1$ and in general $x_{30+k} = x_k$ for every k. Thus the vector to which DMD is effectively applied at $t+1$ is $\mathbf{x}_{t+1} = (x_2, x_3, ..., x_{30}, x_1)$ and at t it is $\mathbf{x}_t = (x_1, x_2, ..., x_{29}, x_{30})$ and the matrix $\mathbf{A}$ such that $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$ becomes simply $[\mathbf{e}_{30}, \mathbf{e}_1, \cdots, \mathbf{e}_{29}]$. The eigenvalues of this matrix are $\lambda_k = e^{ki\pi/30}$ for $k = 0, ..., 29$, so they are equally spaced on the unit circle and include $\lambda_0 = 1$. We can thus conclude that the DMD is predicting the future evolution of the light curve by leveraging solely the periodicity of the data and nothing else; in an intuitive sense we can claim that an important fraction of the

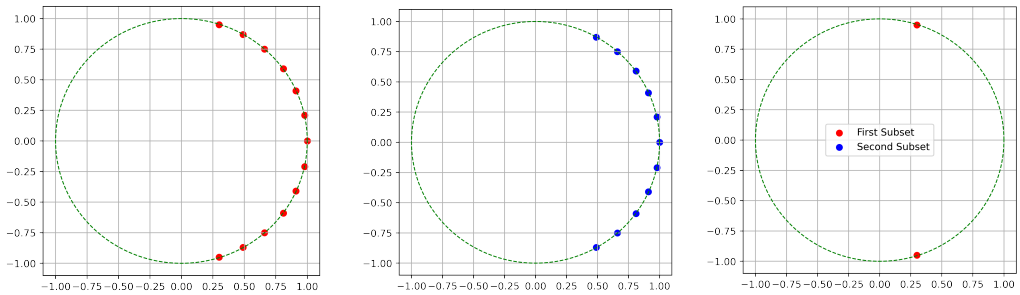


Figure 2.43: Left panel: DMD eigenvalues (red points) for variable V5 calculated between Julian time 2449858.7698 and 2449869.7601, corresponding to the beginning of the observation period. The green dashed line corresponds to the unit circle in the complex plane. Middle panel: DMD eigenvalues (blue points) for variable V5 calculated between Julian time 2450971.5723 and 2450985.6906, corresponding to the end of the observation period. Right panel: DMD eigenvalues that differ between the two.

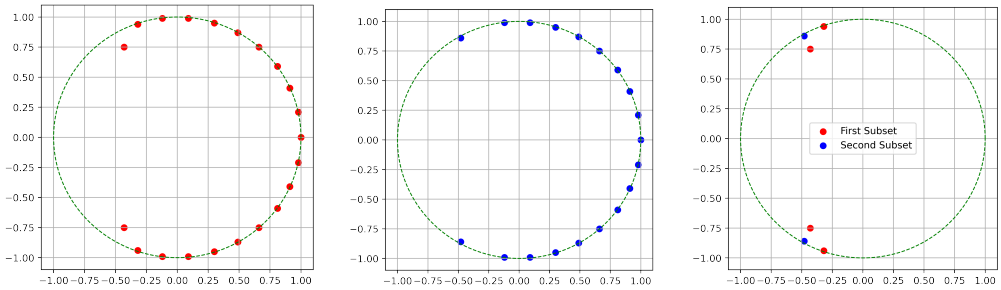


Figure 2.44: Left panel: DMD eigenvalues (red points) for variable V115 calculated between Julian time 2449858.7468 and 2449869.7635, corresponding to the beginning of the observation period. The green dashed line corresponds to the unit circle in the complex plane. Middle panel: DMD eigenvalues (blue points) for variable V115 calculated between Julian time 2450971.5723 and 2450985.6906, corresponding to the end of the observation period. Right panel: DMD eigenvalues that differ between the two.

RRab curves is 'incompressible'. This is no news, since we showed in the previous work, that most RRab curves could not be modeled with SINDy under our choice of basis.

Blazhko effect: A final application of this description that goes beyond distinguishing fundamental and first overtone pulsators is related to Blazhko effect. DMD summarizes a light curve with a handful of complex numbers: do these differ over time for a Blazhko variable, and how? In Fig. 2.43 we show the DMD eigenvalues calculated during the first part of the observation period (left panel) and during the second part (right panel). Most eigenvalues match exactly except for a very limited number.

These eigenvalues may correspond to Koopman modes specifically associated to Blazhko evolution, since V5 is a suspected Blazhko. In any case DMD has revealed a secular change in its light curve. Potentially, this may be used to automatically detect Blazhko variables without the need for direct visual inspection of the light curve. For comparison, in Fig. 2.44 and 2.45 we show the same for two more stars that are deemed affected by the Blazhko effect.

DMD and the physics of the system Theoretical studies of Koopman operator, e.g. Bevanda et al. (2021), show clearly that representations via the Koopman paradigm generalize the notion of mode analysis from linear to non-linear systems allowing for amendable relevance determination of the constituents of the full dynamics. In particular, the spectrum of the Koopman operator allows

H

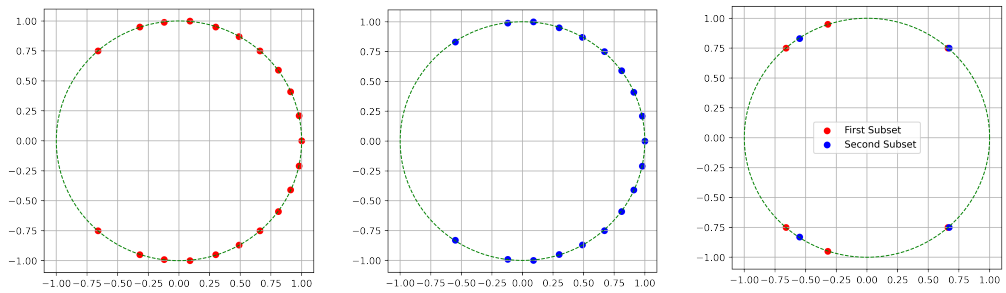


Figure 2.45: Left panel: DMD eigenvalues (red points) for variable V120 calculated between Julian time 2449858.7468 and 2449869.7635, corresponding to the beginning of the observation period. The green dashed line corresponds to the unit circle in the complex plane. Middle panel: DMD eigenvalues (blue points) for variable V120 calculated between Julian time 2450970.5019 and 2450985.6906, corresponding to the end of the observation period. Right panel: DMD eigenvalues that differ between the two.

for a decomposition of the nonlinear system into different dynamic regimes (fast-slow dynamics) while linearly evolving coordinates contain intrinsic information concerning regions of attraction that are relevant for analysis. Figure 2.46 emphasizes how the dynamic is progressively defined by eigenvalues and eigenvectors.

2.6.3 DMD vs Fourier Series:

Several experimental studies provide robust evidence supporting our claims, and a detailed examination of these can be found in the works of Fujii et al. (2020), Pan (2021), Schmid (2022). The Dynamic Mode Decomposition (DMD) stands out due to certain distinct advantages over other prevalent algorithms, such as those based on Fourier methods. Specifically:

- DMD's approach, which involves trading the challenges of infinite-dimensionality for the advantages of linearity, paves the way for the effective use of well-established linear techniques, making the analysis more tractable.
- The inherent linear dynamics associated with the new coordinate system simplifies various challenges. For instance, the complications that arise from numerical integration and the non-convex nature of optimization in state dynamics become more manageable, especially when applied to model predictive control.
- The Koopman operator paradigm offers a comprehensive system description, as opposed to a localized one. Here, a single iteration of the Koopman operator acting on an observable mirrors an iteration along all trajectories of the system. Importantly, this should not be misconstrued as a mere local linearization centered around a specific operational point.

From a hands-on viewpoint, the latter point, offered by the Koopman operator paradigm, facilitates more concise representations of dynamical systems in comparison to conventional methodologies. This representation can be instrumental in establishing a clearer connection with the foundational physics governing these systems. In fact, in the context of variable stars, both theoretical (Bono and Stellingwerf, 1994) and empirical evidences (Lub, 1977) suggest that both fundamental and first overtone RR Lyrae stars exhibit distinct secondary features on their light curves. These features commonly manifest as bumps during the decreasing phase or dips during the ascension phase of

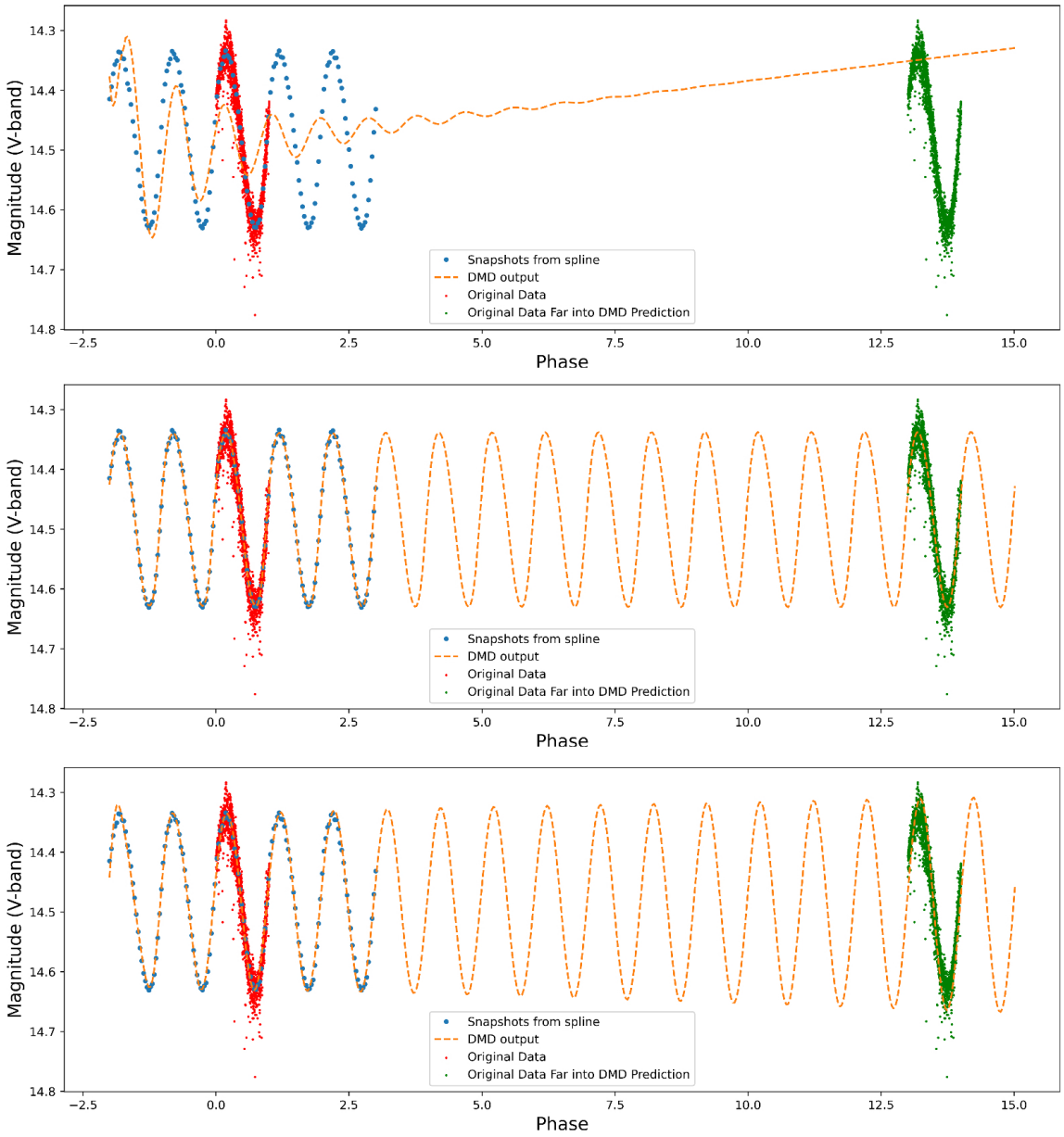


Figure 2.46: Reconstruction of a light curve with an increasing number d of eigenvectors. Top panel : $d = 10$. Center panel : $d = 20$. Bottom panel : $d = 30$.

the light curve. As said in the chapter introduction, nonlinear phenomena take place in RRL stars, especially linked to the generation and movement of shocks in the star's outermost layers, as elaborated by (Gillet et al., 2019a, and references therein). The morphology of these light curves is influenced by various factors including physical parameters (like stellar mass, effective temperature, and surface gravity) and the chemical composition. However, they remain predominantly uninfluenced by uncertainties concerning distance and reddening. Recognizing their significance, the astronomical community has put forth considerable effort, as highlighted by (Jurcsik and Kovacs, 1996a), to parameterize these features. While the bumps can be fit by using either Fourier series or polynomials, the dips can only be fit with a custom base and an limited number of terms while a Fourier series would require a much larger number of terms (Braga et al., 2019b)..

2.6.4 Conclusions

DMD and related algorithms have recently found application to fluid dynamics and other branches of physics that deal with high-dimensional snapshots either from observations or from simulations. In our application to variable stars, snapshots at a given time contain at most a few values, corresponding to magnitudes in various filters and possibly their derivatives. We thus applied the HODMD algorithm, which essentially synthesizes a higher-dimension snapshot from delayed coordinates. This is the first application of DMD-related methods to variable star light curves to date. This was done with the goal of obtaining features from light curves in a principled way that could be amenable to physical interpretation. In DMD the observed dynamics is decomposed into modes, similar to normal modes in a coupled oscillator system; each mode corresponds to an eigenvalue, associated either to exponential decay (or increase) or to periodic oscillations.

We obtained a reconstructed light curve as a linear combination of the individual evolution of such modes, and compared it with the original data over a sample of 151 stars. We measured the discrepancy between the reconstructed light curve and the original by means of the mean squared error normalized to that of a constant fit. We are able to obtain consistently good fits (with nMSE under 10^{-2}) if we let the number of eigenvalues increase as needed. We found that simpler light curves, such as those of RRc variables, require only a limited number of eigenvalues. Conversely RRab stars, which have more complex light curve shapes, require the inclusion of more eigenvalues to be described with good accuracy. We have quantified this, finding that that the Koopman dimension d needed to achieve a given normalized MSE in reconstructing the curve behaves differently for RRab and RRc variables. We used this as a basis for a classifier, choosing a threshold of 0.45 normalized MSE, to achieve which the modal RRab star requires $d = 27$ (resulting in a mode of 11 eigenvalues) and the modal RRc star, $d = 15$ (corresponding to a mode of 7 eigenvalues). The threshold value 0.45 corresponds to the maximum (over all the light curves) of the minimum normalized MSE reached.

We found that the number of eigenvalues that needs to be included to obtain a good representation of the curve varies with amplitude and period in accord to the fact that RRab and RRc stars occupy distinctive loci in the period-amplitude plane.

While the DMD decomposition bears some resemblance to Fourier methods, which have been extensively applied to time domain astronomy and variable star light curves in particular, the advantage of DMD is to identify modes that are potentially physically meaningful. This is the case of the long term period variability due to the Blazhko effect, where indeed a few eigenvalues found by DMD change between the beginning and the end of the observation period for the confirmed

Blazhko variables in our data set. Whether this is indeed an indication of DMD finding specific modes associated to Blazhko evolution will be discussed in further detail in upcoming work.

Final Remarks

In our investigation of variable stars data using interpretable methods, we have highlighted the practical advantages of data-driven tools in astronomy. Though the methodologies employed are new for the field (like sindy and DMD), sometimes the results shown may not be groundbreaking, their significance lies in the emphasis on interpretability and simplicity over traditional performance metrics. These studies also reaffirm the importance of deriving insights from data, which is crucial for the accurate progression of astronomical research.

Chapter 3

Intermediate-Mass Black Hole

The second case study stems from an accepted paper (27 October 2023) and ready to be published in the Astrophysical Journal (ApJ). This research initiative was inspired by a transformative two-month period spent in Montreal at MILA[1]+Université de Montreal+CIELA organizations, collaborating with Yashar Hezaveh and Laurence Perreault-Levasseur group. The impetus for this work was Cynthia Rudin's thought-provoking article, "Please Stop Explaining Black Box Models for High Stakes Decision" (Rudin, 2019) which introduced a paradigm shift in our perception of explainable tools.

While our study's focus pertains to IMBH, which are not subject to high-stakes decisions like those in the judiciary or healthcare systems, the article's perspective and insights hold broader relevance. Rudin's work shows the limits of explainable tools in certain contexts and emphasizes the primacy of interpretability over conventional metrics. Thus, exploring the implications of this viewpoint in the IMBH domain can yield valuable lessons applicable to interpretability challenges across various fields (more on the limits of explainable AI and the need of natively interpretable models in Chapter 1).

The chapter is divided like this: a quick introduction on the IMBH detection problem via a data-driven approach, afterwards I present possible evidences for IMBH, different scenarios of IMBH formation and why star clusters are great environments for the formation and evolution of IMBH. Next, I present the MOCCA data used to train our models and the real GCs used to deploy the aforementioned models. Then, the different ML tools are explained, especially the interpretable model CORELS and the model-agnostic local explanations via anchors of the black box used (XGBoost). Finally, the models are trained and tested on the simulations and deployed on real GCs. In the end, a discussion on the results and some conclusion of this work are presented.

3.1 Introduction

The detection of quasars at high redshift (see e.g., Mortlock et al., 2011, Schindler et al., 2023) has sparked the need for a mechanism to rapidly assemble supermassive black holes. Bridging the gap between stellar-mass remnants and supermassive black holes (mass range between 10^2 to $10^5 M_\odot$), IMBH have emerged as potential seeds for the latter (see Woods et al., 2019, for a review). Consequently, searches for IMBHs in the present-day universe are actively underway, with particular

[1]Quebec AI Institute (originally Montreal Institute for Learning Algorithms.

focus on high-density environments such as the Galactic center (Oka et al., 2016, Ballone et al., 2018, Takekawa et al., 2019, Takekawa et al., 2019, Takekawa et al., 2020, Kaneko et al., 2023, The GRAVITY Collaboration et al., 2023), dwarf galaxies (Mezcua et al., 2016, 2018), and globular clusters (Maccarone and Servillat, 2008, Farrell et al., 2009, Cseh et al., 2010, Strader et al., 2012 Bachetti et al., 2014, Kains et al., 2016, Kızıltan et al., 2017, Lin et al., 2018 Su et al., 2022 among others). While direct IMBH detection remains challenging, indirect methods based on dynamical effects in GCs offer promising candidate hosts for follow-up investigations:

Applying ML models to indirect IMBH detection: Applying ML models to indirect IMBH detection presents both opportunities and challenges: handcrafted physical models offer interpretability but may be outperformed by more flexible data-driven models. However, data-driven approaches have drawbacks, such as high model complexity leading to black-box behavior and the reliance on simulations, which can introduce biases into training data.

Explainability and interpretability: In this context, the distinction between explainability and native interpretability central to this thesis becomes crucial: explainability tools like anchors (Ribeiro et al., 2018) can shed light on the behavior of black-box models locally, providing understandable logical rules for predictions. Meanwhile, natively interpretable models, such as CORELS (Angelino et al., 2017), are inherently understandable by humans without requiring post hoc explanations. In this study, we explore both approaches and assess their performance in detecting IMBH candidates in GCs.

Training on simulations, predicting on observations: Furthermore, we address the challenge of training ML models on simulations and predicting on observations, aiming to gauge how far individual predictions correspond to actual observed physical systems compared to the simulated training data. To ensure reliable predictions, we focus on measuring individual data point distances from the training data, rather than overall generalization error. The combination of interpretability and accurate predictions is vital in advancing our understanding of IMBHs and their role.

3.2 Intermediate Mass Black Holes

IMBH serve as the astrophysical bridge between stellar-sized and supermassive black holes, characterized by a mass range of $10^2 - 10^5 M_\odot$. Their potential role as progenitors to the supermassive black holes observed in the early universe has garnered considerable attention, especially in light of the discovery of high-redshift quasars (z$\sim$ 6) (Fan et al., 2001, Madau and Rees, 2001).

While unambiguous electromagnetic evidence for IMBHs remains elusive, several candidates have been put forth over the past decades. Notably, the dynamics of certain globular clusters, such as 47 Tucanae and G1, infer the presence of IMBHs at their centers (Gebhardt et al., 2005, Kızıltan et al., 2017, among others). One of the most compelling candidates for an IMBH is the hyperluminous X-ray source HLX-1 in the ESO 243-49 galaxy, standing out among ultra-luminous X-ray sources (Farrell et al., 2009, Godet et al., 2014, and references therein). Additionally, the nuclei of certain dwarf galaxies may host black holes with masses around $10^5 M_\odot$ (Filippenko and Sargent, 1989, Barth et al., 2004, Greene and Ho, 2004 for examples). Black holes with masses in

the range $10^4 - 10^5 M_\odot$ are also suspected to reside in high-velocity compact clouds (Tanaka, 2018, Takekawa et al., 2019, 2020). The gravitational wave event, GW190521, associated with a binary black hole merger resulting in a mass of approximately 150 $M_\odot$, has provided strong indications for the existence of IMBHs (Abbott et al., 2020). This detection not only sets a precedent for future observations with existing and forthcoming gravitational wave detectors but also reinforces the merger formation scenario for IMBHs. Such scenarios suggest that smaller black holes, typically of stellar origin, undergo frequent mergers in dense environments like star clusters (Coleman Miller and Hamilton, 2002, Giersz et al., 2015, Rodriguez et al., 2019, among others). Since the data used for this work come from the search of IMBH in star clusters (especially globular clusters), we focus on the formation mechanism in such objects. For comprehensive reviews on IMBHs, the reader is directed to Mezcua (2017) and Greene et al. (2020). For completeness some of the alternative formation mechanisms explored are:

- the collapse of extremely massive metal-poor stars exceeding 230 $M_\odot$ (Bond et al., 1984, Woosley et al., 2002).
- potential primordial IMBH formations from gravitational instabilities in the early Universe (Kawaguchi et al., 2008, Carr et al., 2016).
- runaway star collisions within star clusters (Colgate, 1967, Sanders, 1970, Freitag et al., 2006 among others).

SCs are great environments for the formation and evolution of IMBHs due to a variety of astrophysical processes at play: the repeated mergers of stellar-mass black holes, as outlined by Coleman Miller and Hamilton (2002), the birth of exceptionally massive stars during the early evolutionary stages of the cluster, and results of runaway collisions, as indicated by Fujii and Portegies Zwart (2013).

Crucially, the internal dynamics of SCs significantly influence the efficiency of binary black hole (BBH) mergers. The phenomenon of core collapse in SCs amplifies dynamical interactions and fosters the dynamical hardening of binary systems. This is supported by studies like those of Sugimoto and Bettwieser (1983),Spitzer (1987), Goodman and Hut (1989), Trenti et al. (2007), Hurley and Shara (2012), Beccari et al. (2019), Di Cintio et al. (2022).

A salient feature of the repeated merger mechanism is its longevity. An IMBH can continue to evolve and expand throughout the SC's lifespan, as suggested by Giersz et al. (2015). However, there's an inherent risk: should a BBH merger lead to an ejection from the SC due to gravitational-wave recoil, the merger chain can abruptly cease.

Such kick velocities, which arise from gravitational-wave recoil, might far surpass the escape velocity of the hosting SC, as emphasized by Merritt et al. (2004), Madau and Quataert (2004), Favata et al. (2004), and Lousto and Zlochower (2011). Consequently, the hierarchical development of an IMBH through recurrent BBH mergers becomes more plausible in SCs with substantial escape velocities, notably the colossal nuclear SCs, as indicated by Antonini et al. (2019b), Fragione et al. (2020), and Mapelli et al. (2021).

Figure 3.1: Globular cluster, M10. *Credits:* Till Credner and Sven Kohle, Observatorium Hoher List.

3.3 Globular clusters

As reported in Sec. 3.2, different mechanism undergo in the rich dynamic of a SCs. Our work, focused on Galactic GCs. In order to understand more about these objects and why we choose some features instead of others during the training of our models, some physical description of the relevant physical quantities is necessary. Globular Clusters (GCs) are ancient, spherical conglomerates of 10^5 to 10^6 stars. GCs are found predominantly in the halos of galaxies, orbiting their central bulges. GCs are characterized by a well-defined color-magnitude diagram, largely devoid of young blue stars. Their stellar population can be approximated by a singular stellar isochrone, suggesting a nearly-coeval formation or multiple stellar populations (multiple population phenomenon, Milone and Marino, 2022) depending on the clusters and the host galaxy. When choosing what parameters that characterize a GCs, we can use observational properties of the clusters. We decided to use 6 out of the 63 parameters provided in the MOCCA simulation (see Giersz et al., 2015 for a comprehensive description of the rest of parameters) as features for the models. In particular the half-mass relaxation time, the central velocity dispersion, the total luminosity, the central surface brightness, the core radius, the half-light radius. We chose few parameters aligning with the simplicity-through-sparsity principle. Recognizing that the utilization of all 63 parameters as features could enhance the model's performance, our decision to focus on these six parameters was driven by a preference for simplicity and practical applicability, especially considering their feasibility for real clusters.

Central velocity dispersion: A key parameter to characterize the dynamical state of the cluster is the central velocitiy dispersion. The latter is a measure of the average kinetic energy of stars in the core of the cluster. It provides insights into the gravitational potential and mass distribution within the cluster. For a collection of stars in equilibrium, the velocities will follow a Gaussian distribution, and the width of this distribution gives the velocity dispersion.

Mathematically, it can be defined as the root-mean-square of the velocity deviations from the mean velocity:

$$\sigma_0 = \sqrt{\langle (v - \langle v \rangle)^2 \rangle}$$

Where $\langle v \rangle$ is the mean velocity of the stars in the core.

In the context of globular clusters, the central velocity dispersion holds a particular importance because it is directly related to the total mass of the cluster. By utilizing the virial theorem and making some assumptions about the mass distribution, one can estimate the total mass M of the cluster from its velocity dispersion. Also, σ_0 can provide insights into the potential presence of a central massive object, like a black hole, in the cluster. A higher-than-expected velocity dispersion might hint at the presence of such an object. Not only, it plays an essential role in determining the rate of various dynamical processes, such as two-body relaxation and core collapse. In order to train the model we used a "mass-weighted" velocity dispersion The mass-weighted central velocity dispersion, typically denoted $\sigma_{0,\mathrm{mw}}$, takes into account the individual masses of stars when computing the average velocity dispersion. Instead of treating every star's velocity contribution equally, stars with greater masses have a larger influence on the overall velocity dispersion value.

Mathematically, the mass-weighted velocity dispersion is defined as:

$$\sigma_{0,\mathrm{mw}} = \sqrt{\frac{\sum_{i=1}^{N} m_i (v_i - \langle v \rangle_{\mathrm{mw}})^2}{\sum_{i=1}^{N} m_i}}$$

Where m_i is the mass of the i^{th} star, v_i is the velocity of the i^{th} star and $\langle v \rangle_{\mathrm{mw}}$ is the mass-weighted mean velocity, given by $\langle v \rangle_{\mathrm{mw}} = \frac{\sum_{i=1}^{N} m_i v_i}{\sum_{i=1}^{N} m_i}$.

In the context of GCs, a mass-weighted velocity dispersion can provide a more refined view of the cluster's dynamics, especially if there's a significant spread in stellar masses. This parameter is particularly relevant when considering the influence of massive stars or stellar remnants (like white dwarfs, neutron stars, or black holes) on the overall dynamics of the cluster. The mass-weighted approach ensures that these massive objects are properly accounted for when assessing the kinetic energy or dynamical state of the cluster.

Half-Mass relaxation time: The half-mass relaxation time (t_{rh}) in globular clusters is a measure of the time it takes for stars within the half-mass radius to establish a statistical equilibrium through two-body interactions. This concept emerges from the virial theorem, which, in the context of self-gravitating systems like globular clusters, relates the kinetic energy T and the potential energy W:

$$2T + W = 0$$

Two-body encounters cause stars to exchange energy, leading to a spread in velocities over time. This process drives the system toward energy equipartition, where more massive stars sink to the cluster center while lighter stars gain energy and move outward. The half-mass relaxation time is given by:

$$t_{rh} \approx \frac{0.138N}{\ln(N)} \times \frac{R_h^{3/2}}{\sqrt{GM_h}}$$

Where N is the number of stars, R_h is the half-mass radius (the radius within which half of the total mass of the system is contained), M_h is the total mass within R_h, and G is the gravitational constant. t_{rh} is crucial in globular cluster dynamics because it determines the rate of core collapse, binary formation, and other key evolutionary processes.

King profile: The King profile is a mathematical empirical function introduced by King (1962) to describe the surface brightness profiles of star clusters. The King profile is given by:

$$I(r) = k\left[\left(1+\frac{r^2}{r_c^2}\right)^{-\frac{1}{2}} - \left(1+\frac{r_t^2}{r_c^2}\right)^{-\frac{1}{2}}\right]^2$$

Where:

- $I(r)$ is the surface brightness at a radial distance r from the cluster center.
- r_c is the core radius (sometimes R_c below).
- r_t is the tidal (or truncation) radius.
- k is a normalization constant.

The core radius, r_c, characterizes the central, dense region of the cluster where the star density or surface brightness is relatively flat. Beyond the core radius, the density or brightness drops off more rapidly. The tidal radius r_t defines the outer boundary of a globular cluster. Beyond this radius, stars are unbound from the cluster due to gravitational tidal forces from the host galaxy. Essentially, it marks the limit where the external gravitational field of the host galaxy begins to dominate over the cluster's internal gravity, stripping stars away from the cluster. We can combine these two parameters defining a concentration parameter c as:

$$c = \log\left(\frac{r_t}{r_c}\right)$$

Effectively, to determine the core radius r_c for a specific globular cluster, we should fit the observed surface brightness or star density profile of that cluster to the King profile and then derive r_c (and other parameters) from the fit.

Finally, we can derive the half-light radius for a King profile: if we integrate the profile from the center outwards and find the radius at which the integrated luminosity is half the total luminosity.

3.4 Data

We used simulations for training, validation and testing from the MOCCA-Survey simulations and for deployment and generalization 161 Milky Way GCs from Baumgardt and Hilker (2018).

3.4.1 Training, validation, and test datasets

To construct a training and validation sample for our ML models, and to assess their performance, we utilized findings from the MOCCA-Survey Database I simulations, described in detail by Askar et al. (2017b). (however, also refer to Arca Sedda et al., 2018, Askar et al., 2017a, Belloni et al., 2016, 2017a,b, 2018, Samsing et al., 2017). The MOCCA-Survey Database encompasses almost 2000 star cluster simulations with diverse initial parameters that were conducted using the MOCCA code (Giersz et al., 2013, Hypki and Giersz, 2013), which relies on the Monte Carlo algorithm (see the seminal papers by Hénon, 1971, Stodolkiewicz, 1982, 1986) to model the long-term evolution of star clusters. The orbit-averaged Monte Carlo method combines a statistical approach to address

distant two-body interactions that drive the dynamical evolution of a star cluster with the particle-based approach of N-body methods. This technique enables the simulation of star cluster models containing numerous stars, ranging from hundreds of thousands to millions, within a few days to weeks. MOCCA incorporates prescriptions for stellar and binary evolution based on the SSE/BSE codes (Hurley et al., 2000, 2002). To compute the outcome of close dynamical interactions between binary-single stars and binary-binary stars, MOCCA employs the FEWBODY code (Fregeau et al., 2004), a direct N-body integrator for small-N gravitational dynamics. MOCCA also implements a realistic treatment of escape processes in tidally limited clusters, following the approach by Fukushige and Heggie (2000). In order to model the Galactic potential, MOCCA adopts a straightforward point-mass approximation. The handling of escapers from the tidally limited GC models is based on Fukushige and Heggie (2000). MOCCA has been thoroughly tested and compared with results from direct N-body codes (see, for instance, Giersz et al., 2013, 2019, Heggie, 2014, Madrid et al., 2017, Wang et al., 2016).

The star clusters models simulated in the MOCCA-Survey Database I encompass a broad range of initial parameters with varying number of objects, metallicity, binary fraction and parameter distribution, central concentration, tidal and half-mass radii, as well as different natal kick prescriptions for stellar-mass BHs (see Table 1 in Askar et al., 2017b). Initial stellar masses in each cluster model were sampled using the Kroupa (2001) initial mass function with minimum and maximum stellar masses of 0.08 $M_\odot$ and 100.0 $M_\odot$. The most densely populated ($\rho_c > 10^6$ $M_\odot pc^{-3}$) GC models form an IMBH (≥ 500 $M_\odot$) within a few hundred Myr of dynamical evolution through collisions, mergers, and mass transfer onto a seed BH that typically forms from a merger between a BH and a very massive main-sequence star (> 50 $M_\odot$). The latter form via the runaway merger of stars in the early evolution of these clusters.

In other less dense models (10^4 $M_\odot pc^{-3} \lesssim \rho_c \lesssim 10^6$ $M_\odot pc^{-3}$), the IMBH forms after a few Gyr of dynamical evolution from mergers during binary interactions involving a stellar-mass BH. Both the *fast* and *slow* IMBH formation mechanisms in the MOCCA GC models have been discussed extensively in Giersz et al. (2015). In most of these simulated cluster models, for mergers involving a BH and a star, it is assumed that the BH accretes the entirety of the mass of the star that it merges with (Askar et al., 2021, Hong et al., 2020). Additionally, gravitational wave recoil kick following the merger of two stars is also not implemented (Maliszewski et al., 2022, Morawski et al., 2018). Both these assumptions facilitate the formation and growth of IMBH in these cluster models. It is good to know that our knowledge of real star clusters is not perfect: recent evidence suggests that exploring a wide range of mass-to-light ratios (see Fig. 3.3) may indeed be a feature rather than a bug, given the possibility of, for example, hypercompact star clusters (Greene et al., 2021). The crucial safeguard is to take into account and appropriately mitigate the potential lack of realism of the simulations used to train a ML model, as discussed in Sec. 3.4.2 and 3.5.4.

From the 1298 GC models that evolved up to at least 12 Gyr in the MOCCA-Survey Database I, 388 hosted an IMBH more massive than 500 $M_\odot$; in the following, we adopted a more inclusive definition of IMBH, extending the range to 100 $M_\odot$ and above: this choice makes the classification problem more challenging and therefore more conservative. In the end, 30% of the clusters have an IMBH. Similar to the approach taken in Askar et al. (2019), we used the simulation snapshot at 12 Gyr for each GC model to determine the following features: Spitzer Jr (2014) half-mass relaxation time corresponding to the half-mass radius T_r (HRT), total luminosity of the cluster in units of

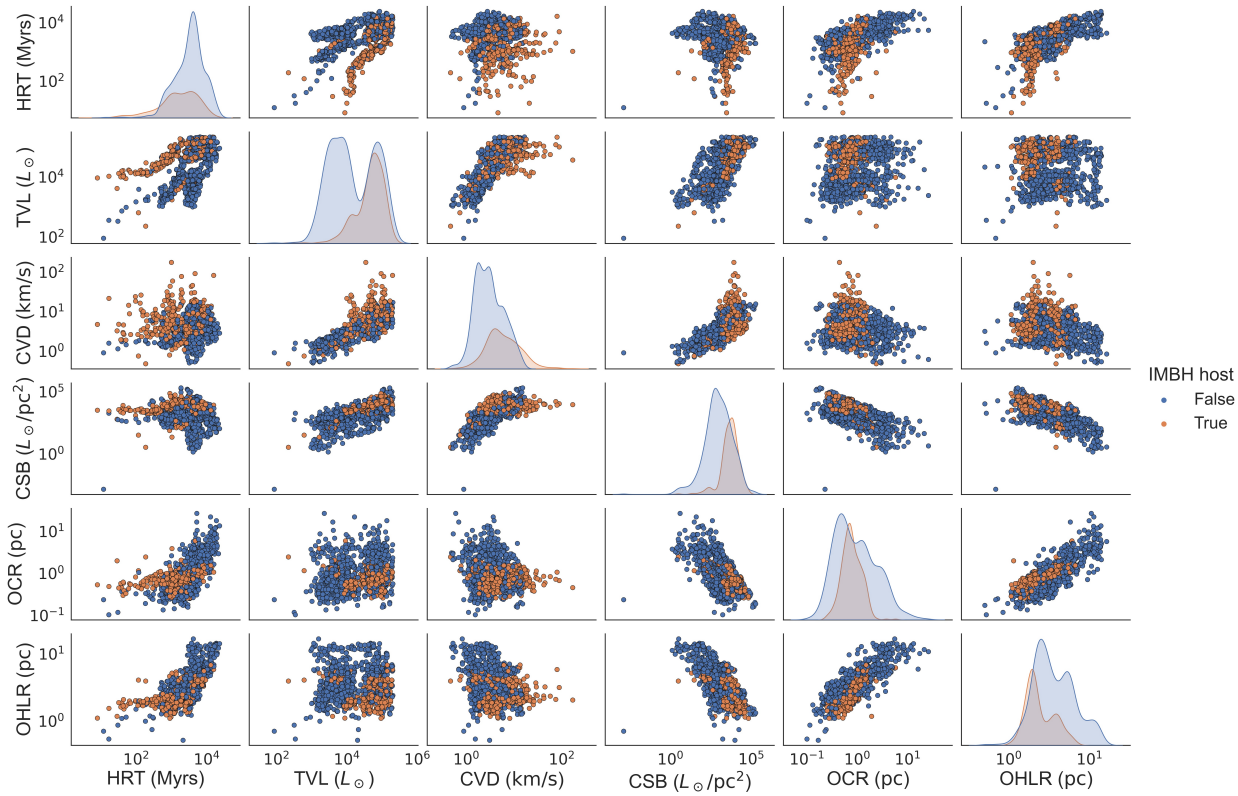


Figure 3.2: Pair plot showing our star cluster simulations in feature space. IMBH hosts are shown as orange points and non-hosts as blue points. In order, the axis are: Half-mass relaxation time (HRT), total luminosity (TVL), central velocity dispersion (CVD), central surface brightness (CSB), Core radius (OCR) and finally half-light radius (OHLR). All the reported values are at 12 Gyr.

solar luminosity L (TVL), mass-weighted central velocity dispersion in units of km/s (CVD), central surface brightness $I_0 = I(0)$ in units of $L_\odot/pc^2$ from King (1962) fitting of the cumulative luminosity profile $I(r)$ (CSB), core radius R_c in pc^2 obtained from King (1962) (OCR), and Half-light radius R_h obtained from King (1962) fitting in pc (OHLR).

The physical meaning of these features is explained in more depth in Sec 3.3.

The simulation data was partitioned into a train-validation dataset, where we used k-fold cross-validation as described below, and an internal test dataset comprising 10% of the initial sample. This was stratified with respect to the label, i.e. each subset was forced to contain approximately the same proportion of IMBH hosts as the whole sample (see Fig. 1.5). Why we need to cross validate, stratify our dataset is explained more in Sec. 1.2.9 and in Sec. 1.2.10.

The features distribution divided between IMBH host and IMBH non-host is shown in Fig.3.2

3.4.2 Deployment Dataset and Data Drift considerations

In ML, the representativeness of a test set to real-world data is crucial for model validity. Within high-dimensional spaces, this evaluation hinges on understanding interpolation and extrapolation. Interpolation involves making predictions within the range of the training data, relying on known patterns to fill gaps. Extrapolation, conversely, projects beyond this known range, making it inherently more uncertain in higher dimensions. The challenge intensifies as the dimensionality increases, with data sparsity becoming prevalent.

Training on simulations and testing on real-world data further amplifies challenges, as simula-

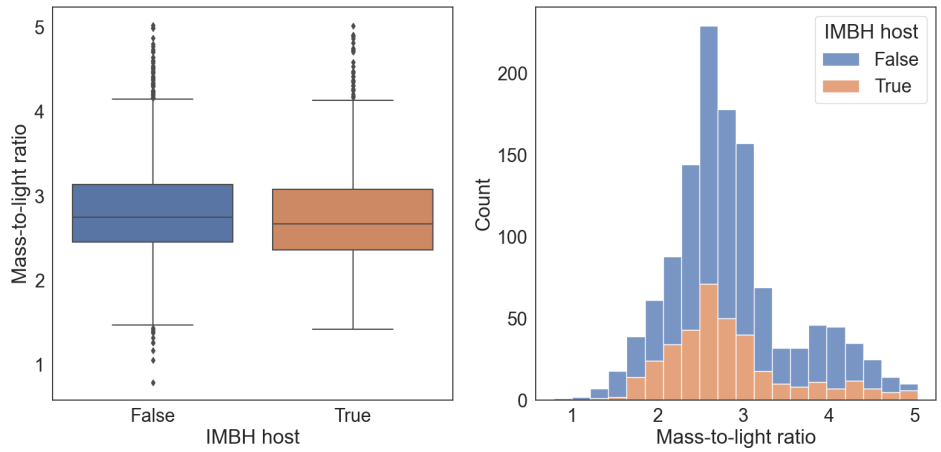


Figure 3.3: Mass-to-light ratio of the simulated clusters color-coded (IMBH hosts are shown in orange and non-hosts in blue)

tions might not capture all complexities inherent in empirical observations (Chattopadhyay and Chattopadhyay, 2014). Additionally, ML, while powerful, is not infallible; models might overfit to specific patterns in training data or falter when encountering novel scenarios, underscoring the importance of robust validation and testing practices (Ng, 2016).

After calibrating our models with simulated data, we applied them to detect IMBHs in real star clusters. We drew upon the observational data from Baumgardt and Hilker (2018), incorporating features from 161 Milky Way globular clusters. The dataset is accessible at `https://people.smp.uq.edu.au/HolgerBaumgardt/globular/parameter.html`. The relevant features of the deployment dataset for this work are shown in comparison with the test set in Fig. 3.8 and in the top panel of Fig. 3.9.

It is essential to note that the catalogue is regularly updated due to improved observations. As a result, the application of the same methods to the same catalogue may yield different outcomes. This data drift issue is a common challenge in real-world ML applications. For a detailed discussion on managing model deployments in the face of changing data, we refer readers to the comprehensive review by Lu et al. (2018). The last data extraction was carried out in April 2023.

In summary, we used simulated data from the MOCCA-Survey I for both our training (with cross-validation) and testing sets. For the deployment phase, however, we switched to using real-world data.

3.5 Methods

I present here, some of the methods applied in our analysis with a focus on the black box model (XGBoost) with its post hoc explanations (Anchors) and the white box (CORELs). Finally, I provide insight on Kernel Density Estimation (KDE).

3.5.1 XGBoost Classifier

XGBoost (Chen and Guestrin, 2016) is a ensemble machine learning algorithm that builds upon the principles of gradient boosting. The idea behind gradient boosting is to iteratively add trees to the ensemble where each new tree tries to correct the errors made by the existing ensemble of trees.

Mathematically, if $f(x)$ is the prediction of the ensemble, and y is the true value, the algorithm tries to add a tree $h(x)$ to minimize the loss function L:

$$L(y, f(x) + h(x))$$

The objective function that XGBoost aims to minimize consists of a loss function and a regularization term. The loss function L measures the difference between the predicted values and the actual values. The regularization term Ω penalizes the complexity of the model to prevent overfitting:

$$Obj = \sum_{i=1}^{m} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where m is the number of samples, K is the number of trees, y_i is the true value for the i-th sample, $\hat{y}_i$ is the predicted value for the i-th sample, and f_k is the k-th tree. The regularization term Ω is defined as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2$$

where T is the number of leaves in the tree, w are the leaf weights, γ is a parameter that controls the penalty for the number of leaves, and λ is a parameter that controls the penalty on the leaf weights. In XGBoost, each f_k represents an individual decision tree that is part of the ensemble. Each tree is a series of decision rules based on the input features that split the data into branches, leading to the final predictions at the leaves (T). The leaves of a tree are the end nodes where predictions are made. T represents the number of leaves in a given tree. The structure of the tree is such that it partitions the input space into T disjoint regions, and each region corresponds to a leaf. The prediction for any instance that falls into a given region is determined by the value associated with that leaf. The leaf weights, denoted before by w, are the values at the leaves that provide the final prediction for the data points that end up in the corresponding leaf after passing through the decision rules of the tree. In classification tasks, it can be transformed into a probability for a certain class. The leaf weight is learned during the training process, and it's the output that minimizes the loss function for the data points in that leaf.

XGBoost uses a gradient-based optimization algorithm to minimize the objective function. The model is trained iteratively, and in each iteration, a tree is added to the ensemble to minimize the objective function.

The gradient and hessian of the loss function with respect to the predictions are calculated, and these are used to determine the optimal weights for the leaves of the new tree. The tree structure is determined by greedily adding splits that result in the maximum reduction in the objective function.

Lack of Interpretability: The strength of XGBoost in capturing complex patterns and making accurate predictions comes at the cost of interpretability. The ensemble of multiple decision trees, each correcting the previous ones' errors, results in a highly complex model. This complexity is further compounded by the algorithm's ability to capture intricate, non-linear relationships and interactions between features in high-dimensional space. Unlike simpler models such as a single decision tree, where one can trace a clear decision path from the root to a leaf, XGBoost's ensemble nature obscures such paths, making it challenging to understand how input features collectively

influence the output. The mathematical intricacies of gradient boosting and regularization add another layer of complexity, making the model's decision-making process less intuitive. While post hoc interpretation techniques like SHAP and Anchors (see next section) can provide insights into individual predictions, they fall short of offering a comprehensive global understanding of the model's behavior, leaving XGBoost as a powerful yet opaque tool in ML[2].

3.5.2 Anchors

In the realm of XAI, a crucial metric emerges: human precision. This refers to the proportion of instances wherein human analysts can accurately predict a model's behavior. Human precision quantifies how often humans are correct in their predictions about a model's outcomes. The significance of high human precision cannot be understated; it becomes a cornerstone for genuine explainability. To genuinely claim an understanding of a model, it is imperative that when humans believe they grasp its behavior, they are predominantly correct in their assertions. This is the core value of explanations in the form of if-then rules named anchors presented by Ribeiro et al. (2018).

These anchors provide high-precision, model-agnostic explanations, designed to *anchor* predictions locally and ensure stability in the prediction despite changes in other feature values. This approach addresses the limitations of existing local explanation methods, particularly the often unclear coverage, which can lead to low precision from the users' perspective.

About local vs global, model-agnostic vs model specific:

- Local explanations aim to provide insight into the model's behavior for an individual prediction or a small group of instances. They answer the question, "Why did the model make a specific prediction for this instance? Global explanations, instead, provide an overall understanding of the model's behavior across the entire dataset. They seek to answer, "How does the model behave in general?" Local explanations are particularly useful when you need to understand critical decisions on a case-by-case basis instead global explanations are particularly useful when we want to understand the general mechanisms in the data or debug a model. The emphasis on local explanations is not merely a matter of preference but a substantive requirement in the pursuit of good interpretability. It is a safeguard against the fallacy of believing that one understands the model comprehensively when, in fact, the understanding is cursory at best. This is so important, that even the European Union's General Data Protection Regulation and the AI act (Hacker and Passoth, 2020), which underscores the necessity for transparency, in a section asks for explanations at the individual decision level.

- Model-agnostic vs model-specific explanations: Model-agnostic explanations are XAI techniques that can be applied to any machine learning model, regardless of its internal workings. They treat the model as a black box and do not require access to its internal structure. On the contrary, model-specific explanatins are designed to explain a particular type of model and rely on the internal structure and parameters of the model. They can often provide more detailed insights but are limited to specific model types.

[2]In Appendix C, I address the fact that even using a full explainer dashboard to interact with the model, XGBoost lacks human understandability.

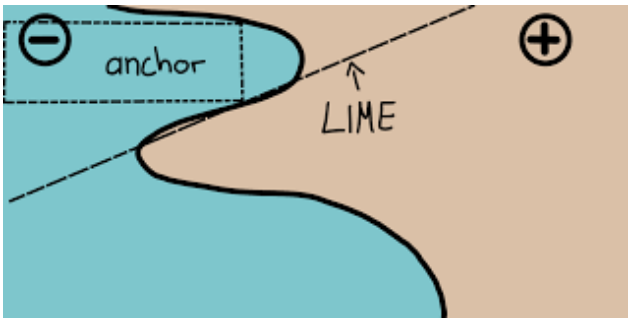


Figure 3.4: LIME vs. Anchors: A Toy Visualization LIME solely learns a linear decision boundary and produces local surrogate models. LIME has a problem of instability of the explanations in addition to the difficult to define the neighbourhood of an explanation. Figure from Ribeiro et al. (2018).

Let's imagine that for a given instance x and its associated black-box prediction $f(x)$, the objective of an anchor is to identify a if-then rule A. This rule, grounded in the feature space of x, aims to encapsulate the reasoning behind $f(x)$. The formal criterion for an anchor A is such that, within a localized perturbation region $D(x)$ x:

$$\mathbb{P}(f(x') = f(x), |, A(x') = 1, x' \in D(x)) \geq \tau \tag{3.1}$$

Here, x' represents instances perturbed from x. The user-defined precision threshold τ ensures that the rule A is robust and reliable within the local region $D(x)$.

The authors emphasize the users' preference for a comprehensive set of explanations that cover most of the model with minimal effort required on their part. To address this, they propose a submodular procedure to select an optimal set of explanations, ensuring that the chosen explanations provide a broad understanding of the model's behavior.

Interesting enough, a user study involving 26 participants with a background in machine learning was conducted to validate the effectiveness of anchors. Utilizing two datasets and a multiple-choice Visual Question Answering (Antol et al., 2015) system on two images, participants were tasked with predicting the behavior of the classifier on unseen instances, both before and after being provided with explanations. The results from this study indicated that anchors not only provide better coverage but also enhance the accuracy of predictions made by users, outperforming random explanations. Also anchors are model-agnostic, meaning they can be applied to any ML model without requiring access to its internal workings. This is a significant advantage, as it provides flexibility and applicability to a wide range of models. However, model-agnostic methods might not be able to leverage specific characteristics of certain models, potentially leading to less accurate or less informative explanations compared to model-specific methods.

While anchors provide a robust and explanations, they are not without their challenges and limitations: anchors can sometimes be too specific, especially for predictions near the decision boundary of a model or for very rare classes. This specificity can lead to complex and low-coverage explanations, which might not generalize well to other instances. For example, a very specific anchor might communicate the conditions necessary for a particular prediction accurately, but its applicability might be limited due to its narrowness. Also in practical applications, there might be situations where two or more anchors with different predictions apply to the same test instance. The unlikelihood of this is addressed in the paper due to the high-probability precision guarantee of

anchors and the submodular objective in the selection procedure, but it is a possibility that needs to be considered when we apply these model explanations to our data.

3.5.3 CORELS

I present here, the white box model used in the work. The primary goal of CORELS (Angelino et al., 2017) is to find a decision list, d, that optimizes an objective function, represented as:

$$R(d, x, y) = \ell(d, x, y) + \lambda K$$

The decision list is defined by $d = (d_p, \delta_p, q_0, K)$ where:

- d_p represents the *predicates* of the decision list. Predicates are essentially the conditions or rules that are checked in sequence. In the context of the CORELS algorithm, a predicate can be thought of as a simple "if" condition.
- δ_p These are the *decisions* or predictions associated with each predicate. For every predicate d_p, there's a corresponding decision in δ_p.
- q_0 This is the *default prediction*. It is the prediction that's made if none of the predicates in d_p are satisfied.
- K This represents the *length* of the decision list, i.e., the number of predicates in the list. In the context of interpretability, a shorter list (smaller K) is usually preferable because it is easier to understand and explain.
- $\ell(d, x, y)$ is the loss function. It calculates the difference between the predictions made by the decision list d and the true labels y for the given training data x. The specific form of the loss function can vary depending on the problem. For classification problems, this could be a simple count of misclassifications.
- λ is a regularization parameter.

CORELS employs a branch-and-bound strategy to search for the optimal decision list. Instead of searching through all possible decision lists, which would be computationally prohibitive, the algorithm leverages the structure of the problem to explore the space efficiently. It starts with an empty list and builds on it, branching on possible rules to add. By bounding the best possible score that can be achieved by extensions of a current partial rule list, CORELS can discard many potential candidates without explicitly evaluating them, significantly speeding up the search. In depth: for each prefix (partial rule list), the algorithm computes both a lower and upper bound on the objective value of the best possible completion of that prefix. If the lower bound of a prefix exceeds the upper bound of the current best solution, the prefix and its extensions can be safely discarded. These bounds are computed using optimistic and pessimistic estimates of the loss and the number of rules in the completed list.

To ensure the algorithm doesn't waste time exploring rule lists that are essentially the same, CORELS uses a "symmetry-aware map." This data structure stores the best objective value known for any rule list that has the same set of predicates, regardless of their order. If the algorithm

encounters a rule list with a set of predicates it has seen before and the new list's objective value is not better than the stored value, the algorithm can disregard this new list.

One of the key contributions of CORELS is its ability to provide an optimality certificate. When the algorithm concludes, this certificate verifies that the decision list found is indeed optimal for the given objective function. The optimality is confirmed when the objective value of the best complete rule list is less than or equal to the lower bound on any remaining prefixes in the priority queue. Down below, a schematic view of how CORELs algorithm works: Where in this pseudo code:

Algorithm 1 CORELs

Input: Dataset D, Binarized features F, Labels L
Output: Optimal decision list M^*
procedure CORELS(D, F, L)
  $R \leftarrow$ GenerateCandidateRules(F, L) {Generate all possible rules}
  $M^* \leftarrow \emptyset$ {Initialize best model as empty}
  $V^* \leftarrow$ Evaluate(M^*, D, L) {Evaluate the initial model}
  $Q \leftarrow$ InitializeQueue() {Initialize priority queue}
  Q.enqueue(M^*)
  while Q is not empty **do**
    $M \leftarrow Q$.dequeue()
    $V \leftarrow$ Evaluate(M, D, L)
    if $V < V^*$ **then**
      $M^* \leftarrow M$
      $V^* \leftarrow V$
    end if
    for each rule $r \in R$ **do**
      $M' \leftarrow M \cup \{r\}$ {Add rule to model}
      $V' \leftarrow$ Bound(M', D, L) {Calculate bound}
      if $V' < V^*$ **then**
        Q.enqueue(M')
      end if
    end for
  end while
  $M^* \leftarrow$ Simplify(M^*, D, L) {Simplify the model}
  return M^*
end procedure=0

- GenerateCandidateRules generates all possible rules based on the binarized features.
- Evaluate evaluates the performance of a decision list on the dataset.
- InitializeQueue initializes a priority queue to store models to be evaluated.
- Bound calculates a bound on the best objective value that can be achieved with a given model. Simplify simplifies the decision list if possible without hurting its performance.

Effectively, when we implemented CORELS we had to specify also these parameters:

1. **max_card**:Refers to the maximum cardinality or length of the rules generated. In simple terms, this parameter sets the maximum number of conditions that a rule can have.

For example, if `max_card=2`, the rules generated can have at most two conditions, like `(Condition1 AND Condition2) => Outcome`.

2. **c**: This is the regularization parameter λ. It is a non-negative value that balances the rule's complexity with its accuracy. The higher the value of `c`, the more the algorithm penalizes rule length, favoring shorter, more interpretable rules at the possible expense of accuracy. For instance, if you have two rules with similar accuracy, the one with fewer conditions might be preferred if `c` is high, as the algorithm will try to minimize the number of conditions used.

3. **min_support**: Specifies the minimum fraction of the training data a rule must cover to be considered. If a rule only applies to a tiny fraction of the data (less than `min_support`), then it is discarded. This parameter ensures that the generated rules are relevant to a reasonable portion of the data, filtering out rules that might only apply to very specific or rare cases.

CORELS shines in scenarios demanding transparency in decision-making. We are happy to introduce it in astrophysics for the first time.

3.5.4 Kernel Density Estimation

Kernel Density Estimation (KDE) stands as a non-parametric approach to approximate the probability density function of a random variable. Unlike the simplicity of histograms, KDE overcomes issues of binning choice and discontinuities, delivering a smooth estimate.

At its core, KDE sums up the contributions from a kernel function K centered at each data point. The estimator for a univariate dataset $x_1, x_2, ...x_n$ is given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3.2}$$

where h is the bandwidth, a crucial parameter determining the smoothness of the density estimate. The kernel function K satisfies:

$$\int K(u)du = 1 \quad \text{and} \quad K(u) \approx 0 \text{ for } |u| \to \infty \tag{3.3}$$

The bandwidth h strikes a balance between bias and variance: a smaller h captures intricate details but is noisy, while a larger h is smoother but may obscure genuine features. It can be optimized through cross-validation or heuristic methods, like Silverman's rule:

$$h = \left(\frac{4\sigma^5}{3n}\right)^{\frac{1}{5}} \tag{3.4}$$

where σ is the standard deviation of the data.

KDE, while robust in the univariate setting, encounters challenges as we venture into higher-dimensional spaces.

For a dataset in d-dimensions, the kernel function K also becomes d-dimensional, and we need a $d \times d$ bandwidth matrix H (as opposed to a single bandwidth value h in the univariate case). This bandwidth matrix helps in controlling the smoothness of the density estimate in every dimension.

Table 3.1: Rules found by CORELS during training in 5-fold cross-validation. The first column reports the number of folds in which a given rule was found. The corresponding rule is reported in the second column.

Number of folds	Rule
4	$T_r < 2000$ and $R_h < 3.0$ and both $L > 10000$ and $R_c/R_h > 0.2$
1	$T_r < 5000$ and $I_0 > 1000$ and $L > 10000$ and $R_c/R_h > 0.2$

The estimator for a d-dimensional dataset $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n$ (where each $\mathbf{x}_i$ is a d-dimensional vector) is:

$$\hat{f}(\mathbf{x}) = \frac{1}{n\sqrt{|H|}} \sum_{i=1}^{n} K\left(H^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right)$$

where $|H|$ is the determinant of the matrix H, $H^{-1/2}$ is a matrix square root (the concept analogous to the square root for scalars, but applied to matrices) and $K\left(H^{-1/2}(\mathbf{x} - \mathbf{x}_i)\right)$ is the multivariate kernel function.

A big problem raises here known as the "curse of dimensionality". Essentially, as the number of dimensions increases, the volume of the space increases so rapidly that the available data becomes sparse. This sparsity makes it difficult to get a good density estimate because the data points are too spread out. The bandwidth matrix H can help to some extent, but it also introduces complexities because it now has to be a $d \times d$ matrix, which is harder to optimize than a single bandwidth value h.

The core idea behind multivariate KDE is similar to the univariate case, but the complexities arise from the need to deal with multiple dimensions simultaneously. The kernel function and bandwidth need to be adapted to this multivariate setting, introducing more complexity and challenges, especially as the number of dimensions grows.

3.6 Results

We trained CORELS with the following parameters: `max_card`=4, `c` = 0.04, `min_support`=0.01 (for the description of these parameters see Sec. 3.5.3). CORELs does not work with smooth features but with binary features. As reference with histograms like Fig.3.5, we binned our feature space in conditions d_p: practically for every example x_i there are 72 features of the type: HRT$<$ 500, HRT$<$ 750 . . . , OHLR$<$ 5.0. Every examples x^i of our dataset fills of boolean values these binary features.

In Fig 3.6 we show the precision-recall curve of XGBoost on each of the five cross validation folds. On each fold we also trained a CORELS model. Both the performance of XGBoost and CORELS change across folds, but in the case of CORELS it is immediately evident *why*. The only fold in which CORELS learned a different rule from the others (corresponding to the blue dot in Fig 3.6) achieves a much greater recall at the cost of a lowered precision. We show below that the rule learned by CORELS in this case is less strict in weeding out star clusters that have a longer relaxation time.

It is also quite evident from Fig 3.6 that the folds on which XGBoost performs best (green and black solid lines respectively), see also a higher performance by CORELS, which achieves the best (green dot) and second best (black dot) precision in each of them.

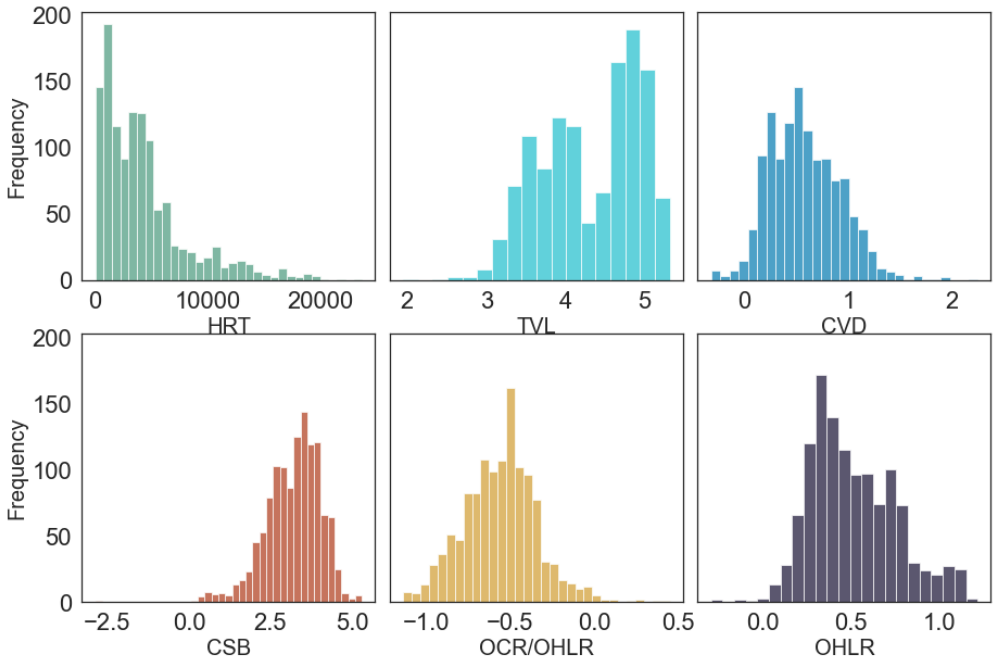


Figure 3.5: Histograms of features for the training dataset. The acronyms are: half-mass relaxation time (HRT), total luminosity (TVL), central velocity dispersion (CVD), central surface brightness (CSB), Core radius (OCR) and finally half-light radius (OHLR). All the reported values are at 12 Gyr. We used the percentile to create the CORELS features.

Table 3.2: Sub-rules for $T_r < 2000$ Myr and $R_h < 3.0$ pc and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$ and the corresponding precision-recall

Conditions	Precision	Recall
$T_r < 2000$ Myr	0.441558	0.521472
$R_h < 3.0$ pc	0.396161	0.696319
$L > 10^4 L_\odot$	0.417355	0.929448
$R_c/R_h > 0.2$	0.378641	0.957055
$T_r < 2000$ Myr and $R_h < 3.0$ pc	0.480597	0.493865
$T_r < 2000$ Myr and $L > 10^4 L_\odot$	0.802083	0.472393
$T_r < 2000$ Myr and $R_c/R_h > 0.2$	0.632184	0.506135
$R_h < 3.0$ pc and $L > 10^4 L_\odot$	0.558376	0.674847
$R_h < 3.0$ pc and $R_c/R_h > 0.2$	0.536946	0.668712
$L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.468498	0.889571
$T_r < 2000$ Myr and $R_h < 3.0$ pc and $L > 10^4 L_\odot$	0.814815	0.472393
$T_r < 2000$ Myr and $R_h < 3.0$ pc and $R_c/R_h > 0.2$	0.672414	0.478528
$T_r < 2000$ Myr and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.867052	0.460123
$R_h < 3.0$ pc and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.636637	0.650307
$T_r < 2000$ Myr and $R_h < 3.0$ pc and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.882353	0.460123

Table 3.3: Sub-rules forT_r < 5000 and I_0 > 1000 $L_\odot/pc^2$ and L > 10000 and R_c/R_h > 0.2 and the corresponding precision-recall

Conditions	Precision	Recall
$T_r < 2000$ Myr	0.338750	0.831288
$I_0 > 1000 L_\odot/pc^2$	0.442982	0.929448
$L > 10^4 L_\odot$	0.417355	0.929448
$R_c/R_h > 0.2$	0.378641	0.957055
$T_r < 2000$ Myr and $I_0 > 1000 L_\odot/pc^2$	0.485493	0.769939
$T_r < 2000$ Myr and $L > 10^4 L_\odot$	0.548458	0.763804
$T_r < 2000$ Myr and $R_c/R_h > 0.2$	0.474453	0.797546
$I_0 > 1000 L_\odot/pc^2$ and $L > 10^4 L_\odot$	0.505922	0.917178
$I_0 > 1000 L_\odot/pc^2$ and $R_c/R_h > 0.2$	0.567515	0.889571
$L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.468498	0.889571
$T_r < 2000$ Myr and $I_0 > 1000 L_\odot/pc^2$ and $L > 10^4 L_\odot$	0.582547	0.757669
$T_r < 2000$ Myr and $I_0 > 1000 L_\odot/pc^2$ and $R_c/R_h > 0.2$	0.654891	0.739264
$T_r < 2000$ Myr and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.635638	0.733129
$I_0 > 1000 L_\odot/pc^2$ and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.583333	0.880368
$T_r < 2000$ Myr and $I_0 > 1000 L_\odot/pc^2$ and $L > 10^4 L_\odot$ and $R_c/R_h > 0.2$	0.681948	0.730061

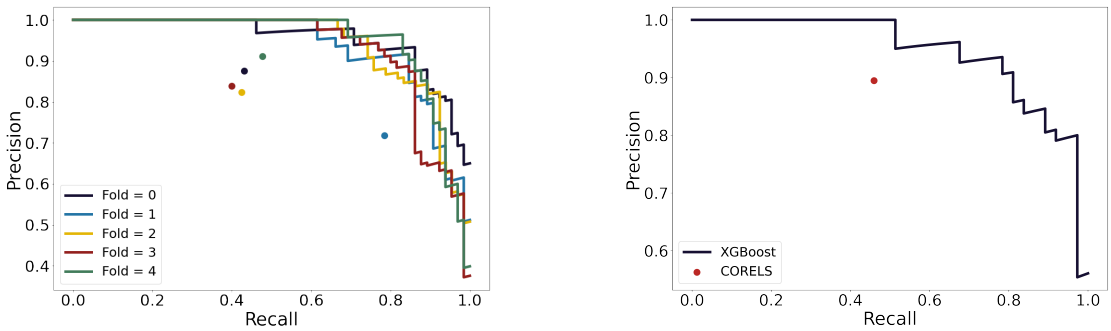


Figure 3.6: Left-hand panel: precision recall curve for our XGBoost classifier on each cross-validation fold (solid lines). The dots represent the recall and precision achieved by a CORELS classifier trained on the same fold. The colours identify the fold for both XGBoost and CORELS. Right-hand panel: Precision recall curve for our XGBoost classifier on the test set. The red dot represents the recall and precision achieved by the CORELS classifier on the same test set.

The rules found by CORELS are reported in Tab. 3.1 and shown graphically (in projection on relevant feature planes) in Fig 3.7. They were learned on one of the cross-validation folds each. Remarkably, an identical rule is learned on four folds out of five. The rule learned on the remaining fold is still quite similar to it. All rules stipulate that an IMBH candidate must have an inflated core, in particular requiring that $R_c/R_h > 0.2$. Let's remember what we have learn in the IMBH section: a large core is produced by dynamical heating due to a binary containing the IMBH. The IMBH has a high probability of ending up in a binary system in the first place, because its large mass ensures that exchange interactions are favoured in IMBH-binary encounters. In addition, mass segregation makes it likely that the secondary is a stellar-mass black hole. The resulting binary system hardens through interactions with the other objects in the core, releasing energy that ultimately results in a bigger core.

All rules also demand that the total luminosity of a potential candidate exceeds 10^4 $L_\odot$. This is a very low bar for GCs, whose typical luminosity is about one order of magnitude larger. This rule thus filters out star clusters that are likely too small to be able to assemble an IMBH, either because they do not have a high enough density to trigger runaway mergers or because their escape velocity is too small to confine dark remnants that may eventually merge to form an IMBH.

High density is also a criterion for all rules, but it is enforced in one fold by a direct cutoff in central surface brightness, and in the other four by a condition on the half mass radius, i.e. $R_h < 3.0$. As shown in 3.7 this combines with the luminosity constraint common to all rules to select high density star clusters.

Finally, all rules require that the relaxation time be short, but the cutoffs differ, being set to either 2 Gyr (in four folds out of five) or to 5 Gyr (in the remaining one). This is likely the reason for the higher recall in the corresponding fold, as shown in Fig. 3.6. A short initial relaxation time corresponds to higher interaction frequency between stars, which results in faster mass segregation and makes both runaway mergers and stellar-mass black hole coalescences more likely. Relaxation times observed at 12 Gyr are essentially a proxy for initial relaxation times, which are unobservable. Additionally, the swelling of the core due to an IMBH binary, is observable only if the system is relaxed, because the associated dynamical heating requires time to take place. Thus a swollen core in the absence of a short relaxation time is an unreliable indicator of IMBH presence.

As discussed before, if we look only at performances (without considering transparency, human readability, debug, domain-field applicability) there is always a trade-off between interpretability and performance as shown in Sec. 2.4 (**?**, among others)Sarkar et al., 2016, van der Veer et al., 2021, van der Veer et al., 2021 among others). But as the physical discussion before shown, we were able to interpret the results (even fold by fold) of CORELS easily; instead the results of XGBoost without explanations are only a product of a powerful model. We provide below explanations in the form of anchors (see Sec. 3.5.2) for XGBoost prediction on real GCs.

3.6.1 Kernel Density Estimate and real data

There is a potential mismatch between our simulated training data and actual star clusters, which can lead to biased predictions. As we can see in Fig. 3.8, the data on which our models were tested and the real clusters are not perfectly matched. It is thus crucial to assess how far out-of-distribution the star clusters we cast predictions for are with respect to the training data. A multitude of out-of-distribution detection approaches have been discussed in the ML literature (see Yang et al.,

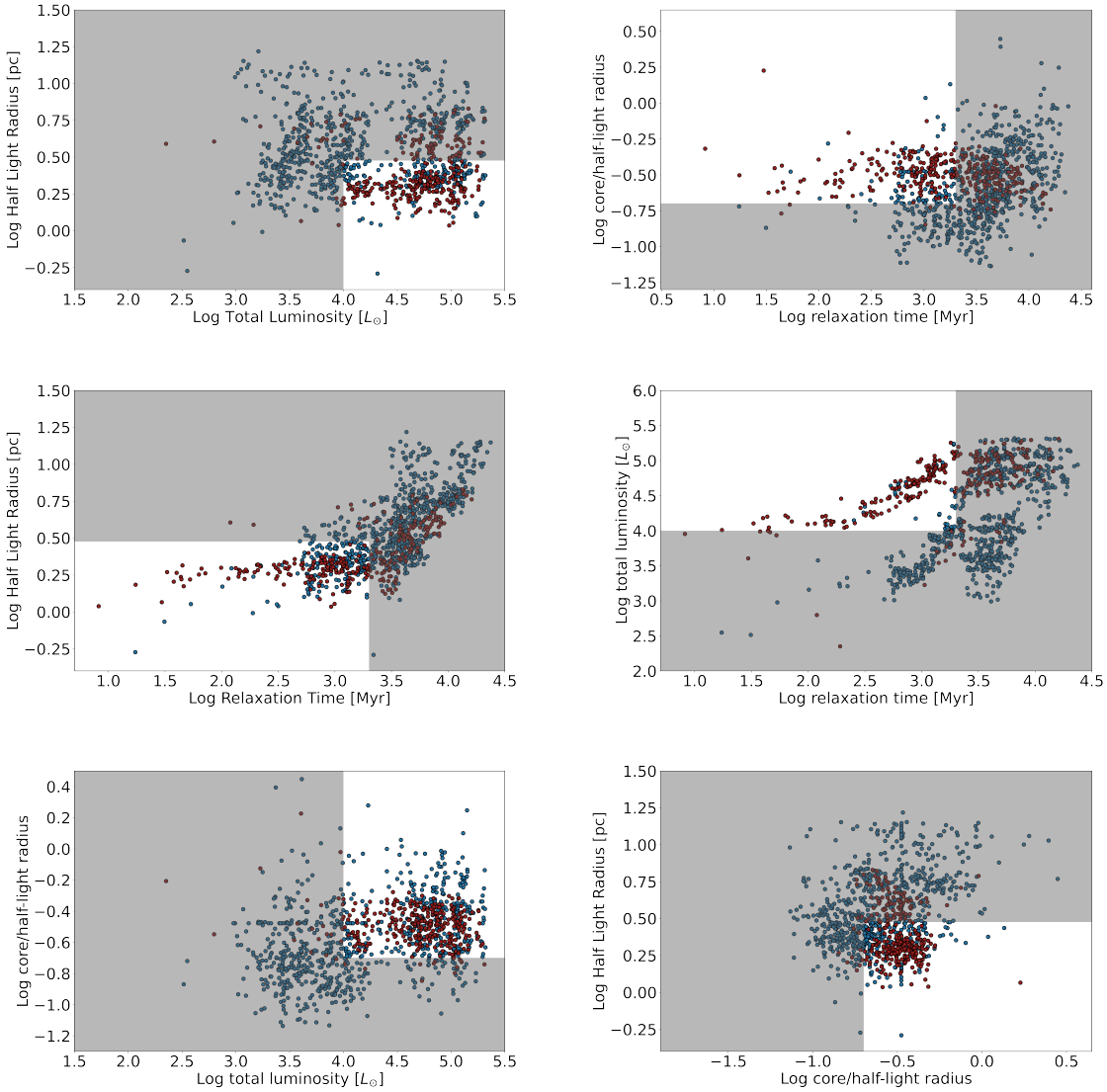


Figure 3.7: The rules found by CORELS, displayed on the planes corresponding each to the relevant couple of features. The white area corresponds to the region that respects the rule. The red points correspond to simulated star clusters in the training set that host an IMBH and the blue points correspond to non-hosts.

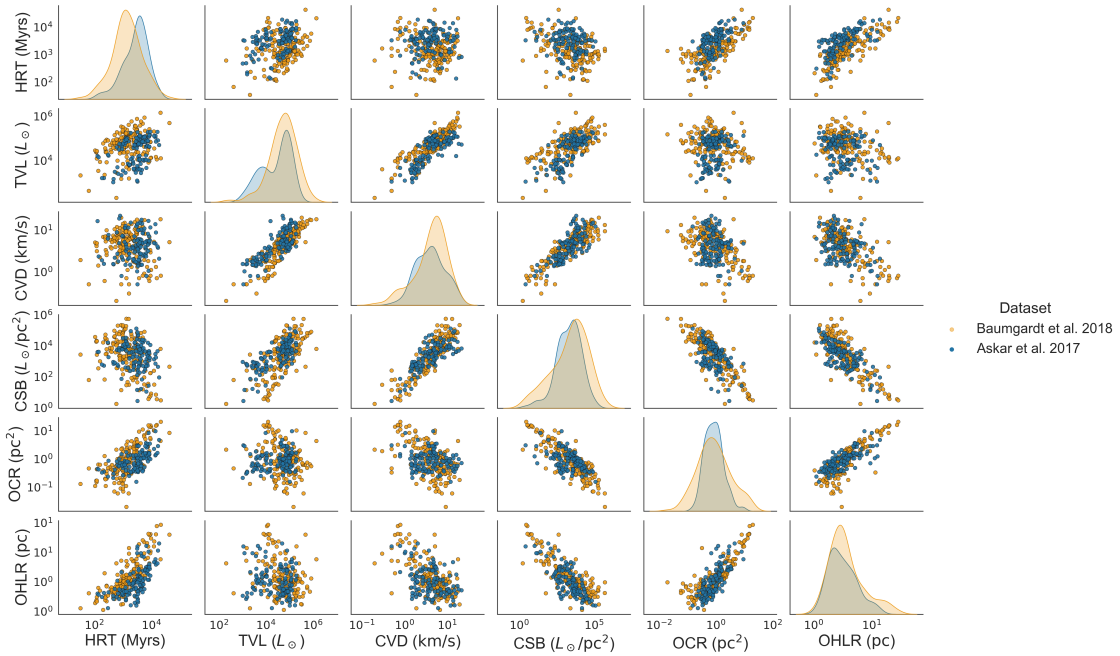


Figure 3.8: Pair plot showing test data (a representative sample of the simulations, shown in blue) versus real star cluster data (orange). In order, the axis are: Half relaxation time (HRT), total Luminosity (TVL), central velocity dispersion (CVD), central surface brightness (CSB), Core radius (OCR) and finally half-light radius (OHLR).

2021, for a recent review). In our case, we stuck to a simple density-based approach, which is generally recognized as effective in low-dimensional feature spaces. For this reason we utilized a Kernel Density estimation (KDE). KDE is a non-parametric technique for estimating the probability density function (PDF) of a random variable from independent, identically distributed samples. The basic idea behind KDE is to place a smooth, symmetric kernel function at each data point, and then sum these kernel functions to estimate the underlying PDF (to see more on KDE see Sec. 3.5.4). While other approaches to density estimation exits, such as e.g. Gaussian mixture models (McLachlan and Basford, 1988), mixture density networks (Bishop, 1994), and autoregressive flows (Huang et al., 2018), KDE is a simple and time tested method that works well in low dimension. We used a simple Gaussian kernel KDE relying on the scikit-learn package (Pedregosa et al., 2011). We fitted the density estimation on the training set and evaluated on both the test set and the real data set. The distributions of the KDE score are shown in Fig. 3.8. Low KDE score corresponds to weird -i.e. unlikely, out of distribution- data points with respect to the training set. We calculated the KDE scores for the points in our test set (randomly extracted from the simulation data set) and used its first decile as a cutoff to reject the out-of-distribution data points among the actual star clusters. This way we remove from the pool of IMBH host candidates those whose prediction would require us to trust our models in regions where the data on which they were trained is scarce.

Lastly, we present a table of GCs that are predicted to be IMBH hosts both by CORELS and by XGBoost and also are in-distribution based on our KDE criterion. This is shown in Tab. 3.4 alongside with the upper limits on IMBH mass obtained for each candidate by T18 (if any are present) and with any reference we found presenting evidence of IMBH presence or absence in the candidate. Candidate selection for this table is summarized by Figs. 3.10 and 3.11.

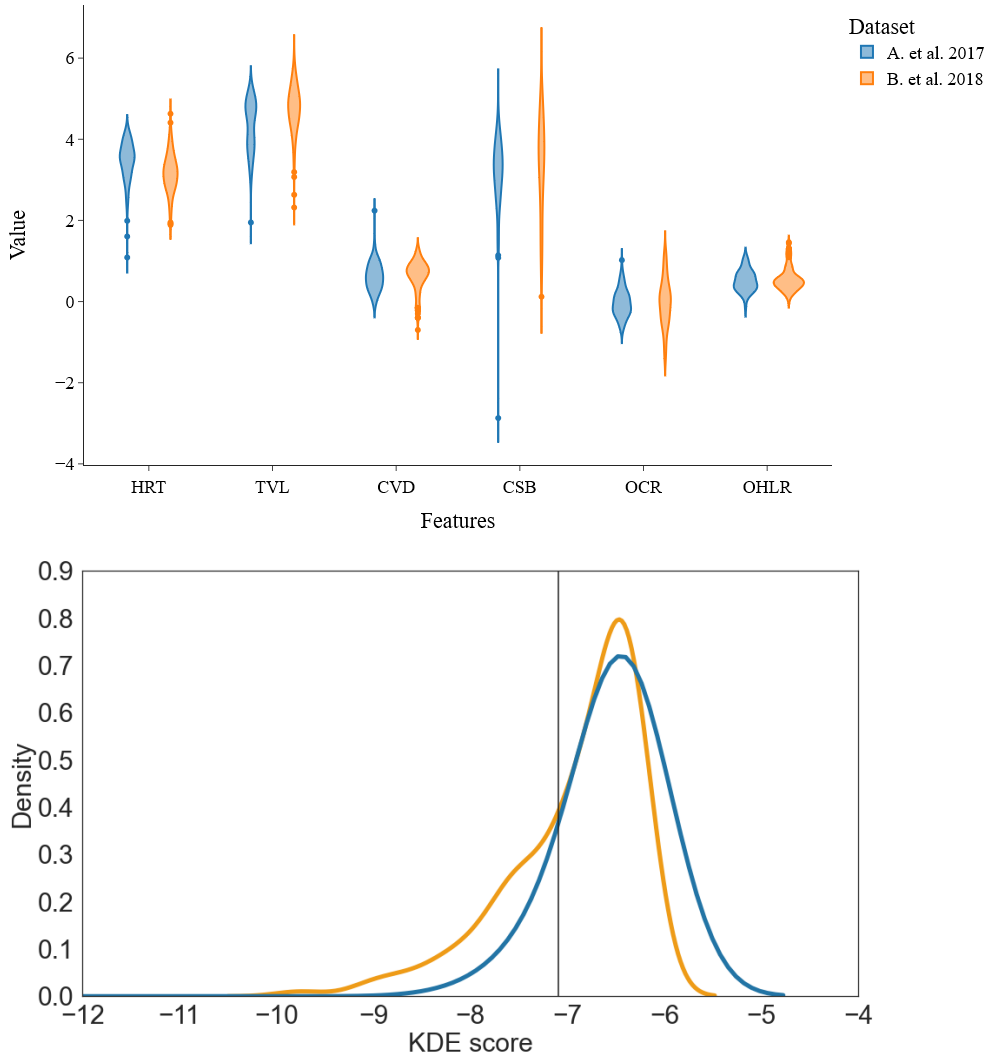


Figure 3.9: *Upper panel*: violin plot for the relevant features. We can see that the central surface brightness span more order of magnitudes.*Lower panel*: Distribution of the KDE score for test data (orange) and actual star cluster data (blue). The vertical line is the cutoff adopted for inclusion in the final sample of candidates, which corresponds to the first decile of the KDE score for test data.

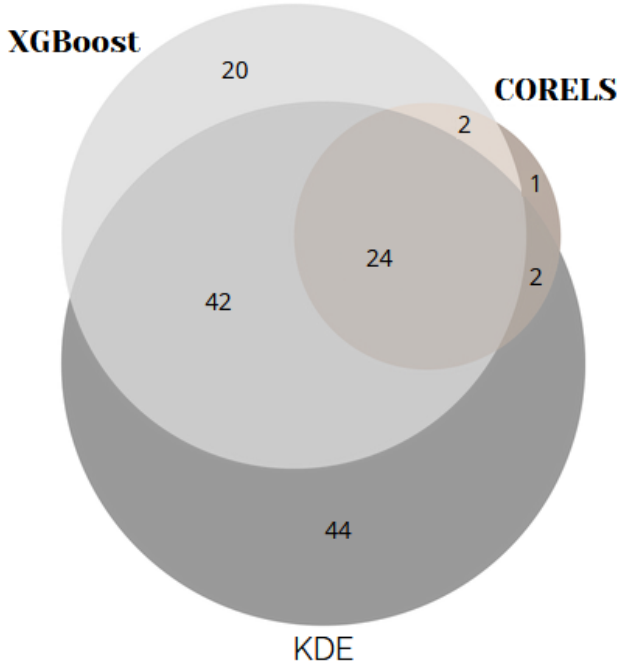


Figure 3.10: Euler-Venn diagram summarizing the inclusion rules for GCs in the final candidate list. The dark gray set includes all GCs who fall in-distribution with respect to the simulations according to the criterion based on KDE we discussed in the text (112 in total); the light gray set includes the XGBoost candidates (88 in total); and the brown set the CORELS candidates (29 in total). The clusters in the mutual intersection of all three sets are included in the final candidate list (24 in total).

Table 3.4: Final table of IMBH candidate hosts. The first column reports the cluster name, columns 3-8 report the (rounded, log) values of the relevant features. Columns 9-11 show the classification prediction by XGBoost, CORELS and the in-distribution criterion based on KDE and column 12 shows the confidence of the classification by XGBoost, based on which the rows are ordered in decreasing order. Column 13 reports the IMBH upper mass limit based on radio observations by T18. The last column lists any previous claims or other relevant literature discussing IMBHs in the cluster. The full table, including non-hosts, is presented in Appendix B.1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14
Cluster ID	Other ID	HRT	TVL	CVD	CSB	OCR	OHLR	CORELS	XGB	KDE	Pred. proba	T18 upper limit	References
NGC 6569		3.06	5.02	0.86	3.68	0.04	0.41	True	True	True	0.9999		
Pal 6		2.67	4.79	0.72	3.64	-0.02	0.35	True	True	True	0.9998		
NGC 6638		2.49	4.9	0.84	4.21	-0.34	0.27	True	True	True	0.9998		
NGC 6333	M9	3.22	5.18	0.92	3.9	-0.04	0.44	True	True	True	0.9996	$1030M_{\odot}$	Miocchi (2007)
NGC 6712		2.73	4.78	0.71	3.54	0.03	0.4	True	True	True	0.9988	$1150M_{\odot}$	Miocchi (2007) Lützgendorf et al. (2013)
NGC 6254	M10	3.25	5.04	0.79	3.65	0.02	0.47	True	True	True	0.9985	$740M_{\odot}$	Beccari et al. (2010) Umbreit and Rasio (2013)
FSR 1735		2.63	4.69	0.67	3.83	-0.29	0.33	True	True	True	0.9969		
NGC 2298		2.65	4.47	0.59	3.55	-0.28	0.38	True	True	True	0.9955		Pasquato et al. (2009)
NGC 5986		3.25	5.24	0.91	3.83	0.07	0.44	True	True	True	0.9954		
NGC 6218	M12	2.94	4.75	0.68	3.34	0.08	0.45	True	True	True	0.9947	$800M_{\odot}$	Sollima et al. (2016)
NGC 6316		3.27	5.09	0.91	4.04	-0.22	0.47	True	True	True	0.9913		
VVV-CL001		2.75	4.54	0.81	3.65	-0.27	0.34	True	True	True	0.9889		
Ton 2		3.01	4.51	0.53	3.04	0.06	0.46	True	True	True	0.9889		
NGC 6352		2.96	4.47	0.54	3.06	0.05	0.46	True	True	True	0.9871		
NGC 6779	M56	3.17	5.03	0.78	3.59	0.07	0.47	True	True	True	0.9845		
NGC 6171	M107	2.87	4.53	0.61	3.32	-0.08	0.46	True	True	True	0.9843	$990M_{\odot}$	
NGC 6553		3.26	4.95	0.88	3.67	-0.06	0.37	True	True	True	0.9837		Kains et al. (2016)
NGC 6934		3.26	4.95	0.68	3.46	0.09	0.47	True	True	True	0.9815		
NGC 6637	M69	2.95	4.95	0.79	3.76	-0.07	0.38	True	True	True	0.9724		Miocchi (2007)
UKS 1		2.85	4.63	0.62	3.29	0.04	0.46	True	True	True	0.9518		
Ter 1		2.54	4.94	0.91	4.44	-0.49	0.18	True	True	True	0.9502		
NGC 6342		2.22	4.36	0.64	3.85	-0.47	0.17	True	True	True	0.9288		
HP 1		2.89	4.58	0.72	3.18	0.1	0.46	True	True	True	0.8894		
NGC 6401		2.86	4.65	0.81	3.73	-0.28	0.39	True	True	True	0.5381		

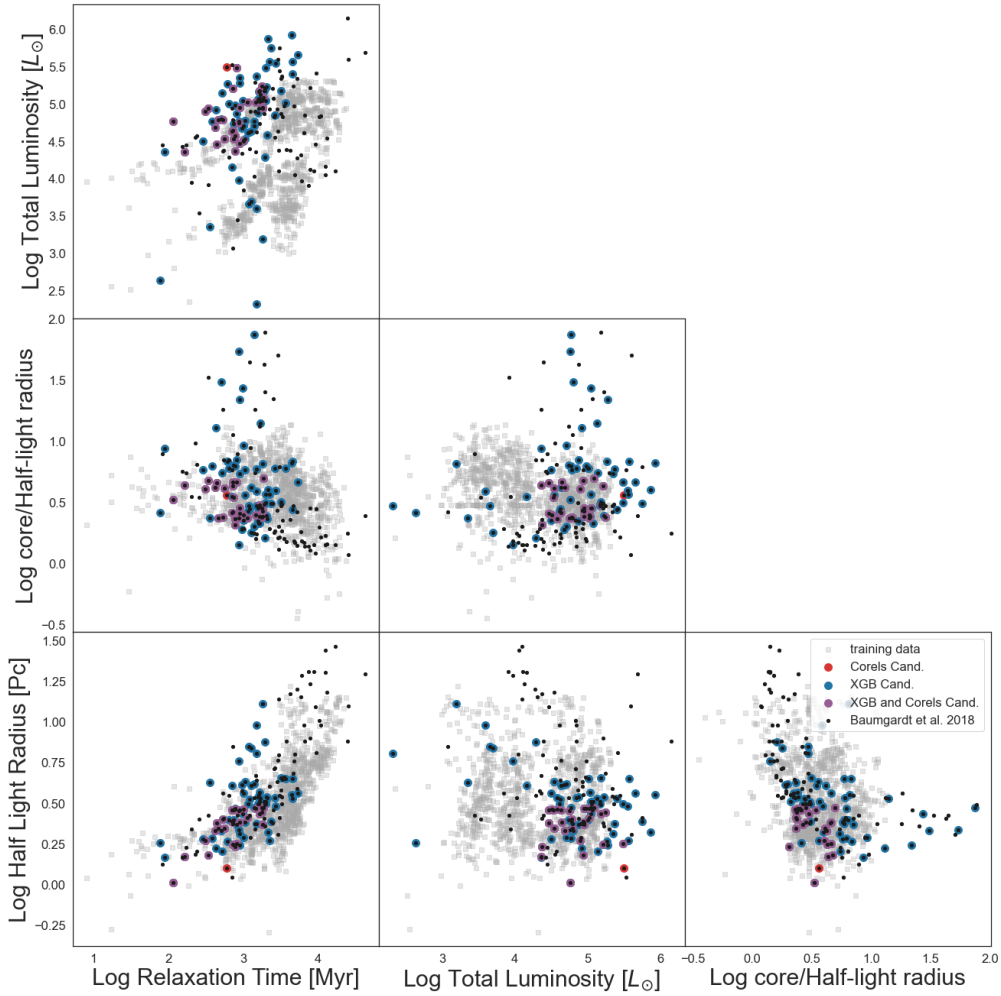


Figure 3.11: Candidate selection shown in the space of relaxation time, total luminosity, log half-light radius, and log core- over half-light radius. The gray squares represent the training data set, the black dots the data from actual star clusters compiled by Baumgardt and Hilker (2018). Star clusters selected as IMBH host candidates only by XGBoost (CORELS) are shown in blue (red), and candidates selected by both models are shown in purple. Units are shown in square brackets except for dimensionless quantities.

3.6.2 Physical interpretation of the CORELS rule list and of the anchors

To be classified as an IMBH host by CORELS, a GC has to meet the following conditions: $T_r < 2$ Gyr, $R_h < 3.0$ pc, $L > 10^4\ M_\odot$, and $R_c/R_h > 0.2$. The first rule selects dynamically relaxed GCs. The second and third rules, in combination, correspond to a condition on cluster density. The last rule selects GCs whose core is inflated with respect to the overall size of the cluster. All of these rules can be understood in terms of IMBH formation physics: mass segregation (which brings heavy stars and compact remnants to the cluster centre where they can merge to give rise to an IMBH) happens on a timescale proportional to the two-body relaxation time. The typical age of a GC is of the order of ≈ 12 Gyr, so clusters with relaxation times of 2 Gyr or shorter had ample time to undergo mass segregation.

High density (captured by the requirement to have a small radius for a minimum mass) is instrumental in increasing the frequency of close encounters that can result in IMBH formation. Finally, a dynamically puffed-up core was identified as a sign of IMBH presence early on (Baumgardt et al., 2005, Heggie et al., 2007, Hurley, 2007), as IMBHs tend to undergo dynamical exchange reactions with stellar-mass black hole (BH) binaries, forming an IMBH-BH binary that is confined to the core by its heavy mass and releases energy in the core through close encounters, heating it up. Additionally, close interactions with the IMBH can also result in stars merging with it, or they can also get ejected out of the central part of the cluster due to strong encounters. These effects contribute in depleting stars from the central part of the cluster.

There is some overlap between the CORELS rules and the anchors listed in Tab. 3.5, suggesting that the latter may also have a similar physical interpretation. However, the anchors explaining the behavior of XGBoost often include a large number of features (even all of them at times) and a disparate array of cutoffs, making a straightforward interpretation harder. These are still a potentially useful tool for investigating a specific candidate in the process of planning a follow-up observation.

For more post hoc explanations, I implemented an interactive dashboard[3] that shows feature importance, SHAP values, individual prediction analysis, what if, feature dependence. Screenshot of this dashboard are shown in appendix C. We preferred anchors for the human-understandability and the fact that no prior knowledge of ML tools (SHAP values for example) is necessary in order to understand if-then rules. In any case, as stressed in the introduction and in the introduction of this chapter: we want to shift towards natively interpretable models. Black box models can give us a false sense of security especially when we place too much trust on the explanations.

3.7 Discussion and conclusion

In the recent years, we witnessed an explosion of ML tools in astronomy. Still, most contributions avoid or only briefly touch upon the issue of interpretability and explainability of the models they adopt. In this part, we distinguish between the two approaches, with interpretability meaning that a model is natively simple enough that a human can understand its inner workings and explainability meaning that a complex or otherwise opaque model is explained post hoc in terms that a human can understand. This is a terminological distinction that has not yet been received by the astronomical

[3] 10.5281/zenodo.6407091.

Table 3.5: Anchors explaining XGBoost decision for real GCs. The first column reports the GC name, the second column the anchor as a list of conditions (each feature has been given a different colour for ease of comparison), the third column reports the anchor's precision (different from the model precision) and the fourth column its coverage.

Cluster	Anchor	Precision	Coverage
NGC 2298	['HRT ≤ 741.31', 'CSB > 458.30', '0.39 < OCR ≤ 1.05', 'OHLR ≤ 3.15', 'TVL > 26722.22', 'CVD ≤ 5.10']	0.992	0.019
NGC 5986	['TVL > 139719.63', 'OHLR ≤ 3.15', 'OCR > 1.05']	1.000	0.006
NGC 6171	['HRT ≤ 741.31', 'CSB > 458.30', 'OCR > 0.39', '2.37 < OHLR ≤ 3.15', 'CVD ≤ 5.10', 'TVL > 26722.22']	1.000	0.025
NGC 6218	['HRT ≤ 1513.56', 'CSB > 458.30', 'OCR > 1.05', 'OHLR ≤ 3.15']	0.990	0.056
FSR 1735	['HRT ≤ 741.31', 'CSB > 4371.87', '0.39 < OCR ≤ 1.05', 'OHLR ≤ 2.37', 'TVL > 26722.22', 'CVD ≤ 5.10']	0.995	0.006
NGC 6254	['TVL > 61963.19', 'OHLR ≤ 3.15', 'CSB ≤ 16864.96', '5.10 < CVD ≤ 7.60']	1.000	0.050
NGC 6316	['TVL > 61963.19', 'OHLR ≤ 3.15', 'CSB ≤ 16864.96', 'HRT > 1513.56']	0.984	0.068
NGC 6333	['TVL > 139719.63', 'OHLR ≤ 3.15', 'CSB ≤ 16864.96', 'HRT > 1513.56']	1.000	0.037
NGC 6342	['HRT ≤ 741.31', '4371.87 < CSB ≤ 16864.96', 'CVD > 3.10', 'OCR ≤ 2.15']	0.870	0.074
NGC 6352	['HRT ≤ 1513.56', '0.39 < OCR ≤ 2.15', 'TVL > 26722.22', 'OHLR ≤ 3.15', 'CVD ≤ 5.10']	1.000	0.056
HP 1	['HRT ≤ 1513.56', 'CVD > 5.10', '0.39 < OCR ≤ 2.15', '2.37 < OHLR ≤ 4.54', 'CSB > 458.30']	0.970	0.056
Ter 1	['HRT ≤ 741.31', 'CVD > 7.60', '61963.19 < TVL ≤ 139719.63', 'OCR ≤ 1.05', 'OHLR ≤ 3.15', 'CSB > 16864.96']	0.936	0.019
Ton 2	['HRT ≤ 1513.56', 'OCR > 1.05', 'OHLR ≤ 3.15']	0.984	0.056
NGC 6401	['HRT ≤ 741.31', '5.10 < CVD ≤ 7.60', '0.39 < OCR ≤ 1.05', 'OHLR ≤ 4.54', '4371.87 < CSB ≤ 16864.96', 'TVL > 26722.22']	0.969	0.019
Pal 6	['HRT ≤ 741.31', '5.10 < CVD ≤ 7.60', '0.39 < OCR ≤ 1.05', 'OHLR ≤ 2.37', 'CSB > 458.30', 'TVL > 26722.22']	0.982	0.019
UKS 1	['HRT ≤ 741.31', '3.10 < CVD ≤ 7.60', '0.39 < OCR ≤ 2.15', 'TVL > 26722.22', 'CSB > 458.30']	0.945	0.062
VVV-CL001	['HRT ≤ 741.31', '5.10 < CVD ≤ 7.60', 'OCR > 0.39', 'CSB > 4371.87']	0.981	0.025
NGC 6553	['TVL > 61963.19', 'OHLR ≤ 2.37', 'CSB ≤ 16864.96']	1.000	0.012
NGC 6569	['HRT ≤ 1513.56', 'CVD > 5.10', 'OHLR > 2.37', 'CSB > 4371.87', 'OCR > 1.05']	0.992	0.006
NGC 6637	['HRT ≤ 1513.56', '5.10 < CVD ≤ 7.60', 'TVL > 61963.19', 'CSB > 4371.87', 'OHLR > 2.37']	0.993	0.031
NGC 6638	['HRT ≤ 741.31', 'CVD > 5.10', 'OCR > 0.39', 'OHLR ≤ 2.37']	0.983	0.025
NGC 6712	['HRT ≤ 741.31', '3.10 < CVD ≤ 5.10', '0.39 < OCR ≤ 2.15', 'OHLR ≤ 3.15', 'TVL > 26722.22']	1.000	0.031
NGC 6779	['HRT ≤ 1513.56', '5.10 < CVD ≤ 7.60', 'OCR > 0.39', 'OHLR ≤ 3.15', 'TVL > 61963.19']	1.000	0.037
NGC 6934	['TVL > 61963.19', 'OHLR ≤ 3.15', 'OCR > 1.05']	1.000	0.031

community, and the trickle of works that address this issue at some level of depth often conflates the two, using interpretability as an umbrella term.

With this in mind, in the present work, we have applied both kinds of techniques, comparing a natively interpretable model (CORELs) with a black box+explanation model (XGBoost explained using the local model agnostic explanation rules known as anchors) on the problem of IMBH detection in GCs. We trained and validated both models on features measured on GC simulations, and deployed the models in prediction on actual GCs. The features we selected are standard, easily measurable characteristics of GCs and are in fact readily available in multiple catalogues. The meaning of each feature is clearly understood in theoretical terms by astronomers.

We produced a list of GCs that are likely to contain an IMBH according to both models, presenting for each one the anchor rule, the confidence score from XGBoost, and a measure of how far out of distribution each actual GC is with respect to the training data. We compare the rule list learned from the intepretable CORELS model to the anchor rules and conclude that there is a qualitative resemblance but the latter are usually more complex. Moreover, the rule list learned by CORELS has an immediate physical interpretation, some aspects of which are shared with a subset of the anchor rules.

We evaluated the performance of both CORELS and XGBoost on a test set comprised of simulations not seen in training. We find that the CORELS model we learned, based on a rule list comprised of four rules, achieves 90% precision at $\approx 50\%$ recall, whereas the precision achieved by XGBoost at the same recall is nominally perfect (100%). However, this comes at the cost of relying on a black box, which is only moderately mitigated by post hoc explanations.

Despite the simulator's best efforts, our training data and actual GCs are characterized by distributions that do not perfectly overlap in feature space. This raises the issue of whether we can trust the predictions of models trained on simulations and deployed on observational data. In this context, understanding the inner workings of a model provides a crucial advantage: we can decide whether to trust it based on how its behavior matches our subject-matter knowledge, i.e. the

physics of IMBH formation. Still, as an additional safeguard, we removed from our final candidate list GCs that were too far out of distribution with respect to the training data. This was achieved by estimating the density of training data in feature space with KDE and placing a relevant cutoff. This procedure also implicitly yields a measure of the realism of the simulations we considered, with respect to the actual observational data, and taken as a whole. The predictions cast by our models with respect to actual GCs cannot be directly tested, since we do not know which GCs host an IMBH. However, they can be used as guidance for further investigation, knowing that they are based on an interpretable model which is not working in extrapolation.

Chapter 4

Conclusion and final words

The overarching narrative of this thesis has been the deliberate and thoughtful application of native interpretable machine learning models with a clear focus on simplicity through sparsity. The journey through variable stars and the search for IMBHs in GCs has underscored the importance of choosing models that are not just powerful, but also transparent and understandable. By adopting a novel perspective on a remarkably straightforward model (logistic regression), and placing a strong emphasis on interpretability we have managed to classify variable stars, provide the governing equations (SINDy) and the dynamical mode starting from the light curves of RR Lyrae stars and finally searching for IMBH in GCs.

The classifier study demonstrates that a simple logistic regression classifier, trained on CSS light curves, can effectively distinguish RRc variables from eclipsing binaries, achieving 0.87 precision and 0.78 recall with unseen CSS light curves. The classifier also performs well on ASAS/ASAS-SN light curves, with increased recall of 0.96 and precision of 0.85, indicating strong generalization to data from different instruments. This is notable given the different proportions of RRcs in the CSS (35%) and ASAS/ASAS-SN (12%) samples, suggesting the model's simplicity helps prevent overfitting. The classifier operates solely on the shape of the light curve, as it uses normalized magnitudes at various phases without information on amplitude, absolute mean magnitude, or period. Visualization of the logistic regression coefficients reveals that the classifier relies on the phase near the light curve's minima, using the steepness of the descent to make predictions. This is physically meaningful since non-contact eclipsing binaries typically have flatter maxima and steeper minima due to the visibility and occlusion of the components, respectively. Precision-recall curves are provided for both the full and penalized classifiers, allowing for threshold adjustment based on the cost of false positives and negatives. The study also acknowledges limitations, such as the reliance on interpolating irregularly sampled light curves into regular time series using Gaussian processes with fixed parameters, which resulted in some poor fits that were excluded from the analysis

The application of SINDY algorithm on RRab, RRc and δ Scuti stars demonstrates that while simple equations derived from light-curve data can replicate the dynamics of RRc and Scuti stars, they fail to do so for RRab stars, with only a few from the sample fitting successfully. This difficulty with RRab stars may be due to nonlinear effects like shock waves, particularly as radial velocities become sonic near peak luminosity. The quality of data and outliers in light curves likely affect both RRab and RRc stars equally, so these factors don't explain the different responses to the modeling approach. When the modeling does work, the successful models are distilled into the generalized

Liénard equation, expressed as $\ddot{x} + (a + bx + c\dot{x})\dot{x} + x = 0$. When a and b are set to zero, this equation is solvable exactly and allows for both periodic and non-periodic solutions. A condition has been identified for the general equation's coefficients that must be met for a limit cycle to exist, and this behavior has also been confirmed through numerical observations in various cases. the resulting equations resemble a perturbed harmonic oscillator with complex nonlinear terms. In a specific case that was solved, two solution types were found, but only one exhibited periodic behavior. The success of finding effective governing equations seems independent of a star's position on the Bailey diagram, suggesting that the method could provide new insights into variable star classification by revealing aspects of light curve shapes not directly related to period and amplitude.

The HODMD algorithm was used to analyze variable star light curves, marking the first time DMD-based methods have been applied in this context. The aim was to extract features from the light curves that could be physically interpreted. In this process, dynamics are broken down into modes, each linked to an eigenvalue indicating either growth, decay, or oscillation. A sample of 151 stars was studied, reconstructing their light curves from these modes and comparing them to original data using the mean squared error normalized against a constant fit. Good fits with nMSE below 0.01 were achieved by adjusting the number of eigenvalues. Simpler RRc variable stars needed fewer eigenvalues, while more complex RRab stars required more for accurate representation. The Koopman dimension necessary for a given nMSE varied between RRab and RRc stars, leading to a classification threshold of 0.45 nMSE. To meet this threshold, a typical RRab star needed 27 eigenvalues (with a mode of 11), while a typical RRc star required 15 (with a mode of 7). This threshold was set based on the highest minimum nMSE across all curves. It was also noted that the number of eigenvalues for an accurate curve representation changes with the star's amplitude and period, reflecting the distinct positions of RRab and RRc stars in the period-amplitude plane.

The novel application of CORELS for the detection of IMBHs in GCs has further entrenched the thesis's dedication to interpretable machine learning. The transparent and succinct rules generated by CORELS highlight the strength of simplicity and the capability of interpretable models to yield reliable and intelligible results in complex astrophysical settings. The rule list learned by CORELS has an immediate physical interpretation in the theory of cluster dynamics. Not only, the discrepancy between the distributions of our training data and the actual Globular Clusters (GCs) in the feature space when working in a full transparent framework can be addressed: this discrepancy rightly brings into question the reliability of models that are trained on simulated data but are applied to observational data. Also, the simulations may exhibit a predisposition towards the formation of IMBHs, and this bias could influence the outcomes of the algorithms: the significant number of identified hosts, contrasted with a smaller fraction having corresponding observational claims of IMBH presence, raises considerations about the generalization of the results. This bias underscores the importance of interpreting the algorithmic outcomes with caution, recognizing the distinction between identified hosts and cluster with real observational claims. Furthermore, while the performances of CORELS and XGBoost are noteworthy, it is essential to provide a nuanced perspective. The interpretability of CORELS and the domain-specific understanding contribute to their utility, allowing for rejection of spurious claims. Precision and recall metrics, although commendable, should be evaluated within the context of the field's expertise. In these circumstances, gaining insights into the workings of a model becomes vitally important. We can decide whether to trust it based on how its behavior matches our subject-matter knowledge, i.e. the physics of IMBH

formation. This was also the case for the the featuers of the interpretable classifier on variable stars.

This, in my opinion, answer the initial question: what is interpretability in astronomical ML problems? Our definition of interpretability should match the intuition of astronomers (and in particular field-domain experts: variable stars, IMBH formation in my case).

This thesis has also innovatively (at least for the astronomy community) addressed the nuanced difference between native interpretability and post hoc explanations, a distinction often blurred in existing literature. By demonstrating the effectiveness of models that are intrinsically interpretable, this work hopes to contribute to the burgeoning of a trend that values transparency from the onset, rather than relying on post hoc explanations of black boxes.

In reflecting on this journey, I am optimistic about the evolving synergy between machine learning and astrophysics, anticipating mutual advancements in both disciplines. My commitment to clarity and interpretability in this work stems from the belief that these principles are integral for meaningful scientific progress. Looking ahead, I see opportunities to build upon the foundations laid in this thesis. Future endeavors may involve the use of more refined set of simulations for globular clusters with intermediate-mass black holes or the development of updated models for variable stars. By staying engaged in these pursuits, I aim to contribute to the ongoing intersection of machine learning and astrophysics, ensuring that the scientific imperative for clarity remains a guiding force in my own research.

Bibliography

R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, M. Agathos, et al. Gw190521: a binary black hole merger with a total mass of 150 m. *Physical review letters*, 125(10):101102, 2020.

H. K. Abdullah. Oscillation criteria of a class of generalized Lienard equation. In *Mathematical Methods and Computational Techniques in Science and Engineering*, volume 1872 of *American Institute of Physics Conference Series*, page 020006, Sept. 2017. doi: 10.1063/1.4996663.

W. Addo-Asah, H. C. Akpati, and R. E. Mickens. Investigation of a generalized van der Pol oscillator differential equation. *Journal of Sound Vibration*, 179(4):733–735, Jan. 1995. doi: 10.1006/jsvi.1995.0048.

C. Aguirre, K. Pichara, and I. Becker. Deep multi-survey classification of variable stars. *Monthly Notices of the Royal Astronomical Society*, 482(4):5078–5092, Feb. 2019. doi: 10.1093/mnras/sty2836.

A. B. Albidah, W. Brevis, V. Fedun, I. Ballai, D. B. Jess, M. Stangalini, J. Higham, and G. Verth. Proper orthogonal and dynamic mode decomposition of sunspot data. *Philosophical Transactions of the Royal Society of London Series A*, 379(2190):20200181, Feb. 2021. doi: 10.1098/rsta.2020.0181.

A. B. Albidah, V. Fedun, A. A. Aldhafeeri, I. Ballai, W. Brevis, D. B. Jess, J. Higham, M. Stangalini, S. S. A. Silva, and G. Verth. Magnetohydrodynamic Wave Mode Identification in Circular and Elliptical Sunspot Umbrae: Evidence for High-order Modes. *The Astrophysical Journal*, 927(2):201, Mar. 2022. doi: 10.3847/1538-4357/ac51d9.

C. Alcock, R. A. Allsman, D. R. Alves, T. S. Axelrod, A. C. Becker, D. P. Bennett, K. H. Cook, K. C. Freeman, M. Geha, K. Griest, M. J. Lehner, S. L. Marshall, B. J. McNamara, D. Minniti, C. Nelson, B. A. Peterson, P. Popowski, M. R. Pratt, P. J. Quinn, A. W. Rodgers, W. Sutherland, M. R. Templeton, T. Vandehei, and D. L. Welch. The MACHO Project Sample of Galactic Bulge High-Amplitude δ Scuti Stars: Pulsation Behavior and Stellar Properties. *The Astrophysical Journal*, 536(2):798–815, June 2000. doi: 10.1086/308979.

E. Angelino, N. Larus-Stone, D. Alabi, M. Seltzer, and C. Rudin. Learning certifiably optimal rule lists for categorical data. *arXiv preprint arXiv:1704.01701*, 2017.

P. P. Angelov and X. Gu. Deep rule-based classifier with human-level performance and characteristics. *Information Sciences*, 463:196–213, 2018.

S. Antol, A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision*, pages 2425–2433, 2015.

M. Arca Sedda, A. Askar, and M. Giersz. Mocca Survey Database I. Unravelling black hole subsystems in globular clusters. *ArXiv e-prints*, Jan. 2018.

H. C. Arp. Color-magnitude diagrams for seven globular clusters. *The Astronomical Journal*, 60:317, 1955. doi: 10.1086/107232.

A. Arzani and S. T. M. Dawson. Data-driven cardiovascular flow modeling: examples and opportunities. *arXiv e-prints*, art. arXiv:2010.00131, Sept. 2020.

A. Askar, P. Bianchini, R. de Vita, M. Giersz, A. Hypki, and S. Kamann. MOCCA-SURVEY Database I: Is NGC 6535 a dark star cluster harbouring an IMBH? *Monthly Notices of the Royal Astronomical Society*, 464:3090–3100, Jan. 2017a. doi: 10.1093/mnras/stw2573.

A. Askar, M. Szkudlarek, D. Gondek-Rosińska, M. Giersz, and T. Bulik. MOCCA-SURVEY Database - I. Coalescing binary black holes originating from globular clusters. *Monthly Notices of the Royal Astronomical Society*, 464:L36–L40, Jan. 2017b. doi: 10.1093/mnrasl/slw177.

A. Askar, A. Askar, M. Pasquato, and M. Giersz. Finding black holes with black boxes - using machine learning to identify globular clusters with black hole subsystems. *Monthly Notices of the Royal Astronomical Society*, 485(4):5345–5362, Jun 2019. doi: 10.1093/mnras/stz628.

A. Askar, A. Askar, M. Pasquato, and M. Giersz. Finding black holes with black boxes–using machine learning to identify globular clusters with black hole subsystems. *Monthly Notices of the Royal Astronomical Society*, 485(4):5345–5362, 2019.

A. Askar, M. B. Davies, and R. P. Church. Formation of supermassive black holes in galactic nuclei - I. Delivering seed intermediate-mass black holes in massive stellar clusters. *Monthly Notices of the Royal Astronomical Society*, 502(2):2682–2700, Apr. 2021. doi: 10.1093/mnras/stab113.

W. Baade. The Period-Luminosity Relation of the Cepheids. , 68:5, Feb. 1956. doi: 10.1086/126870.

M. Bachetti, F. A. Harrison, D. J. Walton, B. W. Grefenstette, D. Chakrabarty, F. Fürst, D. Barret, A. Beloborodov, S. E. Boggs, and F. E. Christensen. An ultraluminous X-ray source powered by an accreting neutron star. , 514(7521):202–204, Oct 2014. doi: 10.1038/nature13791.

A. Ballone, M. Mapelli, and M. Pasquato. Weighing the IMBH candidate CO-0.40-0.22* in the Galactic Centre. *Monthly Notices of the Royal Astronomical Society*, 480(4):4684–4692, Nov. 2018. doi: 10.1093/mnras/sty2139.

A. J. Barth, L. C. Ho, R. E. Rutledge, and W. L. Sargent. Pox 52: a dwarf seyfert 1 galaxy with an intermediate-mass black hole. *The Astrophysical Journal*, 607(1):90, 2004.

D. Bates, J. Chambers, and T. Hastie. Statistical models in s. *CRC Press: Boca Raton, FL, USA*, pages 421–453, 1992.

H. Baumgardt and M. Hilker. A catalogue of masses, structural parameters, and velocity dispersion profiles of 112 Milky Way globular clusters. *Monthly Notices of the Royal Astronomical Society*, 478(2):1520–1557, Aug. 2018. doi: 10.1093/mnras/sty1057.

H. Baumgardt, J. Makino, and P. Hut. Which Globular Clusters Contain Intermediate-Mass Black Holes? *The Astrophysical Journal*, 620(1):238–243, Feb. 2005. doi: 10.1086/426893.

R. L. Beaton, W. L. Freedman, B. F. Madore, G. Bono, E. K. Carlson, G. Clementini, M. J. Durbin, A. Garofalo, D. Hatt, I. S. Jang, J. A. Kollmeier, M. G. Lee, A. J. Monson, J. A. Rich, V. Scowcroft, M. Seibert, L. Sturch, and S.-C. Yang. The Carnegie-Chicago Hubble Program. I. An Independent Approach to the Extragalactic Distance Scale Using Only Population II Distance Indicators. *The Astrophysical Journal*, 832:210, Dec. 2016. doi: 10.3847/0004-637X/832/2/210.

R. L. Beaton, G. Bono, V. F. Braga, M. Dall'Ora, G. Fiorentino, I. S. Jang, C. E. Martínez-Vázquez, N. Matsunaga, M. Monelli, J. R. Neeley, and M. Salaris. Old-Aged Primary Distance Indicators. , 214(8):113, Dec 2018. doi: 10.1007/s11214-018-0542-1.

G. Beccari, M. Pasquato, G. De Marchi, E. Dalessandro, M. Trenti, and M. Gill. The Dynamical State of the Globular Cluster M10 (NGC 6254). *The Astrophysical Journal*, 713:194–204, Apr. 2010. doi: 10.1088/0004-637X/713/1/194.

G. Beccari, F. R. Ferraro, E. Dalessandro, B. Lanzoni, S. Raso, L. Origlia, E. Vesperini, J. Hong, A. Sills, A. Dieball, et al. Discovery of a double blue straggler sequence in m15: new insight into the core-collapse process. *The Astrophysical Journal*, 876(1):87, 2019.

I. Becker, K. Pichara, M. Catelan, P. Protopapas, C. Aguirre, and F. Nikzat. Scalable end-to-end recurrent neural network for variable star classification. *Monthly Notices of the Royal Astronomical Society*, 493(2):2981–2995, Apr. 2020. doi: 10.1093/mnras/staa350.

A. Bell, I. Solano-Kamaiko, O. Nov, and J. Stoyanovich. It's just not that simple: an empirical study of the accuracy-explainability trade-off in machine learning for public policy. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 248–266, 2022.

D. Belloni, M. Giersz, A. Askar, N. Leigh, and A. Hypki. MOCCA-SURVEY database I. Accreting white dwarf binary systems in globular clusters - I. Cataclysmic variables - present-day population. *Monthly Notices of the Royal Astronomical Society*, 462:2950–2969, Nov. 2016. doi: 10.1093/mnras/stw1841.

D. Belloni, M. Giersz, H. J. Rocha-Pinto, N. W. C. Leigh, and A. Askar. MOCCA-SURVEY database I. Accreting white dwarf binary systems in globular clusters - II. Cataclysmic variables - progenitors and population at birth. *Monthly Notices of the Royal Astronomical Society*, 464:4077–4095, Feb. 2017a. doi: 10.1093/mnras/stw2516.

D. Belloni, M. Zorotovic, M. R. Schreiber, N. W. C. Leigh, M. Giersz, and A. Askar. mocca-SURVEY database I. Accreting white dwarf binary systems in globular clusters - III. Cataclysmic variables - implications of model assumptions. *Monthly Notices of the Royal Astronomical Society*, 468:2429–2446, June 2017b. doi: 10.1093/mnras/stx575.

D. Belloni, P. Kroupa, H. J. Rocha-Pinto, and M. Giersz. Dynamical equivalence, the origin of the Galactic field stellar and binary population, and the initial radius-mass relation of embedded clusters. *Monthly Notices of the Royal Astronomical Society*, 474:3740–3745, Mar. 2018. doi: 10.1093/mnras/stx3034.

P. Bevanda, S. Sosnowski, and S. Hirche. Koopman operator dynamical models: Learning, analysis and control. *Annual Reviews in Control*, 52:197–212, 2021. doi: 10.1016/j.arcontrol.2021.09.002. URL `https://doi.org/10.1016%2Fj.arcontrol.2021.09.002`.

J. Bialopetravičius, D. Narbutis, and V. Vansevičius. Deriving star cluster parameters with convolutional neural networks. I. Age, mass, and size. *Astronomy & Astrophysics*, 621:A103, Jan. 2019. doi: 10.1051/0004-6361/201833833.

C. M. Bishop. Mixture density networks. 1994.

V. Blanco. The reddening of type ab rr lyrae stars. *The Astronomical Journal*, 104:734–747, 1992.

A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

J. Bond, W. Arnett, and B. J. Carr. The evolution and fate of very massive objects. *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 280, May 15, 1984, p. 825-847. Research supported by the Science and Engineering Research Council.*, 280:825–847, 1984.

G. Bono and R. F. Stellingwerf. Cepheid convective models. , 64:559, Jan. 1993.

G. Bono and R. F. Stellingwerf. Pulsation and Stability of RR Lyrae Stars. I. Instability Strip. , 93: 233, July 1994. doi: 10.1086/192054.

G. Bono, F. Caputo, V. Castellani, M. Marconi, J. Storm, and S. Degl'Innocenti. A pulsational approach to near-infrared and visual magnitudes of RR Lyr stars. *Monthly Notices of the Royal Astronomical Society*, 344:1097–1106, Oct. 2003. doi: 10.1046/j.1365-8711.2003.06878.x.

G. Bono, G. Iannicola, V. F. Braga, I. Ferraro, P. B. Stetson, D. Magurno, N. Matsunaga, R. L. Beaton, R. Buonanno, B. Chaboyer, M. Dall'Ora, M. Fabrizio, G. Fiorentino, W. L. Freedman, C. K. Gilligan, B. F. Madore, M. Marconi, M. Marengo, S. Marinoni, P. M. Marrese, C. E. Martinez-Vazquez, M. Mateo, M. Monelli, J. R. Neeley, M. Nonino, C. Sneden, F. Thevenin, E. Valenti, and A. R. Walker. On a New Method to Estimate the Distance, Reddening, and Metallicity of RR Lyrae Stars Using Optical/Near-infrared (B, V, I, J, H, K) Mean Magnitudes: ω Centauri as a First Test Case. *The Astrophysical Journal*, 870:115, Jan. 2019. doi: 10.3847/1538-4357/aaf23f.

V. F. Braga, M. Dall'Ora, G. Bono, P. B. Stetson, I. Ferraro, G. Iannicola, M. Marengo, J. Neeley, S. E. Persson, R. Buonanno, G. Coppola, W. Freedman, B. F. Madore, M. Marconi, N. Matsunaga, A. Monson, J. Rich, V. Scowcroft, and M. Seibert. On the Distance of the Globular Cluster M4 (NGC6121) Using RR Lyrae Stars. I. Optical and Near-infrared Period-Luminosity and Period-Wesenheit Relations. *The Astrophysical Journal*, 799:165, Feb. 2015. doi: 10.1088/0004-637X/799/2/165.

V. F. Braga, P. B. Stetson, G. Bono, M. Dall'Ora, I. Ferraro, G. Fiorentino, L. M. Freyhammer, G. Iannicola, M. Marengo, J. Neeley, E. Valenti, R. Buonanno, A. Calamida, M. Castellani, R. da Silva, S. Degl'Innocenti, A. Di Cecco, M. Fabrizio, W. L. Freedman, G. Giuffrida, J. Lub, B. F. Madore, M. Marconi, S. Marinoni, N. Matsunaga, M. Monelli, S. E. Persson, A. M. Piersimoni, A. Pietrinferni, P. Prada-Moroni, L. Pulone, R. Stellingwerf, E. Tognelli, and A. R. Walker. On the RR Lyrae Stars in Globulars. IV. ω Centauri Optical UBVRI Photometry. *The Astronomical Journal*, 152:170, Dec. 2016. doi: 10.3847/0004-6256/152/6/170.

V. F. Braga, P. B. Stetson, G. Bono, M. Dall'Ora, I. Ferraro, G. Fiorentino, G. Iannicola, L. Inno, M. Marengo, J. Neeley, R. L. Beaton, R. Buonanno, A. Calamida, R. Contreras Ramos, B. Chaboyer, M. Fabrizio, W. L. Freedman, C. K. Gilligan, K. V. Johnston, J. Lub, B. F. Madore, D. Magurno, M. Marconi, S. Marinoni, P. M. Marrese, M. Mateo, N. Matsunaga, D. Minniti, A. J. Monson, M. Monelli, M. Nonino, S. E. Persson, A. Pietrinferni, C. Sneden, J. Storm, A. R. Walker, E. Valenti, and M. Zoccali. New near-infrared JHK$_s$ light-curve templates for RR Lyrae variables. *Astronomy & Astrophysics*, 625:A1, May 2019a. doi: 10.1051/0004-6361/201834893.

V. F. Braga, P. B. Stetson, G. Bono, M. Dall'Ora, I. Ferraro, G. Fiorentino, G. Iannicola, L. Inno, M. Marengo, J. Neeley, R. L. Beaton, R. Buonanno, A. Calamida, R. Contreras Ramos, B. Chaboyer, M. Fabrizio, W. L. Freedman, C. K. Gilligan, K. V. Johnston, J. Lub, B. F. Madore, D. Magurno, M. Marconi, S. Marinoni, P. M. Marrese, M. Mateo, N. Matsunaga, D. Minniti, A. J. Monson, M. Monelli, M. Nonino, S. E. Persson, A. Pietrinferni, C. Sneden, J. Storm, A. R. Walker, E. Valenti, and M. Zoccali. New near-infrared JHK$_s$ light-curve templates for RR Lyrae variables. *Astronomy & Astrophysics*, 625:A1, May 2019b. doi: 10.1051/0004-6361/201834893.

H. Brink, J. W. Richards, D. Poznanski, J. S. Bloom, J. Rice, S. Negahban, and M. Wainwright. Using machine learning for discovery in synoptic survey imaging data. *Monthly Notices of the Royal Astronomical Society*, 435(2):1047–1060, Oct. 2013. doi: 10.1093/mnras/stt1306.

S. L. Brunton. Sparse identification of nonlinear dynamics (sindy): Sparse machine learning models 5 years later!, August 27 2021. Video presentation.

S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1517384113. URL `https://www.pnas.org/content/113/15/3932`.

G. Cabrera, S. E. Hong, L. Nakazono, D. Parkinson, and Y.-S. Ting. Panel discussion: Practical problem solving for machine learning. *arXiv preprint arXiv:2211.16782*, 2022.

B. Carr, F. Kühnel, and M. Sandstad. Primordial black holes as dark matter. *Physical Review D*, 94 (8):083504, 2016.

N. Castro, P. Protopapas, and K. Pichara. Uncertain Classification of Variable Stars: Handling Observational GAPS and Noise. *The Astronomical Journal*, 155(1):16, Jan. 2018. doi: 10.3847/1538-3881/aa9ab8.

M. Catelan and H. A. Smith. *Pulsating stars*. John Wiley & Sons, 2015.

M. Catelan, B. J. Pritzl, and H. A. Smith. The RR Lyrae Period-Luminosity Relation. I. Theoretical Calibration. , 154(2):633–649, Oct. 2004. doi: 10.1086/422916.

J. Chardin and P. Bianchini. Predicting images for the dynamics of stellar clusters (π-DOC): a deep learning framework to predict mass, distance, and age of globular clusters. *Monthly Notices of the Royal Astronomical Society*, 504(4):5656–5670, July 2021. doi: 10.1093/mnras/stab737.

A. K. Chattopadhyay and T. Chattopadhyay. *Statistical methods for astronomical data analysis*, volume 3. Springer, 2014.

C. Chen, K. Lin, C. Rudin, Y. Shaposhnik, S. Wang, and T. Wang. An interpretable model with globally consistent explanations for credit risk. *arXiv preprint arXiv:1811.12615*, 2018a.

J. Chen, L. Song, M. Wainwright, and M. Jordan. Learning to explain: An information-theoretic perspective on model interpretation. In *International conference on machine learning*, pages 883–892. PMLR, 2018b.

L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018c.

T. Chen and C. Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4232-2. doi: 10.1145/2939672.2939785. URL `http://doi.acm.org/10.1145/2939672.2939785`.

M. Coleman Miller and D. P. Hamilton. Production of intermediate-mass black holes in globular clusters. *Monthly Notices of the Royal Astronomical Society*, 330(1):232–240, 2002.

S. A. Colgate. Stellar coalescence and the multiple supernova interpretation of quasi-stellar sources. *Astrophysical Journal, vol. 150, p. 163*, 150:163, 1967.

R. Contreras Ramos, D. Minniti, F. Gran, M. Zoccali, J. Alonso-García, P. Huijse, M. G. Navarro, Á. Rojas-Arriagada, and E. Valenti. The VVV Survey RR Lyrae Population in the Galactic Center Region. *The Astrophysical Journal*, 863:79, Aug. 2018. doi: 10.3847/1538-4357/aacf90.

K. Course and P. B. Nair. State estimation of a physical system with unknown governing equations. *Nature*, 622(7982):261–267, 2023.

J. Crestani, V. F. Braga, M. Fabrizio, G. Bono, C. Sneden, G. W. Preston, I. Ferraro, G. Iannicola, M. Nonino, G. Fiorentino, F. Thévenin, B. Lemasle, Z. Prudil, A. Alves-Brito, G. Altavilla, B. Chaboyer, M. Dall'Ora, V. D'Orazi, C. K. Gilligan, E. Grebel, A. J. Koch-Hansen, H. Lala, M. Marengo, S. Marinoni, P. M. Marrese, C. E. Martínez-Vázquez, N. Matsunaga, M. Monelli, J. P. Mullen, J. Neeley, R. da Silva, P. B. Stetson, M. Salaris, J. Storm, E. Valenti, and M. Zoccali. On the Use of Field RR Lyrae as Galactic Probes. III. The α-element abundances. *arXiv e-prints*, art. arXiv:2104.08113, Apr. 2021a.

J. Crestani, M. Fabrizio, V. F. Braga, C. Sneden, G. Preston, I. Ferraro, G. Iannicola, G. Bono, A. Alves-Brito, M. Nonino, V. D'Orazi, L. Inno, M. Monelli, J. Storm, G. Altavilla, B. Chaboyer,

M. Dall'Ora, G. Fiorentino, C. Gilligan, E. K. Grebel, H. Lala, B. Lemasle, M. Marengo, S. Marinoni, P. M. Marrese, C. E. Martínez-Vázquez, N. Matsunaga, J. P. Mullen, J. Neeley, Z. Prudil, R. da Silva, P. B. Stetson, F. Thévenin, E. Valenti, A. Walker, and M. Zoccali. On the Use of Field RR Lyrae as Galactic Probes. II. A New ΔS Calibration to Estimate Their Metallicity. *The Astrophysical Journal*, 908(1):20, Feb. 2021b. doi: 10.3847/1538-4357/abd183.

D. Cseh, P. Kaaret, S. Corbel, E. Körding, M. Coriat, A. Tzioumis, and B. Lanzoni. Radio observations of NGC 6388: an upper limit on the mass of its central black hole. *Monthly Notices of the Royal Astronomical Society*, 406:1049–1054, Aug. 2010. doi: 10.1111/j.1365-2966.2010.16726.x.

K. Darling and L. M. Widrow. Eigenfunctions of Galactic phase space spirals from dynamic mode decomposition. *Monthly Notices of the Royal Astronomical Society*, 490(1):114–123, Nov. 2019. doi: 10.1093/mnras/stz2539.

B. de Silva, K. Champion, M. Quade, J.-C. Loiseau, J. Kutz, and S. Brunton. PySINDy: A Python package for the sparse identification of nonlinear dynamical systems from data. *The Journal of Open Source Software*, 5(49):2104, May 2020. doi: 10.21105/joss.02104.

J. Debosscher, L. M. Sarro, C. Aerts, J. Cuypers, B. Vandenbussche, R. Garrido, and E. Solano. Automated supervised classification of variable stars. I. Methodology. *Astronomy & Astrophysics*, 475(3):1159–1183, Dec. 2007. doi: 10.1051/0004-6361:20077638.

N. Demo, M. Tezzele, and G. Rozza. Pydmd: Python dynamic mode decomposition. *Journal of Open Source Software*, 3(22):530, 2018.

P. Di Cintio, M. Pasquato, A. Simon-Petit, and S.-J. Yoon. Introducing a new multi-particle collision method for the evolution of dense stellar systems-ii. core collapse. *Astronomy & Astrophysics*, 659: A19, 2022.

X. Dong, Y.-L. Bai, Y. Lu, and M. Fan. An improved sparse identification of nonlinear dynamics with akaike information criterion and group sparsity. *Nonlinear Dynamics*, 111(2):1485–1510, 2023.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

A. J. Drake, M. Catelan, S. G. Djorgovski, G. Torrealba, M. J. Graham, V. Belokurov, S. E. Koposov, A. Mahabal, J. L. Prieto, C. Donalek, R. Williams, S. Larson, E. Christensen, and E. Beshore. Probing the Outer Galactic Halo with RR Lyrae from the Catalina Surveys. *The Astrophysical Journal*, 763(1):32, Jan. 2013a. doi: 10.1088/0004-637X/763/1/32.

A. J. Drake, M. Catelan, S. G. Djorgovski, G. Torrealba, M. J. Graham, A. Mahabal, J. L. Prieto, C. Donalek, R. Williams, S. Larson, E. Christensen, and E. Beshore. Evidence for a Milky Way Tidal Stream Reaching Beyond 100 kpc. *The Astrophysical Journal*, 765(2):154, Mar. 2013b. doi: 10.1088/0004-637X/765/2/154.

A. J. Drake, M. J. Graham, S. G. Djorgovski, M. Catelan, A. A. Mahabal, G. Torrealba, D. García-Álvarez, C. Donalek, J. L. Prieto, R. Williams, S. Larson, E. Christen sen, V. Belokurov, S. E. Koposov, E. Beshore, A. Boattini, A. Gibbs, R. Hill, R. Kowalski, J. Johnson, and F. Shelly. The Catalina Surveys Periodic Variable Star Catalog. , 213(1):9, July 2014a. doi: 10.1088/0067-0049/213/1/9.

A. J. Drake, M. J. Graham, S. G. Djorgovski, M. Catelan, A. A. Mahabal, G. Torrealba, D. García-Álvarez, C. Donalek, J. L. Prieto, R. Williams, S. Larson, E. Christen sen, V. Belokurov, S. E. Koposov, E. Beshore, A. Boattini, A. Gibbs, R. Hill, R. Kowalski, J. Johnson, and F. Shelly. The Catalina Surveys Periodic Variable Star Catalog. , 213(1):9, July 2014b. doi: 10.1088/0067-0049/213/1/9.

A. J. Drake, S. G. Djorgovski, M. Catelan, M. J. Graham, A. A. Mahabal, S. Larson, E. Christensen, G. Torrealba, E. Beshore, R. H. McNaught, G. Garradd, V. Belokurov, and S. E. Koposov. The Catalina Surveys Southern periodic variable star catalogue. *Monthly Notices of the Royal Astronomical Society*, 469(3):3688–3712, Aug. 2017. doi: 10.1093/mnras/stx1085.

P. Dubath, L. Rimoldini, M. Süveges, J. Blomme, M. López, L. M. Sarro, J. De Ridder, J. Cuypers, L. Guy, I. Lecoeur, K. Nienartowicz, A. Jan, M. Beck, N. Mowlavi, P. De Cat, T. Lebzelter, and L. Eyer. Random forest automated supervised classification of Hipparcos periodic variable stars. *Monthly Notices of the Royal Astronomical Society*, 414(3):2602–2617, July 2011. doi: 10.1111/j.1365-2966.2011.18575.x.

R. M. Dudley. *Real analysis and probability*. CRC Press, 2018.

G. K. Dziugaite, S. Ben-David, and D. M. Roy. Enforcing interpretability and its statistical impacts: Trade-offs between accuracy and interpretability. *arXiv preprint arXiv:2010.13764*, 2020.

A. E. Evrard, P. Arnault, D. Huterer, and A. Farahi. A model for multiproperty galaxy cluster statistics. *Monthly Notices of the Royal Astronomical Society*, 441(4):3562–3569, 2014.

M. Fabrizio, V. F. Braga, J. Crestani, G. Bono, I. Ferraro, G. Fiorentino, G. Iannicola, G. W. Preston, C. Sneden, F. Thévenin, G. Altavilla, B. Chaboyer, M. Dall'Ora, R. da Silva, E. K. Grebel, C. K. Gilligan, H. Lala, B. Lemasle, D. Magurno, M. Marengo, S. Marinoni, P. M. Marrese, C. E. Martínez-Vázquez, N. Matsunaga, M. Monelli, J. P. Mullen, J. Neeley, M. Nonino, Z. Prudil, M. Salaris, P. B. Stetson, E. Valenti, and M. Zoccali. On the Use of Field RR Lyrae As Galactic Probes: IV. New Insights Into and Around the Oosterhoff Dichotomy. *The Astrophysical Journal*, 919(2):118, Oct. 2021. doi: 10.3847/1538-4357/ac1115.

X. Fan, V. K. Narayanan, R. H. Lupton, M. A. Strauss, G. R. Knapp, R. H. Becker, R. L. White, L. Pentericci, S. Leggett, Z. Haiman, et al. A survey of z$>$ 5.8 quasars in the sloan digital sky survey. i. discovery of three new quasars and the spatial density of luminous quasars at z 6. *The Astronomical Journal*, 122(6):2833, 2001.

S. A. Farrell, N. A. Webb, D. Barret, O. Godet, and J. M. Rodrigues. An intermediate-mass black hole of over 500 solar masses in the galaxy ESO243-49. , 460(7251):73–75, Jul 2009. doi: 10.1038/nature08083.

S. A. Farrell, N. A. Webb, D. Barret, O. Godet, and J. M. Rodrigues. An intermediate-mass black hole of over 500 solar masses in the galaxy eso 243-49. *Nature*, 460(7251):73–75, 2009.

A. V. Filippenko and W. L. Sargent. Discovery of an extremely low luminosity seyfert 1 nucleus in the dwarf galaxy ngc 4395. *Astrophysical Journal, Part 2-Letters (ISSN 0004-637X), vol. 342, July 1, 1989, p. L11-L14. Research supported by the University of California.*, 342:L11–L14, 1989.

J. M. Fregeau, P. Cheung, S. F. Portegies Zwart, and F. A. Rasio. Stellar collisions during binary-binary and binary-single star interactions. *Monthly Notices of the Royal Astronomical Society*, 352:1–19, July 2004. doi: 10.1111/j.1365-2966.2004.07914.x.

M. Freitag, P. Amaro-Seoane, and V. Kalogera. Stellar remnants in galactic nuclei: mass segregation. *The Astrophysical Journal*, 649(1):91, 2006.

K. P. F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 1901. doi: 10.1080/14786440109462720.

K. Fujii, N. Takeishi, M. Hojo, Y. Inaba, and Y. Kawahara. Physically-interpretable classification of biological network dynamics for complex collective motions. *Scientific Reports*, 10(1), feb 2020. doi: 10.1038/s41598-020-58064-w. URL `https://doi.org/10.1038%2Fs41598-020-58064-w`.

M. Fujii and S. Portegies Zwart. The growth of massive stars via stellar collisions in ensemble star clusters. *Monthly Notices of the Royal Astronomical Society*, 430(2):1018–1029, 2013.

T. Fukushige and D. C. Heggie. The time-scale of escape from star clusters. *Monthly Notices of the Royal Astronomical Society*, 318:753–761, Nov. 2000. doi: 10.1046/j.1365-8711.2000.03811.x.

Gaia Collaboration, T. Prusti, J. H. J. de Bruijne, A. G. A. Brown, A. Vallenari, C. Babusiaux, C. A. L. Bailer-Jones, U. Bastian, M. Biermann, D. W. Evans, L. Eyer, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Mignard, D. J. Milligan, C. Panem, V. Poinsignon, D. Pourbaix, S. Randich, G. Sarri, P. Sartoretti, H. I. Siddiqui, C. Soubiran, V. Valette, F. van Leeuwen, N. A. Walton, C. Aerts, F. Arenou, M. Cropper, R. Drimmel, E. Høg, D. Katz, M. G. Lattanzi, W. O'Mullane, E. K. Grebel, A. D. Holland, C. Huc, X. Passot, L. Bramante, C. Cacciari, J. Castañeda, L. Chaoul, N. Cheek, F. De Angeli, C. Fabricius, R. Guerra, J. Hernández, A. Jean-Antoine-Piccolo, E. Masana, R. Messineo, N. Mowlavi, K. Nienartowicz, D. Ordóñez-Blanco, P. Panuzzo, J. Portell, P. J. Richards, M. Riello, G. M. Seabroke, P. Tanga, F. Thévenin, J. Torra, S. G. Els, G. Gracia-Abril, G. Comoretto, M. Garcia-Reinaldos, T. Lock, E. Mercier, M. Altmann, R. Andrae, T. L. Astraatmadja, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, G. Busso, B. Carry, A. Cellino, G. Clementini, S. Cowell, O. Creevey, J. Cuypers, M. Davidson, J. De Ridder, A. de Torres, L. Delchambre, A. Dell'Oro, C. Ducourant, Y. Frémat, M. García-Torres, E. Gosset, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, M. Hauser, D. Hestroffer, S. T. Hodgkin, H. E. Huckle, A. Hutton, G. Jasniewicz, S. Jordan, M. Kontizas, A. J. Korn, A. C. Lanzafame, M. Manteiga, A. Moitinho, K. Muinonen, J. Osinde, E. Pancino, T. Pauwels, J. M. Petit, A. Recio-Blanco, A. C. Robin, L. M. Sarro, C. Siopis, M. Smith, K. W. Smith, A. Sozzetti, W. Thuillot, W. van Reeven, Y. Viala, U. Abbas, A. Abreu Aramburu, S. Accart, J. J. Aguado, P. M. Allan, W. Allasia, G. Altavilla, M. A. Álvarez, J. Alves, R. I. Anderson, A. H. Andrei,

E. Anglada Varela, E. Antiche, T. Antoja, S. Antón, B. Arcay, A. Atzei, L. Ayache, N. Bach, S. G. Baker, L. Balaguer-Núñez, C. Barache, C. Barata, A. Barbier, F. Barblan, M. Baroni, D. Barrado y Navascués, M. Barros, M. A. Barstow, U. Becciani, M. Bellazzini, G. Bellei, A. Bello García, V. Belokurov, P. Bendjoya, A. Berihuete, L. Bianchi, O. Bienaymé, F. Billebaud, N. Blagorodnova, S. Blanco-Cuaresma, T. Boch, A. Bombrun, R. Borrachero, S. Bouquillon, G. Bourda, H. Bouy, A. Bragaglia, M. A. Breddels, N. Brouillet, T. Brüsemeister, B. Bucciarelli, F. Budnik, P. Burgess, R. Burgon, A. Burlacu, D. Busonero, R. Buzzi, E. Caffau, J. Cambras, H. Campbell, R. Cancelliere, T. Cantat-Gaudin, T. Carlucci, J. M. Carrasco, M. Castellani, P. Charlot, J. Charnas, P. Charvet, F. Chassat, A. Chiavassa, M. Clotet, G. Cocozza, R. S. Collins, P. Collins, G. Costigan, F. Crifo, N. J. G. Cross, M. Crosta, C. Crowley, C. Dafonte, Y. Damerdji, A. Dapergolas, P. David, M. David, P. De Cat, F. de Felice, P. de Laverny, F. De Luise, R. De March, D. de Martino, R. de Souza, J. Debosscher, E. del Pozo, M. Delbo, A. Delgado, H. E. Delgado, F. di Marco, P. Di Matteo, S. Diakite, E. Distefano, C. Dolding, S. Dos Anjos, P. Drazinos, J. Durán, Y. Dzigan, E. Ecale, B. Edvardsson, H. Enke, M. Erdmann, D. Escolar, M. Espina, N. W. Evans, G. Eynard Bontemps, C. Fabre, M. Fabrizio, S. Faigler, A. J. Falcão, M. Farràs Casas, F. Faye, L. Federici, G. Fedorets, J. Fernández-Hernández, P. Fernique, A. Fienga, F. Figueras, F. Filippi, K. Findeisen, A. Fonti, M. Fouesneau, E. Fraile, M. Fraser, J. Fuchs, R. Furnell, M. Gai, S. Galleti, L. Galluccio, D. Garabato, F. García-Sedano, P. Garé, A. Garofalo, N. Garralda, P. Gavras, J. Gerssen, R. Geyer, G. Gilmore, S. Girona, G. Giuffrida, M. Gomes, A. González-Marcos, J. González-Núñez, J. J. González-Vidal, M. Granvik, A. Guerrier, P. Guillout, J. Guiraud, A. Gúrpide, R. Gutiérrez-Sánchez, L. P. Guy, R. Haigron, D. Hatzidimitriou, M. Haywood, U. Heiter, A. Helmi, D. Hobbs, W. Hofmann, B. Holl, G. Holland, J. A. S. Hunt, A. Hypki, V. Icardi, M. Irwin, G. Jevardat de Fombelle, P. Jofré, P. G. Jonker, A. Jorissen, F. Julbe, A. Karampelas, A. Kochoska, R. Kohley, K. Kolenberg, E. Kontizas, S. E. Koposov, G. Kordopatis, P. Koubsky, A. Kowalczyk, A. Krone-Martins, M. Kudryashova, I. Kull, R. K. Bachchan, F. Lacoste-Seris, A. F. Lanza, J. B. Lavigne, C. Le Poncin-Lafitte, Y. Lebreton, T. Lebzelter, S. Leccia, N. Leclerc, I. Lecoeur-Taibi, V. Lemaitre, H. Lenhardt, F. Leroux, S. Liao, E. Licata, H. E. P. Lindstrøm, T. A. Lister, E. Livanou, A. Lobel, W. Löffler, M. López, A. Lopez-Lozano, D. Lorenz, T. Loureiro, I. MacDonald, T. Magalhães Fernandes, S. Managau, R. G. Mann, G. Mantelet, O. Marchal, J. M. Marchant, M. Marconi, J. Marie, S. Marinoni, P. M. Marrese, G. Marschalkó, D. J. Marshall, J. M. Martín-Fleitas, M. Martino, N. Mary, G. Matijevič, T. Mazeh, P. J. McMillan, S. Messina, A. Mestre, D. Michalik, N. R. Millar, B. M. H. Miranda, D. Molina, R. Molinaro, M. Molinaro, L. Molnár, M. Moniez, P. Montegriffo, D. Monteiro, R. Mor, A. Mora, R. Morbidelli, T. Morel, S. Morgenthaler, T. Morley, D. Morris, A. F. Mulone, T. Muraveva, I. Musella, J. Narbonne, G. Nelemans, L. Nicastro, L. Noval, C. Ordénovic, J. Ordieres-Meré, P. Osborne, C. Pagani, I. Pagano, F. Pailler, H. Palacin, L. Palaversa, P. Parsons, T. Paulsen, M. Pecoraro, R. Pedrosa, H. Pentikäinen, J. Pereira, B. Pichon, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poujoulet, A. Prša, L. Pulone, S. Ragaini, S. Rago, N. Rambaux, M. Ramos-Lerate, P. Ranalli, G. Rauw, A. Read, S. Regibo, F. Renk, C. Reylé, R. A. Ribeiro, L. Rimoldini, V. Ripepi, A. Riva, G. Rixon, M. Roelens, M. Romero-Gómez, N. Rowell, F. Royer, A. Rudolph, L. Ruiz-Dern, G. Sadowski, T. Sagristà Sellés, J. Sahlmann, J. Salgado, E. Salguero, M. Sarasso, H. Savietto, A. Schnorhk, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D. Segransan, E. Serpell, I. C. Shih, R. Smareglia, R. L. Smart, C. Smith, E. Solano, F. Solitro, R. Sordo, S. Soria Nieto, J. Souchay,

A. Spagna, F. Spoto, U. Stampa, I. A. Steele, H. Steidelmüller, C. A. Stephenson, H. Stoev, F. F. Suess, M. Süveges, J. Surdej, L. Szabados, E. Szegedi-Elek, D. Tapiador, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, D. Terrett, B. Tingley, S. C. Trager, C. Turon, A. Ulla, E. Utrilla, G. Valentini, A. van Elteren, E. Van Hemelryck, M. van Leeuwen, M. Varadi, A. Vecchiato, J. Veljanoski, T. Via, D. Vicente, S. Vogt, H. Voss, V. Votruba, S. Voutsinas, G. Walmsley, M. Weiler, K. Weingrill, D. Werner, T. Wevers, G. Whitehead, . Wyrzykowski, A. Yoldas, M. Žerjal, S. Zucker, C. Zurbach, T. Zwitter, A. Alecu, M. Allen, C. Allende Prieto, A. Amorim, G. Anglada-Escudé, V. Arsenijevic, S. Azaz, P. Balm, M. Beck, H. H. Bernstein, L. Bigot, A. Bijaoui, C. Blasco, M. Bonfigli, G. Bono, S. Boudreault, A. Bressan, S. Brown, P. M. Brunet, P. Bunclark, R. Buonanno, A. G. Butkevich, C. Carret, C. Carrion, L. Chemin, F. Chéreau, L. Corcione, E. Darmigny, K. S. de Boer, P. de Teodoro, P. T. de Zeeuw, C. Delle Luche, C. D. Domingues, P. Dubath, F. Fodor, B. Frézouls, A. Fries, D. Fustes, D. Fyfe, E. Gallardo, J. Gallegos, D. Gardiol, M. Gebran, A. Gomboc, A. Gómez, E. Grux, A. Gueguen, A. Heyrovsky, J. Hoar, G. Iannicola, Y. Isasi Parache, A. M. Janotto, E. Joliet, A. Jonckheere, R. Keil, D. W. Kim, P. Klagyivik, J. Klar, J. Knude, O. Kochukhov, I. Kolka, J. Kos, A. Kutka, V. Lainey, D. LeBouquin, C. Liu, D. Loreggia, V. V. Makarov, M. G. Marseille, C. Martayan, O. Martinez-Rubi, B. Massart, F. Meynadier, S. Mignot, U. Munari, A. T. Nguyen, T. Nordlander, P. Ocvirk, K. S. O'Flaherty, A. Olias Sanz, P. Ortiz, J. Osorio, D. Oszkiewicz, A. Ouzounis, M. Palmer, P. Park, E. Pasquato, C. Peltzer, J. Peralta, F. Péturaud, T. Pieniluoma, E. Pigozzi, J. Poels, G. Prat, T. Prod'homme, F. Raison, J. M. Rebordao, D. Risquez, B. Rocca-Volmerange, S. Rosen, M. I. Ruiz-Fuertes, F. Russo, S. Sembay, I. Serraller Vizcaino, A. Short, A. Siebert, H. Silva, D. Sinachopoulos, E. Slezak, M. Soffel, D. Sosnowska, V. Straižys, M. ter Linden, D. Terrell, S. Theil, C. Tiede, L. Troisi, P. Tsalmantza, D. Tur, M. Vaccari, F. Vachier, P. Valles, W. Van Hamme, L. Veltz, J. Virtanen, J. M. Wallut, R. Wichmann, M. I. Wilkinson, H. Ziaeepour, and S. Zschocke. The Gaia mission. *Astronomy & Astrophysics*, 595:A1, Nov. 2016. doi: 10.1051/0004-6361/201629272.

Gaia Collaboration et al. Gaia data release 3. summary of the content and survey properties. *Astronomy & Astrophysics*, 2022.

K. Gebhardt, R. M. Rich, and L. C. Ho. An intermediate-mass black hole in the globular cluster g1: improved significance from new keck and hubble space telescope observations. *The Astrophysical Journal*, 634(2):1093, 2005.

M. Giersz, D. C. Heggie, J. R. Hurley, and A. Hypki. MOCCA code for star cluster simulations - II. Comparison with N-body simulations. *Monthly Notices of the Royal Astronomical Society*, 431: 2184–2199, May 2013. doi: 10.1093/mnras/stt307.

M. Giersz, N. Leigh, A. Hypki, N. Lützgendorf, and A. Askar. MOCCA code for star cluster simulations - IV. A new scenario for intermediate mass black hole formation in globular clusters. *Monthly Notices of the Royal Astronomical Society*, 454:3150–3165, Dec. 2015. doi: 10.1093/mnras/stv2162.

M. Giersz, N. Leigh, A. Hypki, N. Lützgendorf, and A. Askar. Mocca code for star cluster simulations–iv. a new scenario for intermediate mass black hole formation in globular clusters. *Monthly Notices of the Royal Astronomical Society*, 454(3):3150–3165, 2015.

M. Giersz, A. Askar, L. Wang, A. Hypki, A. Leveque, and R. Spurzem. MOCCA survey data base-I. Dissolution of tidally filling star clusters harbouring black hole subsystems. *Monthly Notices of the Royal Astronomical Society*, 487(2):2412–2423, Aug 2019. doi: 10.1093/mnras/stz1460.

D. Gillet and A. B. Fokin. Emission lines and shock waves in RR Lyrae stars. *Astronomy & Astrophysics*, 565:A73, May 2014. doi: 10.1051/0004-6361/201322938.

D. Gillet, B. Mauclaire, T. Lemoult, P. Mathias, J. S. Devaux, T. de France, and T. Garrel. Dynamical structure of the pulsating atmosphere of RR Lyrae. I. A typical pulsation cycle. *Astronomy & Astrophysics*, 623:A109, Mar. 2019a. doi: 10.1051/0004-6361/201833869.

N. Gillet, A. Mesinger, B. Greig, A. Liu, and G. Ucci. Deep learning from 21-cm tomography of the cosmic dawn and reionization. *Monthly Notices of the Royal Astronomical Society*, 484(1): 282–293, Mar. 2019b. doi: 10.1093/mnras/stz010.

P. Glendinning. *Stability, instability and chaos: an introduction to the theory of nonlinear differential equations*. Cambridge university press, 1994.

O. Godet, J. Lombardi, F. Antonini, D. Barret, N. A. Webb, J. Vingless, and M. Thomas. Implications of the delayed 2013 outburst of eso 243-49 hlx-1. *The Astrophysical Journal*, 793(2):105, 2014.

J. Goodman and P. Hut. Primordial binaries and globular cluster evolution. *Nature*, 339(6219): 40–42, 1989.

J. E. Greene and L. C. Ho. Active galactic nuclei with candidate intermediate-mass black holes. *The Astrophysical Journal*, 610(2):722, 2004.

J. E. Greene, J. Strader, and L. C. Ho. Intermediate-mass black holes. *Annual Review of Astronomy and Astrophysics*, 58:257–312, 2020.

J. E. Greene, L. Lancaster, Y.-S. Ting, S. E. Koposov, S. Danieli, S. Huang, F. Jiang, J. P. Greco, and J. Strader. A Search for Wandering Black Holes in the Milky Way with Gaia and DECaLS. *The Astrophysical Journal*, 917(1):17, Aug. 2021. doi: 10.3847/1538-4357/ac0896.

B. M. Greenwell, B. C. Boehmke, and A. J. McCarthy. A simple and effective model-based variable importance measure. *arXiv preprint arXiv:1805.04755*, 2018.

Y. Guan, S. L. Brunton, and I. Novosselov. Sparse nonlinear models of chaotic electroconvection. *arXiv e-prints*, art. arXiv:2009.11862, Sept. 2020.

P. Hacker and J.-H. Passoth. Varieties of ai explanations under the law. from the gdpr to the aia, and beyond. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, pages 343–373. Springer, 2020.

E. Haghighat and R. Juanes. Sciann: A keras/tensorflow wrapper for scientific computations and physics-informed deep learning using artificial neural networks. *Computer Methods in Applied Mechanics and Engineering*, 373:113552, 2021.

D. J. Hand. Classifier technology and the illusion of progress. 2006.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

D. C. Heggie. Towards an N-body model for the globular cluster M4. *Monthly Notices of the Royal Astronomical Society*, 445:3435–3443, Dec. 2014. doi: 10.1093/mnras/stu1976.

D. C. Heggie, P. Hut, S. Mineshige, J. Makino, and H. Baumgardt. The Core Radius of a Star Cluster Containing a Massive Black Hole. *Publications of the Astronomical Society of Japan*, 59: L11–L14, June 2007. doi: 10.1093/pasj/59.3.L11.

M. H. Hénon. The Monte Carlo Method (Papers appear in the Proceedings of IAU Colloquium No. 10 Gravitational N-Body Problem (ed. by Myron Lecar), R. Reidel Publ. Co. , Dordrecht-Holland.). , 14:151–167, Nov. 1971. doi: 10.1007/BF00649201.

T. A. Hinners, K. Tat, and R. Thorp. Machine Learning Techniques for Stellar Light Curve Classification. *The Astronomical Journal*, 156(1):7, July 2018. doi: 10.3847/1538-3881/aac16d.

T. K. Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

T. Hoefler, D. Alistarh, T. Ben-Nun, N. Dryden, and A. Peste. Sparsity in deep learning: Pruning and growth for efficient inference and training in neural networks. *The Journal of Machine Learning Research*, 22(1):10882–11005, 2021.

J. Hong, A. Askar, M. Giersz, A. Hypki, and S.-J. Yoon. MOCCA-SURVEY Database I: Binary black hole mergers from globular clusters with intermediate mass black holes. *Monthly Notices of the Royal Astronomical Society*, 498(3):4287–4294, Nov. 2020. doi: 10.1093/mnras/staa2677.

K. Hori, S. M. Tobias, and R. J. Teed. Dynamic mode decomposition to retrieve torsional Alfvén waves. *arXiv e-prints*, art. arXiv:2009.13095, Sept. 2020.

J. Horrocks and C. T. Bauch. Algorithmic discovery of dynamic models from infectious disease data. *Scientific Reports*, 10:7061, Apr. 2020. doi: 10.1038/s41598-020-63877-w.

C.-W. Huang, D. Krueger, A. Lacoste, and A. Courville. Neural autoregressive flows. In *International Conference on Machine Learning*, pages 2078–2087. PMLR, 2018.

D. Huppenkothen, M. Ntampaka, M. Ho, M. Fouesneau, B. Nord, J. Peek, M. Walmsley, J. F. Wu, C. Avestruz, T. Buck, et al. Constructing impactful machine learning research for astronomy: Best practices for researchers and reviewers. *arXiv preprint arXiv:2310.12528*, 2023.

J. R. Hurley. Ratios of star cluster core and half-mass radii: a cautionary note on intermediate-mass black holes in star clusters. *Monthly Notices of the Royal Astronomical Society*, 379(1):93–99, July 2007. doi: 10.1111/j.1365-2966.2007.11912.x.

J. R. Hurley and M. M. Shara. A direct n-body model of core-collapse and core oscillations. *Monthly Notices of the Royal Astronomical Society*, 425(4):2872–2879, 2012.

J. R. Hurley, O. R. Pols, and C. A. Tout. Comprehensive analytic formulae for stellar evolution as a function of mass and metallicity. *Monthly Notices of the Royal Astronomical Society*, 315:543–569, July 2000. doi: 10.1046/j.1365-8711.2000.03426.x.

J. R. Hurley, C. A. Tout, and O. R. Pols. Evolution of binary stars and the effect of tides on binary populations. *Monthly Notices of the Royal Astronomical Society*, 329:897–928, Feb. 2002. doi: 10.1046/j.1365-8711.2002.05038.x.

A. Hypki and M. Giersz. MOCCA code for star cluster simulations - I. Blue stragglers, first results. *Monthly Notices of the Royal Astronomical Society*, 429:1221–1243, Feb. 2013. doi: 10.1093/mnras/sts415.

P. T. Inc. Collaborative data science, 2015. URL `https://plot.ly`.

E. Ishida, R. Beck, S. González-Gaitán, R. de Souza, A. Krone-Martins, J. Barrett, N. Kennamer, R. Vilalta, J. Burgess, B. Quint, et al. Optimizing spectroscopic follow-up strategies for supernova photometric classification with active learning. *Monthly Notices of the Royal Astronomical Society*, 483(1):2–18, 2019.

Ž. Ivezić, S. M. Kahn, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, D. Alonso, Y. AlSayyad, S. F. Anderson, J. Andrew, et al. Lsst: from science drivers to reference design and anticipated data products. *The Astrophysical Journal*, 873(2):111, 2019.

T. Jayasinghe, K. Z. Stanek, C. S. Kochanek, B. J. Shappee, T. W. S. Holoien, T. A. Thompson, J. L. Prieto, S. Dong, M. Pawlak, O. Pejcha, J. V. Shields, G. Pojmanski, S. Otero, N. Hurst, C. A. Britt, and D. Will. The ASAS-SN catalogue of variable stars III: variables in the southern TESS continuous viewing zone. *Monthly Notices of the Royal Astronomical Society*, 485(1):961–971, May 2019. doi: 10.1093/mnras/stz444.

C. I. Johnson and C. A. Pilachowski. Chemical Abundances for 855 Giants in the Globular Cluster Omega Centauri (NGC 5139). *The Astrophysical Journal*, 722(2):1373–1410, Oct. 2010. doi: 10.1088/0004-637X/722/2/1373.

J. Jurcsik and G. Kovacs. Determination of [Fe/H] from the light curves of RR Lyrae stars. *Astronomy & Astrophysics*, 312:111–120, Aug. 1996a.

J. Jurcsik and G. Kovacs. Determination of [Fe/H] from the light curves of RR Lyrae stars. *Astronomy & Astrophysics*, 312:111–120, Aug. 1996b.

N. Kains, D. M. Bramich, K. C. Sahu, and A. Calamida. Searching for intermediate-mass black holes in globular clusters with gravitational microlensing. *Monthly Notices of the Royal Astronomical Society*, 460(2):2025–2035, Aug. 2016. doi: 10.1093/mnras/stw1137.

M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian processes and kernel methods: A review on connections and equivalences. *arXiv preprint arXiv:1807.02582*, 2018.

M. Kaneko, T. Oka, H. Yokozuka, R. Enokiya, S. Takekawa, Y. Iwata, and S. Tsujimoto. Discovery of the Tadpole Molecular Cloud near the Galactic Nucleus. *The Astrophysical Journal*, 942(1):46, Jan. 2023. doi: 10.3847/1538-4357/aca66a.

T. Kawaguchi, M. Kawasaki, T. Takayama, M. Yamaguchi, and J. Yokoyama. Formation of intermediate-mass black holes as primordial black holes in the inflationary cosmology with running spectral index. *Monthly Notices of the Royal Astronomical Society*, 388(3):1426–1432, 2008.

K. Kinemuchi, H. Smith, P. Woźniak, T. McKay, R. Collaboration, et al. Analysis of rr lyrae stars in the northern sky variability survey. *The Astronomical Journal*, 132(3):1202, 2006.

K. Kinemuchi, H. A. Smith, P. R. Woźniak, T. A. McKay, and ROTSE Collaboration. Analysis of RR Lyrae Stars in the Northern Sky Variability Survey. *The Astronomical Journal*, 132(3): 1202–1220, Sept. 2006. doi: 10.1086/506198.

I. King. The structure of star clusters. i. an empirical density law. *Astronomical Journal, Vol. 67, p. 471 (1962)*, 67:471, 1962.

B. Kızıltan, H. Baumgardt, and A. Loeb. An intermediate-mass black hole in the centre of the globular cluster 47 Tucanae. , 542(7640):203–205, Feb 2017. doi: 10.1038/nature21361.

B. Kızıltan, H. Baumgardt, and A. Loeb. An intermediate-mass black hole in the centre of the globular cluster 47 tucanae. *Nature*, 542(7640):203–205, 2017.

P. Kontschieder, M. Fiterau, A. Criminisi, and S. R. Bulo. Deep neural decision forests. In *Proceedings of the IEEE international conference on computer vision*, pages 1467–1475, 2015.

R. P. Kraft. Abundance Differences among Globular Cluster Giants: Primordial vs. Evolutionary Scenarios. , 106:553, June 1994. doi: 10.1086/133416.

P. Kroupa. On the variation of the initial mass function. *Monthly Notices of the Royal Astronomical Society*, 322:231–246, Apr. 2001. doi: 10.1046/j.1365-8711.2001.04022.x.

J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

F. Lalande and A. A. Trani. Predicting the Stability of Hierarchical Triple Systems with Convolutional Neural Networks. *arXiv e-prints*, art. arXiv:2206.12402, June 2022.

F. Lanusse, Q. Ma, N. Li, T. E. Collett, C.-L. Li, S. Ravanbakhsh, R. Mandelbaum, and B. Póczos. CMU DeepLens: deep learning for automatic image-based galaxy-galaxy strong lens finding. *Monthly Notices of the Royal Astronomical Society*, 473(3):3895–3906, Jan. 2018. doi: 10.1093/mnras/stx1665.

F. Lanusse, Q. Ma, N. Li, T. E. Collett, C.-L. Li, S. Ravanbakhsh, R. Mandelbaum, and B. Póczos. Cmu deeplens: deep learning for automatic image-based galaxy–galaxy strong lens finding. *Monthly Notices of the Royal Astronomical Society*, 473(3):3895–3906, 2018.

S. Larson, E. Beshore, R. Hill, E. Christensen, D. McLean, S. Kolar, R. McNaught, and G. Garradd. The CSS and SSS NEO surveys. In *AAS/Division for Planetary Sciences Meeting Abstracts #35*, volume 35 of *AAS/Division for Planetary Sciences Meeting Abstracts*, page 36.04, May 2003.

S. Le Clainche and J. M. Vega. Higher order dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 16(2):882–925, 2017. doi: 10.1137/15M1054924. URL `https://doi.org/10.1137/15M1054924`.

C. Le Losq and Y. Feng. gcvsplie. `https://github.com/charlesll/gcvspline`, 2007.

Y. W. Lee, J. M. Joo, Y. J. Sohn, S. C. Rey, H. C. Lee, and A. R. Walker. Multiple stellar populations in the globular cluster ω Centauri as tracers of a merger event. , 402(6757):55–57, Nov. 1999. doi: 10.1038/46985.

Y. Li, Y. Ni, R. A. Croft, T. Di Matteo, S. Bird, and Y. Feng. Ai-assisted superresolution cosmological simulations. *Proceedings of the National Academy of Sciences*, 118(19):e2022038118, 2021.

D. Lin, J. Strader, E. R. Carrasco, D. Page, A. J. Romanowsky, J. Homan, J. A. Irwin, R. A. Remillard, O. Godet, and N. A. Webb. A luminous X-ray outburst from an intermediate-mass black hole in an off-centre star cluster. *Nature Astronomy*, 2:656–661, Jun 2018. doi: 10.1038/s41550-018-0493-1.

M. Linkmann and M. Dikpati. Symmetry-reduced dynamic mode decomposition of solar tachocline dynamics. In *AGU Fall Meeting Abstracts*, volume 2021, pages SH53C–06, Dec. 2021.

M. Linkmann, S. Dallas, and M. Dikpati. Data-driven low-dimensional modelling of solar tachocline dynamics. In *AGU Fall Meeting Abstracts*, volume 2020, pages SH006–02, Dec. 2020.

P. Lipton. Contrastive explanation. *Royal Institute of Philosophy Supplements*, 27:247–266, 1990.

N. R. Lomb. Least-Squares Frequency Analysis of Unequally Spaced Data. , 39(2):447–462, Feb. 1976. doi: 10.1007/BF00648343.

J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang. Learning under concept drift: A review. *IEEE transactions on knowledge and data engineering*, 31(12):2346–2363, 2018.

J. Lub. An atlas of light and colour curves of field RR Lyrae stars. , 29:345–378, Sept. 1977.

S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30, 2017.

N. Lützgendorf, H. Baumgardt, and J. M. D. Kruijssen. N-body simulations of globular clusters in tidal fields: Effects of intermediate-mass black holes. *Astronomy & Astrophysics*, 558:A117, Oct. 2013. doi: 10.1051/0004-6361/201321927.

T. J. Maccarone and M. Servillat. Radio observations of NGC 2808 and other globular clusters: constraints on intermediate-mass black holes. *Monthly Notices of the Royal Astronomical Society*, 389:379–384, Sept. 2008. doi: 10.1111/j.1365-2966.2008.13577.x.

P. Madau and M. J. Rees. Massive black holes as population iii remnants. *The Astrophysical Journal*, 551(1):L27, 2001.

J. P. Madrid, N. W. C. Leigh, J. R. Hurley, and M. Giersz. Mass evaporation rate of globular clusters in a strong tidal field. *Monthly Notices of the Royal Astronomical Society*, 470:1729–1737, Sept. 2017. doi: 10.1093/mnras/stx1350.

K. Małek, A. Solarz, A. Pollo, A. Fritz, B. Garilli, M. Scodeggio, A. Iovino, B. Granett, U. Abbas, C. Adami, et al. The vimos public extragalactic redshift survey (vipers)-a support vector machine classification of galaxies, stars, and agns. *Astronomy & Astrophysics*, 557:A16, 2013.

K. Maliszewski, M. Giersz, D. Gondek-Rosinska, A. Askar, and A. Hypki. MOCCA-SURVEY data base II - Properties of intermediate mass black holes escaping from star clusters. *Monthly Notices of the Royal Astronomical Society*, 514(4):5879–5889, Aug. 2022. doi: 10.1093/mnras/stac1728.

F. J. Masci, R. R. Laher, B. Rusholme, D. L. Shupe, S. Groom, J. Surace, E. Jackson, S. Monkewitz, R. Beck, D. Flynn, et al. The zwicky transient facility: Data processing, products, and archive. *Publications of the Astronomical Society of the Pacific*, 131(995):018003, 2018.

G. J. McLachlan and K. E. Basford. *Mixture models: Inference and applications to clustering*, volume 38. M. Dekker New York, 1988.

M. Mezcua. Observational evidence for intermediate-mass black holes. *International Journal of Modern Physics D*, 26(11):1730021, 2017.

M. Mezcua, F. Civano, G. Fabbiano, T. Miyaji, and S. Marchesi. A Population of Intermediate-mass Black Holes in Dwarf Starburst Galaxies Up to Redshift=1.5. *The Astrophysical Journal*, 817(1): 20, Jan. 2016. doi: 10.3847/0004-637X/817/1/20.

M. Mezcua, F. Civano, S. Marchesi, H. Suh, G. Fabbiano, and M. Volonteri. Intermediate-mass black holes in dwarf galaxies out to redshift ∼2.4 in the Chandra COSMOS-Legacy Survey. *Monthly Notices of the Royal Astronomical Society*, 478(2):2576–2591, Aug. 2018. doi: 10.1093/mnras/sty1163.

A. P. Milone and A. F. Marino. Multiple populations in star clusters. *Universe*, 8(7):359, 2022.

P. Miocchi. The presence of intermediate-mass black holes in globular clusters and their connection with extreme horizontal branch stars. *Monthly Notices of the Royal Astronomical Society*, 381: 103–116, Oct. 2007. doi: 10.1111/j.1365-2966.2007.12165.x.

T. M. Mitchell. Machine learning, 1997.

J. Moehlis, K. Josic, and E. T. Shea-Brown. Periodic orbit. *Scholarpedia*, 1(7):1358, 2006. doi: 10.4249/scholarpedia.1358. revision #153208.

C. Molnar. *Interpretable Machine Learning*. 2 edition, 2022. URL `https://christophm.github.io/interpretable-ml-book`.

J. Morawski, M. Giersz, A. Askar, and K. Belczynski. MOCCA-SURVEY Database I: Assessing GW kick retention fractions for BH-BH mergers in globular clusters. *Monthly Notices of the Royal Astronomical Society*, 481(2):2168–2179, Dec. 2018. doi: 10.1093/mnras/sty2401.

D. J. Mortlock, S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, R. G. McMahon, C. Simpson, T. Theuns, E. A. Gonzáles-Solares, A. Adamson, S. Dye, N. C. Hambly, P. Hirst, M. J. Irwin, E. Kuiper, A. Lawrence, and H. J. A. Röttgering. A luminous quasar at a redshift of z = 7.085. , 474(7353):616–619, June 2011. doi: 10.1038/nature10159.

J. P. Mullen, M. Marengo, C. E. Martínez-Vázquez, J. R. Neeley, G. Bono, M. Dall'Ora, B. Chaboyer, F. Thévenin, V. F. Braga, J. Crestani, M. Fabrizio, G. Fiorentino, C. K. Gilligan, M. Monelli, and P. B. Stetson. Metallicity of galactic RR lyrae from optical and infrared light curves. i.

period–fourier–metallicity relations for fundamental-mode RR lyrae. *The Astrophysical Journal*, 912 (2):144, may 2021. doi: 10.3847/1538-4357/abefd4. URL `https://doi.org/10.3847/1538-4357/abefd4`.

J. P. Mullen, M. Marengo, C. E. Martínez-Vázquez, G. Bono, V. F. Braga, B. Chaboyer, J. Crestani, M. Dall'Ora, M. Fabrizio, G. Fiorentino, M. Monelli, J. R. Neeley, P. B. Stetson, and F. Thévenin. Metallicity of Galactic RR Lyrae from Optical and Infrared Light Curves. II. Period-Fourier-Metallicity Relations for First Overtone RR Lyrae. *The Astrophysical Journal*, 931(2):131, June 2022. doi: 10.3847/1538-4357/ac67ee.

M. Nagy and K. Petrovay. Oscillator models of the solar cycle and the Waldmeier effect. *Astronomische Nachrichten*, 334(9):964, Nov. 2013. doi: 10.1002/asna.201211971.

M. Nauta, R. Van Bree, and C. Seifert. Neural prototype trees for interpretable fine-grained image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14933–14943, 2021.

J. Nemec and M. Mateo. SX Phœnicis stars. In C. Cacciari and G. Clementini, editors, *Confrontation Between Stellar Pulsation and Evolution*, volume 11 of *Astronomical Society of the Pacific Conference Series*, pages 64–84, Jan. 1990.

A. Ng. Nuts and bolts of building ai applications using deep learning. *NIPS Keynote Talk*, 2016.

J. E. Norris and G. S. Da Costa. The Giant Branch of omega Centauri. IV. Abundance Patterns Based on Echelle Spectra of 40 Red Giants. *The Astrophysical Journal*, 447:680, July 1995. doi: 10.1086/175909.

I. Nun, P. Protopapas, B. Sim, M. Zhu, R. Dave, N. Castro, and K. Pichara. FATS: Feature Analysis for Time Series. *arXiv e-prints*, art. arXiv:1506.00010, May 2015.

T. Oka, R. Mizuno, K. Miura, and S. Takekawa. Signature of an Intermediate-mass Black Hole in the Central Molecular Zone of Our Galaxy. , 816(1):L7, Jan. 2016. doi: 10.3847/2041-8205/816/1/L7.

P. T. Oosterhoff. Some remarks on the variable stars in globular clusters. *The Observatory*, 62: 104–109, Apr. 1939.

S. Pan. *Robust and Interpretable Learning for Operator-Theoretic Modeling of Non-linear Dynamics*. PhD thesis, 01 2021.

R. Pantoja, M. Catelan, K. Pichara, and P. Protopapas. Semi-supervised classification and clustering analysis for variable stars. *Monthly Notices of the Royal Astronomical Society*, 517(3):3660–3681, Dec. 2022. doi: 10.1093/mnras/stac2715.

M. Pasquato, M. Trenti, G. De Marchi, M. Gill, D. P. Hamilton, M. C. Miller, M. Stiavelli, and R. P. van der Marel. Mass Segregation in NGC 2298: Limits on the Presence of an Intermediate Mass Black Hole. *The Astrophysical Journal*, 699:1511–1517, July 2009. doi: 10.1088/0004-637X/699/2/1511.

M. Pasquato, M. Abbas, A. A. Trani, M. Nori, J. A. Kwiecinski, P. Trevisan, V. F. Braga, G. Bono, and A. V. Macciò. Sparse identification of variable star dynamics. *The Astrophysical Journal*, 930 (2):161, 2022.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

G. Pojmanski. The All Sky Automated Survey. *Acta Astronomica*, 47:467–481, Oct. 1997.

G. Pojmanski. The All Sky Automated Survey. Catalog of Variable Stars. I. 0 h - 6 hQuarter of the Southern Hemisphere. *Acta Astronomica*, 52:397–427, Dec. 2002.

G. W. Preston, C. Sneden, and M. Chadid. Hydrogen and Helium Shock Phenomena during Rising Light in RR Lyrae Fundamental Mode Pulsators. *The Astronomical Journal*, 163(3):109, Mar. 2022. doi: 10.3847/1538-3881/ac46ca.

Z. Prudil and M. Skarka. Blazhko effect in the Galactic bulge fundamental mode RR Lyrae stars - I. Incidence rate and differences between modulated and non-modulated stars. *Monthly Notices of the Royal Astronomical Society*, 466:2602–2613, Apr. 2017. doi: 10.1093/mnras/stw3231.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL `https://www.R-project.org/`.

M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

M. Raissi, A. Yazdani, and G. E. Karniadakis. Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations. *Science*, 367(6481):1026–1030, 2020.

C. E. Rasmussen. Gaussian processes in machine learning. In *Summer school on machine learning*, pages 63–71. Springer, 2003.

O. Razim, S. Cavuoti, M. Brescia, G. Riccio, M. Salvato, and G. Longo. Improving the reliability of photometric redshift with machine learning. *Monthly Notices of the Royal Astronomical Society*, 507(4):5034–5052, 2021.

S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

M. T. Ribeiro, S. Singh, and C. Guestrin. Model-agnostic interpretability of machine learning. *arXiv preprint arXiv:1606.05386*, 2016.

M. T. Ribeiro, S. Singh, and C. Guestrin. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

J. W. Richards, D. L. Starr, N. R. Butler, J. S. Bloom, J. M. Brewer, A. Crellin-Quick, J. Higgins, R. Kennedy, and M. Rischard. On machine-learned classification of variable stars with sparse and noisy time-series data. *The Astrophysical Journal*, 733(1):10, 2011.

C. L. Rodriguez, M. Zevin, P. Amaro-Seoane, S. Chatterjee, K. Kremer, F. A. Rasio, and S. Y. Claire. Black holes: The next generation—repeated mergers in dense star clusters and their gravitational-wave properties. *Physical Review D*, 100(4):043027, 2019.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.

C. Rudin and Ş. Ertekin. Learning customized and optimized lists of rules with mathematical programming. *Mathematical Programming Computation*, 10:659–702, 2018.

J. Samsing, A. Askar, and M. Giersz. MOCCA-SURVEY Database I: Eccentric Black Hole Gravitational Wave Mergers Forming In Binary-Single Interactions. *ArXiv e-prints*, Dec. 2017.

R. H. Sanders. The effects of stellar collisions in dense stellar systems. *The Astrophysical Journal*, 162:791, 1970.

S. Sarkar, T. Weyde, A. d. Garcez, G. G. Slabaugh, S. Dragicevic, and C. Percy. Accuracy and interpretability trade-offs in machine learning applied to safer gambling. In *CEUR Workshop Proceedings*, volume 1773. CEUR Workshop Proceedings, 2016.

J. D. Scargle. Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, 263:835–853, Dec. 1982. doi: 10.1086/160554.

J.-T. Schindler, E. Bañados, T. Connor, R. Decarli, X. Fan, E. P. Farina, C. Mazzucchelli, R. Nanni, H.-W. Rix, D. Stern, B. P. Venemans, and F. Walter. The Pan-STARRS1 z $>$ 5.6 Quasar Survey. III. The z $\approx$ 6 Quasar Luminosity Function. *The Astrophysical Journal*, 943(1):67, Jan. 2023. doi: 10.3847/1538-4357/aca7ca.

P. J. Schmid. Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics*, 656:5–28, 2010. doi: 10.1017/S0022112010001217.

P. J. Schmid. Dynamic Mode Decomposition and Its Variants. *Annual Review of Fluid Mechanics*, 54:225–254, Jan. 2022. doi: 10.1146/annurev-fluid-030121-015835.

A. Schwarzenberg-Czerny. On the advantage of using analysis of variance for period search. *Monthly Notices of the Royal Astronomical Society*, 241:153–165, Nov. 1989. doi: 10.1093/mnras/241.2.153.

L. Semenova, C. Rudin, and R. Parr. On the existence of simpler machine learning models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1827–1858, 2022.

L. S. Shapley. Notes on the n-person game—ii: The value of an n-person game. 1951.

B. J. Shappee, J. L. Prieto, D. Grupe, C. S. Kochanek, K. Z. Stanek, G. De Rosa, S. Mathur, Y. Zu, B. M. Peterson, R. W. Pogge, S. Komossa, M. Im, J. Jencson, T. W.-S. Holoien, U. Basu, J. F. Beacom, D. M. Szczygieł, J. Brimacombe, S. Adams, A. Campillay, C. Choi, C. Contreras, M. Dietrich, M. Dubberley, M. Elphick, S. Foale, M. Giustini, C. Gonzalez, E. Hawkins, D. A. Howell, E. Y. Hsiao, M. Koss, K. M. Leighly, N. Morrell, D. Mudd, D. Mullins, J. M. Nugent, J. Parrent, M. M. Phillips, G. Pojmanski, W. Rosing, R. Ross, D. Sand, D. M. Terndrup, S. Valenti, Z. Walker, and Y. Yoon. The Man behind the Curtain: X-Rays Drive the UV through NIR Variability in the 2013 Active Galactic Nucleus Outburst in NGC 2617. *The Astrophysical Journal*, 788:48, June 2014. doi: 10.1088/0004-637X/788/1/48.

M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock. The Two Micron All Sky Survey (2MASS). *The Astronomical Journal*, 131:1163–1183, Feb. 2006. doi: 10.1086/498708.

R. Smolec. Metallicity Dependence of the Blazhko Effect. *Acta Astronomica*, 55:59–84, Mar. 2005.

A. Sollima, F. R. Ferraro, L. Lovisi, F. Contenta, E. Vesperini, L. Origlia, E. Lapenna, B. Lanzoni, A. Mucciarelli, E. Dalessandro, and C. Pallanca. The dark mass content of the Milky Way globular clusters NGC288 and NGC6218 . , 87:614, Jan. 2016.

L. Spitzer. *Dynamical evolution of globular clusters*. 1987.

L. S. Spitzer Jr. *Dynamical evolution of globular clusters*, volume 25. Princeton University Press, 2014.

J. S. Stodolkiewicz. Dynamical evolution of globular clusters. I. *Acta Astronomica*, 32(1-2):63–91, Jan 1982.

J. S. Stodolkiewicz. Dynamical evolution of globular clusters. II - Binaries Method. *Acta Astronomica*, 36(1):19–41, Jan 1986.

J. Strader, L. Chomiuk, T. J. Maccarone, J. C. A. Miller-Jones, A. C. Seth, C. O. Heinke, and G. R. Sivakoff. No Evidence for Intermediate-mass Black Holes in Globular Clusters: Strong Constraints from the JVLA. *The Astrophysical Journal*, 750:L27, May 2012. doi: 10.1088/2041-8205/750/2/L27.

C. Sturch. Intrinsic ubv colors of rr lyrae stars. *The Astrophysical Journal*, 143:774, 1966.

J. Su, D. V. Vargas, and K. Sakurai. One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, 2019.

Z. Su, Z. Li, M. Hou, M. Zhang, and Z. Cheng. A Chandra survey of milky way globular clusters - III. Searching for X-ray signature of intermediate-mass black holes. *Monthly Notices of the Royal Astronomical Society*, 516(2):1788–1807, Oct. 2022. doi: 10.1093/mnras/stac2345.

D. Sugimoto and E. Bettwieser. Post-collapse evolution of globular clusters. *Monthly Notices of the Royal Astronomical Society*, 204(1):19P–22P, 1983.

T. Szklenár, A. Bódi, D. Tarczay-Nehéz, K. Vida, G. Mező, and R. Szabó. Variable Star Classification with a Multiple-input Neural Network. *The Astrophysical Journal*, 938(1):37, Oct. 2022. doi: 10.3847/1538-4357/ac8df3.

K. Taira, S. L. Brunton, S. T. M. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley. Modal Analysis of Fluid Flows: An Overview. *AIAA Journal*, 55(12):4013–4041, Dec. 2017. doi: 10.2514/1.J056060.

S. Takekawa, T. Oka, Y. Iwata, S. Tsujimoto, and M. Nomura. Indication of Another Intermediate-mass Black Hole in the Galactic Center. , 871(1):L1, Jan. 2019. doi: 10.3847/2041-8213/aafb07.

S. Takekawa, T. Oka, Y. Iwata, S. Tsujimoto, and M. Nomura. Indication of another intermediate-mass black hole in the galactic center. *The Astrophysical Journal Letters*, 871(1):L1, 2019.

S. Takekawa, T. Oka, S. Tokuyama, K. Tanabe, Y. Iwata, S. Tsujimoto, M. Nomura, and Y. Shibuya. An energetic high-velocity compact cloud: CO-0.31+0.11. *Publications of the Astronomical Society of Japan*, 71:S21, Dec. 2019. doi: 10.1093/pasj/psz027.

S. Takekawa, T. Oka, Y. Iwata, S. Tsujimoto, and M. Nomura. The Fifth Candidate for an Intermediate-mass Black Hole in the Galactic Center. *The Astrophysical Journal*, 890(2):167, Feb. 2020. doi: 10.3847/1538-4357/ab6f6f.

S. Takekawa, T. Oka, Y. Iwata, S. Tsujimoto, and M. Nomura. The fifth candidate for an intermediate-mass black hole in the galactic center. *The Astrophysical Journal*, 890(2):167, 2020.

M. Tan and Q. Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

M. Tan, R. Pang, and Q. V. Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020.

K. Tanaka. Alma images of the host cloud of the intermediate-mass black hole candidate co-0.40–0.22*: No evidence for cloud–black hole interaction, but evidence for a cloud–cloud collision. *The Astrophysical Journal*, 859(2):86, 2018.

Y. Tanaka, K. Seya, and M. Takeuti. Nonlinear coupling models of variable stars. In C. Cacciari and G. Clementini, editors, *Confrontation Between Stellar Pulsation and Evolution*, volume 11 of *Astronomical Society of the Pacific Conference Series*, pages 145–148, Jan. 1990.

The GRAVITY Collaboration, O. Straub, M. Bauböck, R. Abuter, N. Aimar, P. Amaro Seoane, A. Amorim, J. P. Berger, H. Bonnet, G. Bourdarot, W. Brandner, V. Cardoso, Y. Clénet, Y. Dallilar, R. Davies, P. T. de Zeeuw, J. Dexter, A. Drescher, F. Eisenhauer, N. M. Förster Schreiber, A. Foschi, P. Garcia, F. Gao, E. Gendron, R. Genzel, S. Gillessen, M. Habibi, X. Haubois, G. Heißel, T. Henning, S. Hippler, M. Horrobin, L. Jochum, L. Jocou, A. Kaufer, P. Kervella, S. Lacour, V. Lapeyrère, J. B. Le Bouquin, P. Léna, D. Lutz, T. Ott, T. Paumard, K. Perraut,

G. Perrin, O. Pfuhl, S. Rabien, D. C. Ribeiro, M. Sadun Bordoni, S. Scheithauer, J. Shangguan, T. Shimizu, J. Stadler, C. Straubmeier, E. Sturm, L. J. Tacconi, F. Vincent, S. von Fellenberg, F. Widmann, E. Wieprecht, E. Wiezorrek, J. Woillez, and S. Yazici. Where intermediate-mass black holes could hide in the Galactic Centre: A full parameter study with the S2 orbit. *arXiv e-prints*, art. arXiv:2303.04067, Mar. 2023. doi: 10.48550/arXiv.2303.04067.

G. Torrealba, M. Catelan, A. J. Drake, S. G. Djorgovski, R. H. McNaught, V. Belokurov, S. Koposov, M. J. Graham, A. Mahabal, S. Larson, and E. Christensen. Discovery of ∼9000 new RR Lyrae in the southern Catalina surveys. *Monthly Notices of the Royal Astronomical Society*, 446(3): 2251–2266, Jan. 2015. doi: 10.1093/mnras/stu2274.

M. Trenti, E. Ardi, S. Mineshige, and P. Hut. Star clusters with primordial binaries–iii. dynamical interaction between binaries and an intermediate-mass black hole. *Monthly Notices of the Royal Astronomical Society*, 374(3):857–866, 2007.

P. Trevisan, M. Pasquato, A. Ballone, and M. Mapelli. Measuring the spectral index of turbulent gas with deep learning from projected density maps. *Monthly Notices of the Royal Astronomical Society*, 498(4):5798–5803, Nov. 2020. doi: 10.1093/mnras/staa2663.

P. Trevisan, M. Pasquato, G. Carenini, N. Mekhaël, V. F. Braga, G. Bono, and M. Abbas. Sparse logistic regression for rr lyrae versus binaries classification. *The Astrophysical Journal*, 950(2):103, 2023.

S. Umbreit and F. A. Rasio. Constraining Intermediate-mass Black Holes in Globular Clusters. *The Astrophysical Journal*, 768:26, May 2013. doi: 10.1088/0004-637X/768/1/26.

B. Van der Pol. The nonlinear theory of electric oscillations. *Proceedings of the Institute of Radio Engineers*, 22(9):1051–1086, 1934.

S. N. van der Veer, L. Riste, S. Cheraghi-Sohi, D. L. Phipps, M. P. Tully, K. Bozentko, S. Atwood, A. Hubbard, C. Wiper, M. Oswald, et al. Trading off accuracy and explainability in ai decision-making: findings from 2 citizens' juries. *Journal of the American Medical Informatics Association*, 28(10):2128–2138, 2021.

D. Wadekar, L. Thiele, F. Villaescusa-Navarro, J. C. Hill, M. Cranmer, D. N. Spergel, N. Battaglia, D. Anglés-Alcázar, L. Hernquist, and S. Ho. Augmenting astrophysical scaling relations with machine learning: Application to reducing the sunyaev–zeldovich flux–mass scatter. *Proceedings of the National Academy of Sciences*, 120(12):e2202074120, 2023.

L. Wang, R. Spurzem, S. Aarseth, M. Giersz, A. Askar, P. Berczik, T. Naab, R. Schadow, and M. B. N. Kouwenhoven. The DRAGON simulations: globular cluster evolution with a million stars. *Monthly Notices of the Royal Astronomical Society*, 458:1450–1465, May 2016. doi: 10.1093/mnras/stw274.

D. L. Welch and P. B. Stetson. Robust Variable Star Detection Techniques Suitable for Automated searches: New Results for NGC 1866. *The Astronomical Journal*, 105:1813, May 1993. doi: 10.1086/116556.

L. M. Widrow, K. Darling, and H. Li. Warps, Waves, and Phase Spirals in the Milky Way. In M. Valluri and J. A. Sellwood, editors, *Galactic Dynamics in the Era of Large Surveys*, volume 353, pages 65–70, Jan. 2020. doi: 10.1017/S1743921319009049.

T. E. Woods, B. Agarwal, V. Bromm, A. Bunker, K.-J. Chen, S. Chon, A. Ferrara, S. C. O. Glover, L. Haemmerlé, Z. Haiman, T. Hartwig, A. Heger, S. Hirano, T. Hosokawa, K. Inayoshi, R. S. Klessen, C. Kobayashi, F. Koliopanos, M. A. Latif, Y. Li, L. Mayer, M. Mezcua, P. Natarajan, F. Pacucci, M. J. Rees, J. A. Regan, Y. Sakurai, S. Salvadori, R. Schneider, M. Surace, T. L. Tanaka, D. J. Whalen, and N. Yoshida. Titans of the early Universe: The Prato statement on the origin of the first supermassive black holes. *Publications of the Astronomical Society of Australia*, 36:e027, Aug. 2019. doi: 10.1017/pasa.2019.14.

S. E. Woosley, A. Heger, and T. A. Weaver. The evolution and explosion of massive stars. *Reviews of modern physics*, 74(4):1015, 2002.

J. Yang, K. Zhou, Y. Li, and Z. Liu. Generalized Out-of-Distribution Detection: A Survey. *arXiv e-prints*, art. arXiv:2110.11334, Oct. 2021. doi: 10.48550/arXiv.2110.11334.

K. Zhang and J. S. Bloom. Classification of periodic variable stars with novel cyclic-permutation invariant neural networks. *Monthly Notices of the Royal Astronomical Society*, 505(1):515–522, July 2021. doi: 10.1093/mnras/stab1248.

L. Zhang and H. Schaeffer. On the convergence of the sindy algorithm. *Multiscale Modeling & Simulation*, 17(3):948–972, 2019.

Appendix A

Appendix A: Other results from the penalized classifier work

A.1 Performance with principal components

Table A.1: Different metrics for different models trained on the first 4 principal components. The L1 Log Regression is the model that have C=0.056.

Model	Precision (RRc)	Recall (RRc)	Specificity	AUC	F1 Score
Logistic Regression	0.82	0.71	0.98	0.85	0.76
L1 Logistic Regression	0.82	0.70	0.98	0.84	0.75
RandomForest	0.79	0.75	0.98	0.86	0.77
MLP	0.77	0.80	0.97	0.89	0.79
Decision Tree	0.68	0.74	0.96	0.85	0.71

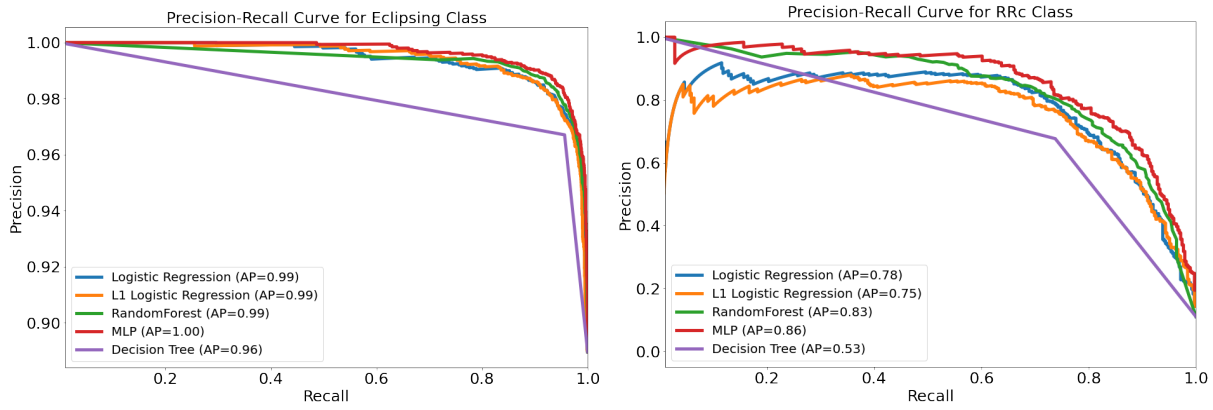


Figure A.1: Precision-recall curve colorcoded for the different models with principal components as training data for both the RRc and EB classes respectively on the left and right panel. When working with unbalanced dataset a precision-recall curve is preferred to a ROC curve.

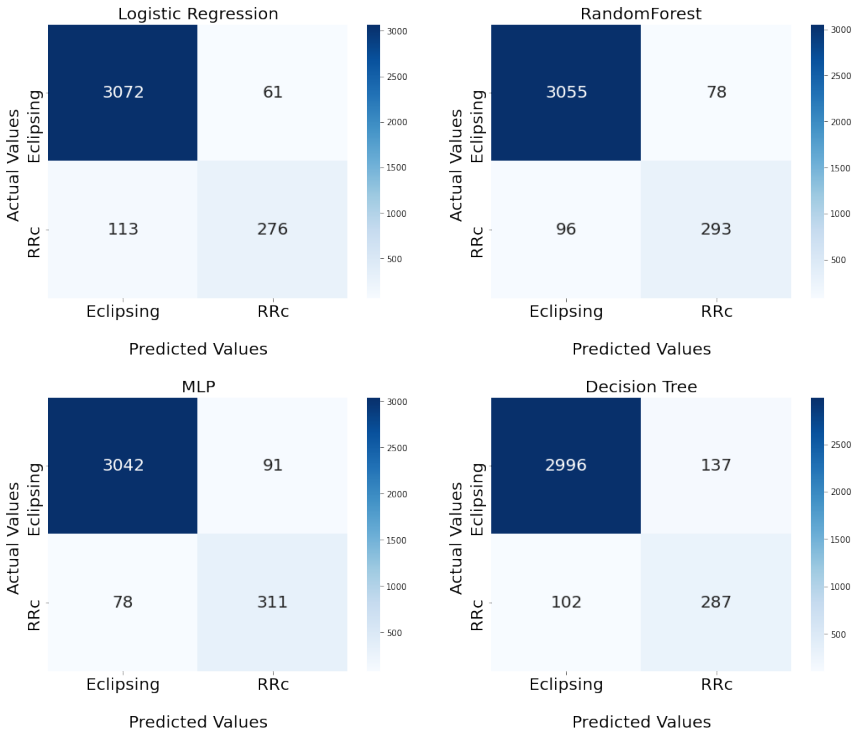


Figure A.2: Confusion matrix for the different models trained on the first 4 principal components. The results are referring to the validation dataset.

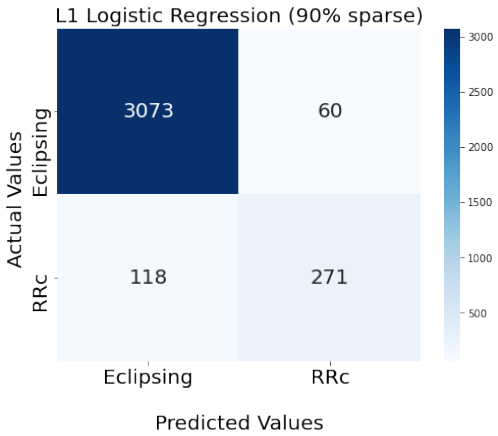


Figure A.3: Confusion matrix for the penalized logistic regression trained on the first 4 principal components. The results are referring to the validation dataset.

PCA are also good for classification but since they are built from a linear combination of *all* features, even if we are classifying with a single PCA we are not gaining a lot of interpretability.

A.2 Performances with information on period and amplitude

Table A.2: Different metrics for different models with the additional information about amplitude and period. The L1 Log Regression model have the same light-curve features as the ones described in the text with the addition of period and amplitude features. We can see that we have a small increment in performance likely due to the stars not in the critical region shown in Fig.2.13.

Model	Precision (RRc)	Recall (RRc)	Specificity	AUC	F1 Score
Logistic Regression	0.92	0.79	0.99	0.89	0.85
L1 Logistic Regression	0.90	0.67	0.99	0.83	0.77
RandomForest	0.93	0.82	0.99	0.91	0.87
MLP	0.93	0.83	0.99	0.91	0.88
Decision Tree	0.80	0.84	0.97	0.91	0.82

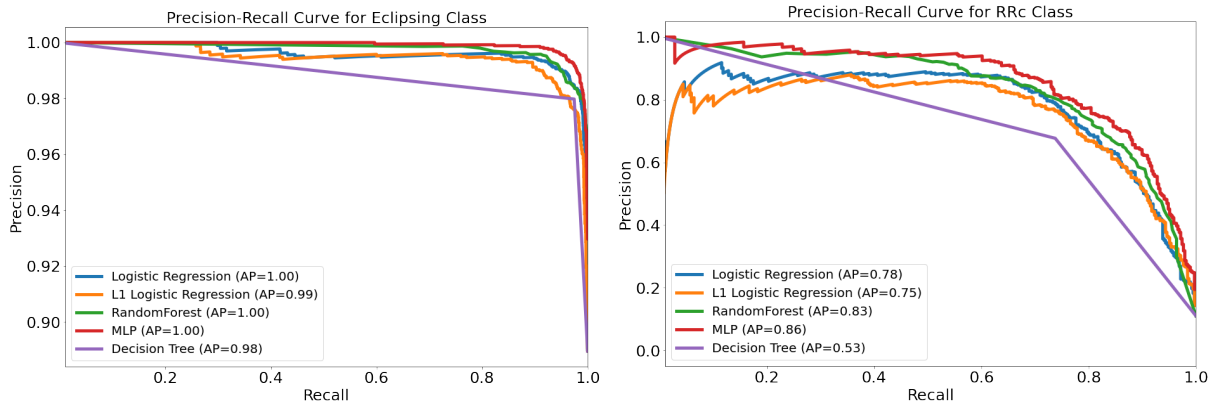


Figure A.4: Precision-recall curve colorcoded for the different models, denoted in the subplot titles, with principal components as training data for both the RRc and EB classes respectively on the left and right panel. When working with unbalanced dataset a precision-recall curve is preferred to a ROC curve.

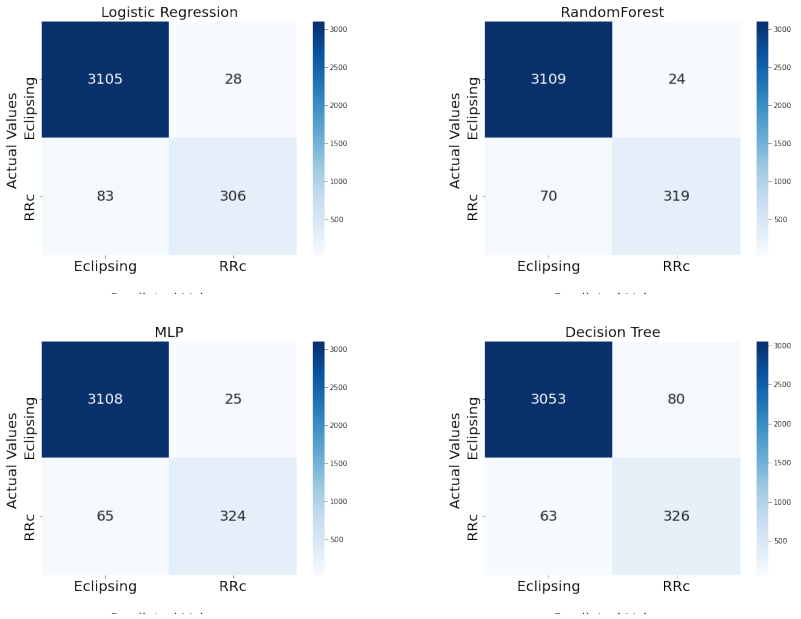


Figure A.5: Confusion matrix for the different models denoted in the subplot titles. with period and amplitude as features as well. The results are referring to the validation dataset.

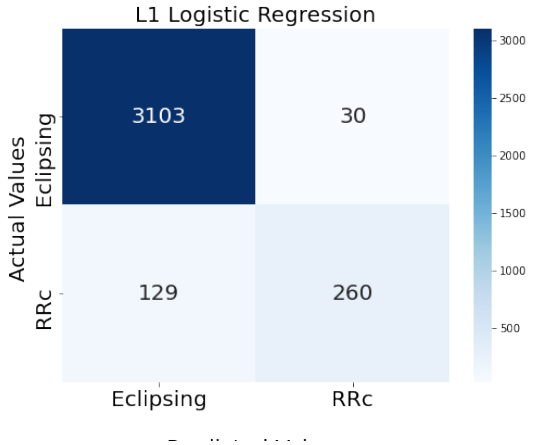


Figure A.6: Confusion matrix for the penalized logistic regression with period and amplitude as features as well. The results are referring to the validation dataset.

Appendix B

Appendix B: Globular Clusters

B.1 Full list of candidates

Table A1: Whole sample of star clusters including deemed non-hosts. The ordering of the star clusters in the table is decreasing with 'final score' (column 2). This is the XGBoost predicted probability for the star clusters that were classified as hosts by CORELS (column 9) and XGBoost (column 10) and were deemed in-distribution based on our KDE criterion (column 11). We also report cluster features (columns 3–8).

Cluster	Final score	HRT	TVL	CVD	CSB	OCR	OHLR	CORELS	XGB	KDE	Pred. prob.
NGC_6569	0.9999	3.06	5.02	0.86	3.68	0.04	0.41	True	True	True	0.9999
Pal_6	0.9998	2.67	4.79	0.72	3.64	-0.02	0.35	True	True	True	0.9998
NGC_6638	0.9998	2.49	4.9	0.84	4.21	-0.34	0.27	True	True	True	0.9998
NGC_6333	0.9996	3.22	5.18	0.92	3.9	-0.04	0.44	True	True	True	0.9996
NGC_6712	0.9988	2.73	4.78	0.71	3.54	0.03	0.4	True	True	True	0.9988
NGC_6254	0.9985	3.25	5.04	0.79	3.65	0.02	0.47	True	True	True	0.9985
FSR_1735	0.9969	2.63	4.69	0.67	3.83	-0.29	0.33	True	True	True	0.9969
NGC_2298	0.9955	2.65	4.47	0.59	3.55	-0.28	0.38	True	True	True	0.9955
NGC_5986	0.9954	3.25	5.24	0.91	3.83	0.07	0.44	True	True	True	0.9954
NGC_6218	0.9947	2.94	4.75	0.68	3.34	0.08	0.45	True	True	True	0.9947
NGC_6316	0.9913	3.27	5.09	0.91	4.04	-0.22	0.47	True	True	True	0.9913
VVV-CL001	0.9889	2.75	4.54	0.81	3.65	-0.27	0.34	True	True	True	0.9889
Ton_2	0.9889	3.01	4.51	0.53	3.04	0.06	0.46	True	True	True	0.9889
NGC_6352	0.9871	2.96	4.47	0.54	3.06	0.05	0.46	True	True	True	0.9871
NGC_6779	0.9845	3.17	5.03	0.78	3.59	0.07	0.47	True	True	True	0.9845
NGC_6171	0.9843	2.87	4.53	0.61	3.32	-0.08	0.46	True	True	True	0.9843
NGC_6553	0.9837	3.26	4.95	0.88	3.67	-0.06	0.37	True	True	True	0.9837
NGC_6934	0.9815	3.26	4.95	0.68	3.46	0.09	0.47	True	True	True	0.9815
NGC_6637	0.9724	2.95	4.95	0.79	3.76	-0.07	0.38	True	True	True	0.9724
UKS_1	0.9518	2.85	4.63	0.62	3.29	0.04	0.46	True	True	True	0.9518
Ter_1	0.9502	2.54	4.94	0.91	4.44	-0.49	0.18	True	True	True	0.9502
NGC_6342	0.9288	2.22	4.36	0.64	3.85	-0.47	0.17	True	True	True	0.9288
HP_1	0.8894	2.89	4.58	0.72	3.18	0.1	0.46	True	True	True	0.8894
NGC_6401	0.5381	2.86	4.65	0.81	3.73	-0.28	0.39	True	True	True	0.5381
NGC_6440	0	2.78	5.5	1.16	4.96	-0.46	0.1	True	False	False	0.1708
Djor_2	0	3.03	4.54	0.63	3.04	0.08	0.63	False	True	True	0.9951
NGC_6539	0	3.22	5.04	0.77	3.46	0.18	0.57	False	True	True	0.9998
NGC_6535	0	2.54	3.92	0.43	3.88	-1.1	0.43	False	False	True	0.0253
NGC_6388	0	3.49	5.76	1.24	4.96	-0.38	0.41	False	True	False	0.6810
NGC_7006	0	3.46	4.93	0.61	3.19	0.2	0.63	False	True	True	0.6639
NGC_6528	0	2.46	4.5	0.67	3.93	-0.48	0.28	False	True	True	0.9979
Laevens_3	0	2.86	3.07	-0.4	0.95	0.43	0.85	False	False	False	0.1948
NGC_6522	0	2.85	4.94	0.91	4.19	-0.38	0.39	False	False	True	0.0972
Djor_1	0	3.11	4.64	0.56	3.22	-0.01	0.61	False	True	True	0.9980
NGC_7078	0	3.47	5.6	1.12	5.85	-1.4	0.31	False	False	False	0.0020
Ter_10	0	3.22	4.74	0.94	4.1	-0.59	0.54	False	False	True	0.2129
NGC_6517	0	2.64	4.92	1	4.85	-0.89	0.23	False	True	False	0.9365
Ter_9	0	2.38	4.57	0.89	4.23	-0.6	0.2	False	False	True	0.1856
NGC_6402	0	3.49	5.45	1	3.67	0.33	0.55	False	False	True	0.0021
NGC_6397	0	2.94	4.76	0.72	5.07	-1.4	0.33	False	True	False	0.9645
Pal_13	0	3.26	3.19	-0.4	1.02	0.29	1.11	False	False	False	0.4264
NGC_6496	0	3.18	4.63	0.48	2.69	0.4	0.64	False	False	True	0.0069
NGC_7089	0	3.44	5.55	1.04	4.34	-0.11	0.48	False	True	True	0.9914
NGC_6540	0	3.09	4.39	0.4	4.23	-1.22	0.42	False	False	True	0.0212
NGC_6453	0	3	5.05	0.86	4.94	-1	0.44	False	True	False	0.9967
NGC_7099	0	3.28	4.87	0.75	4.8	-1.22	0.4	False	False	False	0.0417
Ter_6	0	2.07	4.77	0.92	4.45	-0.51	0.01	True	True	False	0.6464
Pal_12	0	3.11	3.69	-0.15	1.36	0.59	0.84	False	True	False	0.8911

NGC_6441	0	3.33	5.87	1.28	5	-0.28	0.32	False	True	False	0.8584
Ter_5	0	3.35	5.57	1.19	4.64	-0.29	0.27	False	True	True	0.9572
NGC_6426	0	3.43	4.56	0.45	2.48	0.45	0.71	False	False	True	0.0035
NGC_6838	0	3.32	4.58	0.45	2.87	0.17	0.53	False	False	True	0.1592
NGC_6541	0	3.29	5.22	0.93	5.07	-1.05	0.36	False	False	False	0.0490
NGC_6681	0	2.71	4.8	0.85	5.12	-1.15	0.33	False	True	False	0.9957
Pal_11	0	2.94	3.98	0.08	1.75	0.61	0.76	False	True	False	0.9989
Ter_8	0	4.12	4.54	0.2	1.47	1.03	1.18	False	False	False	0.0024
NGC_6809	0	3.55	4.97	0.69	2.95	0.46	0.66	False	False	True	0.0001
Arp_2	0	3.77	4.31	0.11	1.43	0.89	1.15	False	False	False	0.0093
Pal_10	0	3.35	4.94	0.66	3.09	0.32	0.6	False	False	True	0.0040
Ter_7	0	3.69	4.05	0.08	1.57	0.63	0.8	False	False	False	0.0116
NGC_6760	0	3.28	5.17	0.83	3.72	0.03	0.51	False	True	True	0.9649
NGC_6752	0	3.4	5.07	0.88	4.83	-0.89	0.46	False	True	True	0.6079
NGC_6749	0	3.51	5.18	0.71	3.47	0.16	0.66	False	True	True	0.9098
NGC_6723	0	3.21	4.94	0.74	3.3	0.23	0.55	False	False	True	0.1026
NGC_6717	0	2.73	4.36	0.52	3.9	-0.7	0.56	False	False	True	0.1662
NGC_6715	0	3.65	5.93	1.28	4.93	-0.27	0.55	False	True	False	0.9831
NGC_6864	0	2.95	5.35	1.05	4.69	-0.42	0.31	False	True	True	0.9711
RLGC_2	0	3.23	5.13	0.88	4.76	-0.62	0.53	False	True	True	0.9972
Pal_8	0	3.15	4.03	0.49	2.25	0.29	0.54	False	False	True	0.0025
NGC_6544	0	2.36	4.55	0.85	4.49	-0.8	0.18	False	False	False	0.2371
NGC_6656	0	3.52	5.37	0.95	3.99	0.02	0.5	False	True	True	0.9654
NGC_6652	0	2.19	4.43	0.69	4.1	-0.59	0.17	False	False	True	0.3098
NGC_6642	0	1.92	4.45	0.68	4.41	-0.77	0.13	False	False	False	0.3241
NGC_6626	0	2.72	5.15	1.05	4.67	-0.54	0.2	False	True	True	0.9927
NGC_6624	0	2.91	4.87	0.85	4.37	-0.48	0.37	False	False	True	0.3324
NGC_6584	0	3.2	4.88	0.6	3.24	0.19	0.54	False	True	True	0.7763
BH_261	0	2.92	3.45	0.32	2.57	-0.47	0.42	False	False	True	0.0000
NGC_6981	0	3.15	4.71	0.49	2.89	0.32	0.62	False	False	True	0.1884
Ter_12	0	2.89	4.37	0.63	3.15	-0.08	0.24	True	False	True	0.4512
IC_1276	0	3.07	4.61	0.56	3.12	0.07	0.5	False	True	True	0.9385
NGC_6558	0	1.95	4.36	0.61	4.33	-0.77	0.17	False	True	False	0.6577
2MASS-GC02	0	2.31	3.94	0.34	2.98	-0.2	0.34	False	False	True	0.3450
ESO_280-SC06	0	3.08	3.66	-0.1	1.57	0.38	0.85	False	True	False	0.5992
Sagittarius_II	0	4.23	4.1	-0.22	0.47	1.31	1.46	False	False	False	0.0020
2MASS-GC01	0	2.81	4.29	0.43	3.28	-0.3	0.55	False	False	True	0.0202
NGC_104	0	3.73	5.66	1.08	4.34	-0.1	0.56	False	True	True	0.9550
NGC_6380	0	3.2	5.06	0.91	3.68	0.05	0.5	False	True	True	0.9987
NGC_4147	0	2.74	4.38	0.51	3.62	-0.44	0.4	False	False	True	0.0800
Rup_106	0	3.58	4.39	0.2	1.9	0.71	0.88	False	False	True	0.0298
NGC_4590	0	3.53	4.83	0.57	2.96	0.29	0.65	False	False	True	0.0003
BH_140	0	3.48	4.38	0.38	2	0.68	0.79	False	False	True	0.0008
NGC_4833	0	3.22	5.07	0.79	3.49	0.19	0.51	False	True	True	0.9994
NGC_5024	0	3.98	5.41	0.8	3.36	0.36	0.81	False	False	True	0.0005
NGC_5053	0	3.97	4.46	0.3	1.49	1.01	1.09	False	False	False	0.0000
NGC_5139	0	4.4	6.15	1.25	3.82	0.63	0.88	False	False	False	0.0127
NGC_5272	0	3.66	5.4	0.88	3.87	0.09	0.53	False	True	True	0.9363
NGC_5286	0	3.18	5.38	0.99	4.6	-0.4	0.37	False	False	True	0.4562
AM_4	0	3.18	2.32	-0.7	0.12	0.33	0.8	False	True	False	0.7439
NGC_5466	0	3.81	4.62	0.28	1.93	0.81	0.98	False	False	False	0.0001
NGC_5634	0	3.57	5.02	0.75	3.79	-0.12	0.65	False	True	True	0.9919
NGC_5694	0	3.29	5.18	0.96	5.38	-1.4	0.49	False	True	False	0.8416
IC_4499	0	4.02	4.83	0.48	2.22	0.71	1.01	False	False	True	0.0000
NGC_5824	0	3.66	5.57	1.05	4.42	-0.18	0.65	False	True	True	0.9533
NGC_4372	0	3.68	4.98	0.65	2.74	0.58	0.76	False	False	True	0.0002
Crater	0	3.78	3.97	-0.22	0.74	1.1	1.29	False	False	False	0.0562
NGC_5897	0	3.82	4.85	0.54	2.34	0.74	0.88	False	False	True	0.0000
Pal_4	0	3.72	4.28	-0.15	1.17	1.06	1.2	False	False	False	0.1473
NGC_362	0	3.16	5.29	0.93	4.49	-0.4	0.34	False	True	True	0.7864
Whiting_1	0	3.18	3.6	-0.4	1.36	0.39	0.98	False	True	False	0.5067
NGC_1261	0	3.3	5.05	0.76	3.79	-0.06	0.51	False	True	True	0.8986
Pal_1	0	1.89	2.63	-0.3	1.56	-0.16	0.26	False	True	False	0.9337
AM_1	0	3.63	4.37	-0.05	1.54	0.82	1.17	False	False	False	0.1989
Eridanus	0	3.48	4.1	-0.15	1.29	0.83	1.14	False	False	False	0.3368
Pal_2	0	3.67	5.24	0.71	3.44	0.21	0.68	False	False	True	0.1936
NGC_1851	0	2.95	5.28	1.04	5.32	-1.1	0.24	False	True	False	0.6323
NGC_1904	0	2.81	5	0.81	4.3	-0.38	0.4	False	True	True	0.8813
NGC_2419	0	4.63	5.69	0.76	2.66	0.91	1.29	False	False	False	0.0000
Pyxis	0	3.9	4.13	0	0.93	1.13	1.23	False	False	False	0.0135
NGC_2808	0	3.37	5.76	1.15	4.62	-0.1	0.39	False	True	False	0.8892
E_3	0	2.56	3.35	-0.22	1.59	0.26	0.63	False	True	False	0.9953
Pal_3	0	3.95	4.11	-0.1	0.8	1.16	1.31	False	False	False	0.0008
NGC_3201	0	3.51	4.87	0.66	3.33	0.09	0.58	False	False	True	0.0182
Pal_5	0	3.87	3.91	-0.22	0.57	1.17	1.31	False	False	False	0.0028
NGC_5904	0	3.51	5.34	0.9	3.86	0.07	0.55	False	True	True	0.6028
Liller_1	0	2.85	5.53	1.34	5.33	-0.72	0.05	False	False	False	0.1790
NGC_6266	0	2.91	5.49	1.18	4.85	-0.41	0.25	True	True	False	0.9919
NGC_6284	0	2.89	4.99	0.77	4.14	-0.33	0.47	False	True	True	0.9988
NGC_6287	0	3.01	4.71	0.8	4.21	-0.68	0.29	False	True	True	0.7655
NGC_6293	0	3.16	4.91	0.88	4.61	-0.89	0.37	False	False	True	0.4187
NGC_6304	0	3.1	4.9	0.74	4.38	-0.66	0.29	False	True	True	0.7501
NGC_6325	0	2.24	4.45	0.72	4.05	-0.51	0.23	False	False	True	0.3542

NGC_6341	0	3.39	5.21	0.92	4.01	-0.11	0.38	False	True	True	0.6153
NGC_6355	0	2.86	4.79	0.74	4.34	-0.68	0.37	False	False	True	0.0326
NGC_6356	0	3.72	5.34	0.94	3.64	0.17	0.59	False	True	True	0.5412
IC_1257	0	2.85	4.16	0.23	2.66	0.06	0.61	False	True	True	0.7621
Ter_2	0	3.03	4.78	0.76	3.95	-0.38	0.39	False	True	True	0.8764
Ter_4	0	3.43	4.85	0.74	3.67	-0.24	0.54	False	False	True	0.4998
NGC_288	0	3.47	4.64	0.49	2.43	0.56	0.77	False	False	True	0.0024
FSR_1758	0	4.41	5.6	0.76	2.63	1.03	1.1	False	False	False	0.0000
NGC_6362	0	3.42	4.83	0.59	2.78	0.46	0.71	False	False	True	0.0010
NGC_6366	0	2.99	4.34	0.38	2.58	0.29	0.58	False	False	True	0.0010
NGC_6273	0	3.3	5.49	1.08	4.19	0	0.5	False	True	True	0.9990
Pal_15	0	4.1	4.16	0.11	0.86	1.16	1.3	False	False	False	0.0003
NGC_5927	0	3.31	5.23	0.83	3.68	0.15	0.54	False	True	True	0.6543
NGC_6256	0	3.15	4.77	0.73	4.9	-1.4	0.47	False	True	False	0.8825
NGC_5946	0	2.59	4.77	0.78	4.22	-0.51	0.29	False	True	True	0.9940
FSR_1716	0	3.01	3.84	0.52	2.13	0.28	0.56	False	False	True	0.0063
Pal_14	0	4.1	4.05	-0.15	0.55	1.21	1.44	False	False	False	0.0006
Lynga_7	0	2.99	4.68	0.56	2.91	0.33	0.61	False	False	True	0.2345
NGC_6093	0	2.86	5.21	1.04	4.53	-0.39	0.25	True	False	True	0.4276
RLGC_1	0	3.9	5.22	0.66	3	0.45	0.9	False	False	True	0.0000
NGC_6101	0	4.04	4.84	0.52	2.21	0.77	0.98	False	False	True	0.0000
NGC_6121	0	2.87	4.74	0.68	3.96	-0.41	0.4	False	True	True	0.6485
NGC_6139	0	2.79	5.27	1.06	4.87	-0.57	0.27	False	True	False	0.8153
NGC_6144	0	3.01	4.5	0.57	2.88	0.21	0.56	False	True	True	0.7837
Ter_3	0	3.17	4.36	0.34	2.32	0.47	0.67	False	True	True	0.5244
ESO_452-SC11	0	2.41	3.53	0.15	2.42	-0.14	0.4	False	False	True	0.2605
NGC_6205	0	3.52	5.37	0.97	3.66	0.28	0.54	False	False	True	0.0004
NGC_6229	0	3.21	5.17	0.88	3.8	0.02	0.49	False	True	True	0.9999
NGC_6235	0	3.09	4.62	0.64	3.15	0.08	0.52	False	True	True	0.9363
NGC_7492	0	3.29	4.29	0.18	1.89	0.67	0.88	False	True	True	0.8244

Appendix C

Appendix C: Explainer Dashboard for XGBoost

During my PhD period, I experimented with both the developing and the use of dashboard with Dash+Plotly: Plotly (Inc., 2015) is the library we use to plot and Dash[1] is the framework we use to generate interactive dashboards from those plots. Thesere were used both for data exploration and for model comprehension. Sadly, interactive plots do not transfer well to static plots for standard literature. As remarked by Molnar (2022), human friendly explanations should also be interactive. For example, we want to easily understand what happen to the predictions if I change some value of the features: partial dependence plot (Greenwell et al., 2018) and SHAP values (SHapley Additive exPlanations i.e. what is the contributions of each feature to each individual prediction?) by Lundberg and Lee (2017) are good tools to explain individual predictions and get a feeling of the inner working of a model. This can be done by interacting with these explanation in a dynamic way. In the end, in the main work, we decided to opt for anchors for the immediate human-understandability and anchors produce if-then rules that were also the main ingredient for the natively interpretable model (CORELs).

[1] https://dash.plotly.com/.

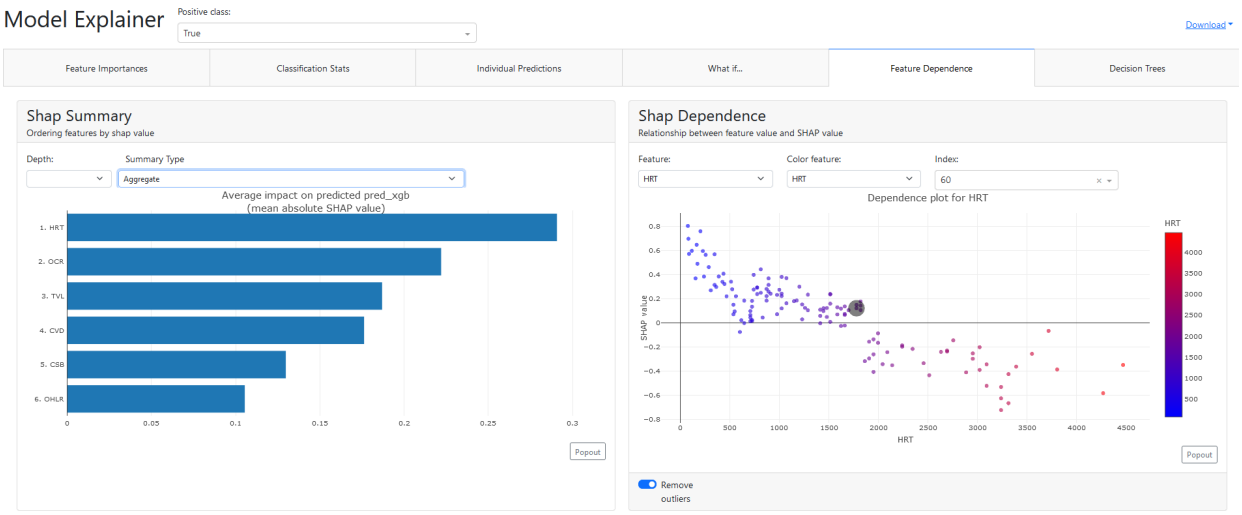


Figure C.1: Screenshots from the explainer dashboard applied to XGBoost model to the real clusters: as we can see we can see the dashboard have different sections such as feature importances, classification stats, individual predictions, what if, feature dependence and also decision tree. Shap summary (left): shows summarizes the shap values per feature. Can be either sorted by absolute SHAP value (average absolute impact of the feature on final prediction) or by permutation importance (how much does the model get worse when you shuffle this feature, rendering it useless?). Instead SHAP dependence (right) shows the relation between feature values and SHAP values. This allows to investigate the general relationship between feature value and impact on the prediction. In the dashboard you can check whether the model uses features in line with your intuitions, or use the plots to learn about the relationships that the model has learned between the input features and the predicted outcome. In the example we are looking at feature:HRT and the grey area is the value for index=60 corresponding to M10.

Acknowledgements

Vorrei ringraziare i miei relatori per i consigli in questi 3 anni. Vorrei ringraziare in particolare Gaia per la sua disponibilità e incredibile lucidità. Vorrei ringraziare la grandiosa città di Thytzë, gli abitanti di Srania e i miei pazzi amici che hanno accettato di entrarci e hanno colorarato e reso vivo quel mondo. Da Andre compagno di corse e riflessioni cosmiche sull'insegnamento, creare storie/mondi/sistemidigiocopiùsemplicimanoncosìsemplici e la vita in generale. A Nica e la sua genuinità e forza d'animo. Da Ame e la sua risata che trascina tutto. A Cande best reparto risorse umane in game e personal trainer IRL. A Raboso catalizzatore di eventi, insegnante dei mille usi di un ippopotamo e in generale fratello con cui condividere tantissime cose. Robolo che con la sua gentilezza e saggezza rende tutti un po' migliori. Vorrei anche ringraziare Marty, Fabio, Marty Rany, Giacomo, Ericattack, Bobia, Daniele, laFede, UNDC, Marco Mastrofini, Andy, miniDely per le risate che mi hanno accompagnato in questi 3 anni. Vorrei ringraziare la mia famiglia da Simbona alla nuova arrivata Thea. Vorrei ringraziare la nuova famiglia acquisita che mi ha accolto nel loro colorato mondo. Il Quaternais: la vera Thytzë. Il rin di Padolina. La maglia glutinica. I bacchettei e in generale le levadas nel mondo. Giulo e le transizioni nei gatti, Muschio e Ghianda per aver fatto esplodere il cuore e aver reso l'angolo più magico di Padova ancora più straordinario. Vorrei ringraziare chiaramente scryfall, Arixmethes turno 3 e in generale il tappastappa. Garp. Could it be a "Gianpaolo Astrorazzo". Gli alberi "belli" del Morandi. Precottini. Le smart sponsorizzate Lidl. La difesa Caro-Kann e la funzione "muta" di Teams. Dodie, Jacob, VALERIOOOO. La maschera dell'uomo tigre.

E ovviamente la mia patata, il mio kalsifer e allo stesso tempo Stromboli, la persona che rende tutto così un'eccezione e tutto la normalità, la mia "casa" e la mia "famiglia" nonché la mia migliore amica. CTC che rompe la realtà facendo ridere come non mai. Senza di lei, non sarei mai riuscito a finire questi tre anni.