The Astrophysical Journal, 930:161 (13pp), 2022 May 10 https://doi.org/10.3847/1538-4357/ac5624

Sparse Identification of Variable Star Dynamics

Mario Pasquato[1,2,3,4,5], Mohamad Abbas[3,4], Alessandro A. Trani[6,7], Matteo Nori[3,4], James A. Kwiecinski[8], Piero Trevisan[9], Vittorio F. Braga[10,11], Giuseppe Bono[9], and Andrea V. Macciò[3,4,12]
[1] Physics and Astronomy Department Galileo Galilei, University of Padova, Vicolo dell'Osservatorio 3, I-35122, Padova, Italy
[2] Département de Physique, Université de Montréal, Montreal, QC H3T 1J4, Canada
[3] Center for Astro, Particle and Planetary Physics (CAP^3), New York University Abu Dhabi
[4] Physic Department, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates
[5] INFN- Sezione di Padova, Via Marzolo 8, I-35131 Padova, Italy
[6] Department of Earth Science and Astronomy, College of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan
[7] Quantum Gravity Unit, Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan
[8] Mathematics, Mechanics, and Materials Unit, Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna-son, Okinawa 904-0495, Japan
[9] Department of Physics, Università di Roma Tor Vergata, via della Ricerca Scientifica 1, I-00133 Roma, Italy
[10] INAF-Osservatorio Astronomico di Roma, via Frascati 33, I-00040 Monte Porzio Catone, Italy
[11] Space Science Data Center, via del Politecnico snc, I-00133 Roma, Italy
[12] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
Received 2021 December 18; revised 2022 February 15; accepted 2022 February 15; published 2022 May 16

Abstract

Variable stars play a crucial role as standard candles and provide valuable insights into stellar physics. They can be modeled either through fully fledged hydrodynamical simulations or analytically as systems of coupled differential equations describing the evolution of relevant physical quantities. Typically, such equations are arrived at by simplified physical assumptions concerning the conservation laws governing stellar interiors. Here we apply a data-driven technique—sparse identification of nonlinear dynamics (SINDy)—to automatically learn governing equations from observed light curves. We apply SINDy to 3100 light curves of three different variable types from the Catalina Sky Survey. 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 $\ddot{x} + (a + bx + c\dot{x})\dot{x} + x = 0$. Members of the Lienard class of ordinary differential equations, such as the well-studied van der Pol oscillator, already saw some application to variable star modeling. For $a, b = 0$ the equation can be solved exactly, and it admits both periodic and nonperiodic 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.

Unified Astronomy Thesaurus concepts: RR Lyrae variable stars (1410); RRab variable stars (1413); RRc variable stars (1415); Pulsating variable stars (1307); Delta Scuti variable stars (370)

1. Introduction

Variable stars have historically been modeled through either detailed hydrodynamical simulations (e.g., Bono & Stellingwerf 1994; Bono et al. 1997) or coupled nonlinear differential equations representing simplified physics (Baker et al. 1966; Moore & Spiegel 1966; Stellingwerf 1972, 1986; Icke et al. 1992; Munteanu et al. 2005). The latter have been studied mathematically as dynamical systems (see Buchler 1993, for an early review).

In the following we will focus on RR Lyrae (RRL) and δ Scuti stars. RRL stars are old, short-period variables with radial pulsation periods of $\lesssim$1 day.[13] These core helium-burning stars occupy the instability strip in the Hertzsprung-Russell (H-R) diagram. They have been used as tracers to map different parts of the Milky Way (e.g., Ivezić et al. 2005; Sesar et al. 2010; Abbas et al. 2014; Pietrukowicz et al. 2015; Neeley et al. 2019), enhancing our understanding of galaxy evolution (Simion et al. 2014; Iorio et al. 2018; Kunder et al. 2019; Prudil et al. 2021). The two most common subtypes of RRL stars are the RRab stars that pulsate in the fundamental mode and the RRc stars that pulsate in the first overtone. RRab stars have larger pulsation amplitudes (A) and periods (P) compared to RRc stars. RRab light curves tend to have a sawtooth shape, while RRc light curves are more sinusoidal. δ Scutis are short-period ($P \approx 0.3$ days) 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 & Mateo 1990). Both δ Scuti and SX Phe are located where the main sequence crosses the instability strip in the H-R diagram. δ Scutis pulsate in radial and nonradial 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.

In this paper, we present governing equations learned automatically from light curves of RRL stars and δ Scuti variables. We use the sparse identification of nonlinear dynamics (SINDy; Brunton et al. 2016) method, which discovers equations that accurately reproduce the observed dynamics with as small a number of terms as possible. SINDy is now available as a Python

[13] One day is the more classical threshold, but 1.4 days would be more realistic when taking into account evolution and chemical composition (see, e.g., Braga et al. 2020).

module (de Silva et al. 2020) and applied to numerous fields (e.g., Arzani & Dawson 2020; Guan et al. 2021; Horrocks & Bauch 2020). Unlike three-dimensional hydrodynamical 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 hyperparameters 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 ordinary 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 & Petrovay 2013).

2. The Data

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 is in Australia. Overall, these telescopes scanned $\sim$33,000 deg^2 of the sky in the -75° to $+65^\circ$ decl. 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's exposure time is $\sim$30 s, and it observes objects in the $\sim$11.5–19.5 mag 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 cutoffs are 13.0 and 21.5 mag, respectively. Finally, the SSS is equipped with the smallest 0.5 m Schmidt telescope and an unfiltered 4k $\times$ 4k CCD and has a 4.2 deg^2 FOV. It observes objects in the $\sim$11.0–19.0 mag 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 were 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, 2013b, 2014; Torrealba et al. 2015; Drake et al. 2017).

To search for variable sources, the authors of the latter studies followed a systematic approach. At first, they distinguished variable from nonvariable sources using the Welch–Stetson variability index (I_{WS}; Welch & 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.

The combined catalogs of variable stars result 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$80%, respectively. These values decrease for fainter objects and for RRc and δ Scuti variables. Finally, the identification numbers (ID_{CRTS}), equatorial J2000.0 R.A. (R.A.) and decl. (Dec.) coordinates, average magnitudes ($\langle V \rangle$), periods in days (P), amplitudes (A), number of observations (N_{CRTS}), and other properties are included in the published catalogs. On our adopted sample the median time elapsed between the first and the last observation was 2794 days for δ Scuti variables and 2787 and 2703 days for RRc and RRab variables, respectively. An issue worth noting is that Catalina data mostly do not include dense regions around the Galactic plane, which might affect the generality of our results given the expected dependency of variable star properties on metallicity.

In the following, we adopted periods from the catalogs and phased light curves shifted to mean zero magnitude. The light curves were fitted with a smoothing spline.[14] This approach results in a smooth curve and tolerates a few outliers in the raw data, as shown, e.g., in Figure 2, top left panel. 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. With this approach SINDy will discover systems of governing equations of the first order that are equivalent to a single second order ODE.

We ran our analysis on a subsample of light curves containing 100 δ Scuti variables, 1500 RRab variables, and 1500 RRc variables. The Bailey diagram for the full sample is shown in Figure 1. Six light curves from this sample are shown in Figure 2 with our smoothing spline superimposed. The relevant parameters for each star in Figure 2 are reported in Table 1.

3. Methods

SINDy leverages sparse regression to learn governing equations for dynamical systems from the measured time series of state variables. Sparse regression is a type of regularized regression where the regularization term is chosen to force many coefficients to be exactly zero (Tibshirani 1996). When predicting a dependent variable by linear regression against a large number of independent variables, this results in a much simpler model than conventional least squares (or even other regularized regression approaches) because just a few independent variables enter the equation.

In SINDy, the measured time series of the derivatives of state variables (relevant variables describing the state of a dynamical system) are predicted via sparse regression against a basis of nonlinear functions of said variables. 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. For simplicity and ease of interpretation we chose a polynomial basis truncated to the second power,

[14] 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.

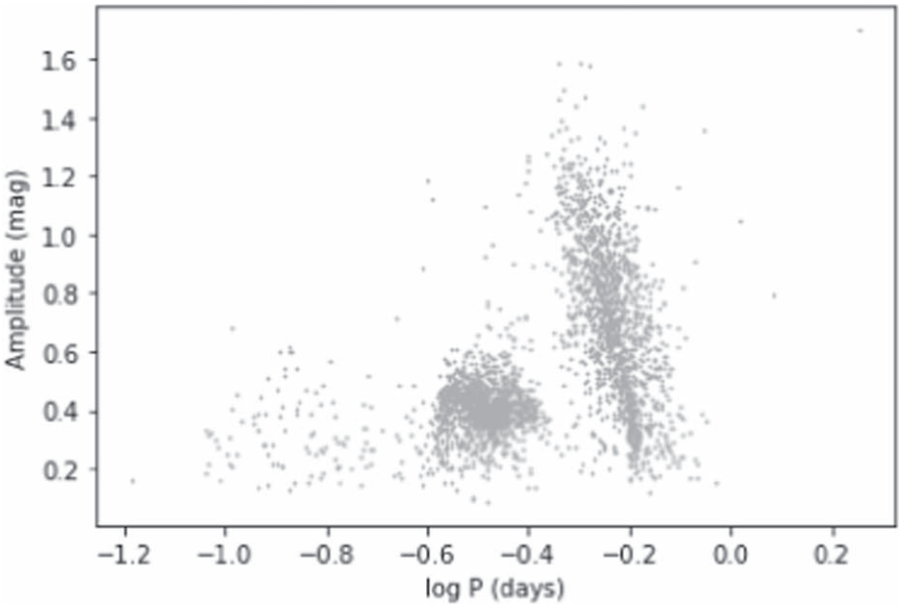

Figure 1. Adopted sample for this study in the period–amplitude plane.

including interaction terms. Thus, the most general system of governing equations that SINDy can learn is in the form

$$\begin{aligned}\dot{x} &= a_{00} + a_{01}x + a_{02}y + a_{03}x^2 \\ &\quad + a_{04}xy + a_{05}y^2 \\ \dot{y} &= a_{10} + a_{11}x + a_{12}y + a_{13}x^2 \\ &\quad + a_{14}xy + a_{15}y^2. \end{aligned} \tag{1}$$

We focused on autonomous governing equations, so time does not appear on the right-hand side of Equations (1).

We used the SINDy implementation for Python *pysindy* (de Silva et al. 2020). Sparse regression was carried out for each light curve using the *Lasso* function from scikit-learn, with regularization parameter $\alpha = 0.01$. Increasing α forces more terms in Equation (1) 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 10 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 10 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 report 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.

4. Results

In Figures 3–5 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 Equation (1) 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 for up to 10 periods of the observational light curve. In each plot we show only the last such period. In Figure 6 we illustrate the effect of slightly perturbing the initial conditions.

The stars presented in Figures 3, 4, and 5 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 Equation (1) 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

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

which reduces to

$$\ddot{u} = -u + (a + bu + c\dot{u})\dot{u} \tag{3}$$

with the substitutions

$$\begin{aligned} u &= (Cx - B)A \\ \Omega &= \sqrt{AC} \\ t' &= \Omega t \\ a &= \frac{1}{\Omega}\left[D + \left(\frac{EB}{C}\right)\right] \\ b &= \frac{E}{\Omega^3} \\ c &= \frac{F}{A\Omega^2}, \end{aligned} \tag{4}$$

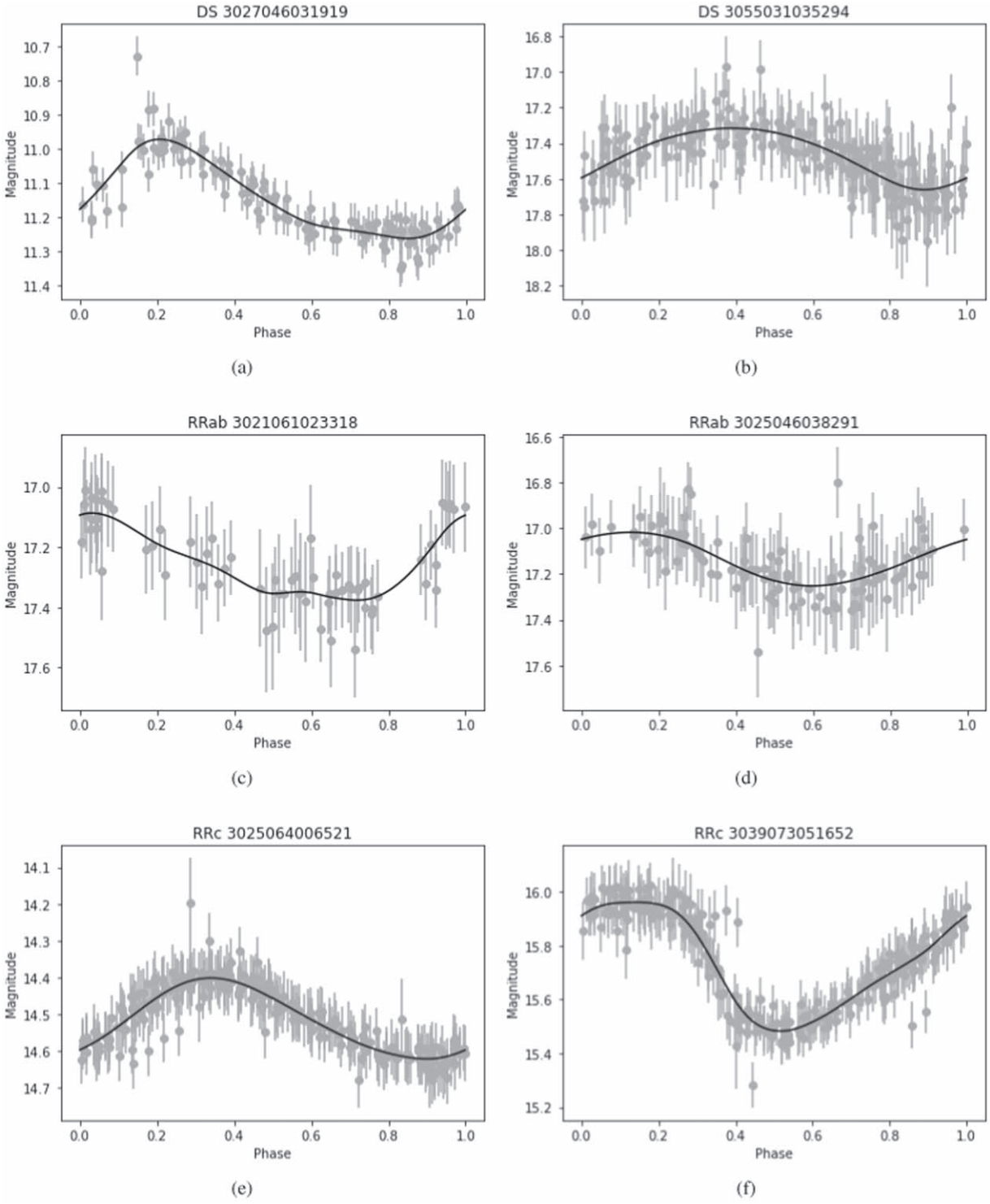

Figure 2. Raw light curves for six stars from our original catalog (gray points with error bars) with our best-fit spline superimposed (black solid curve).

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 (3) with the symbol ∘ (Eastern Arabic numeral for five, pronounced *khamsa*) owing 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 form (or g-∘ form) and the particular case where $a = b = 0$ the simplified form (or s-∘ form). For Equation (2) to reduce to the s-∘ 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.

Equation (3) immediately reveals some interesting properties: neither the g-∘ form nor the s-∘ form is conservative, as can be seen by multiplying both sides by $\dot{u}$, obtaining

$$\frac{1}{2}\frac{d}{dt}[\dot{u}^2 + u^2] = (a + bu + c\dot{u})\dot{u}^2, \tag{5}$$

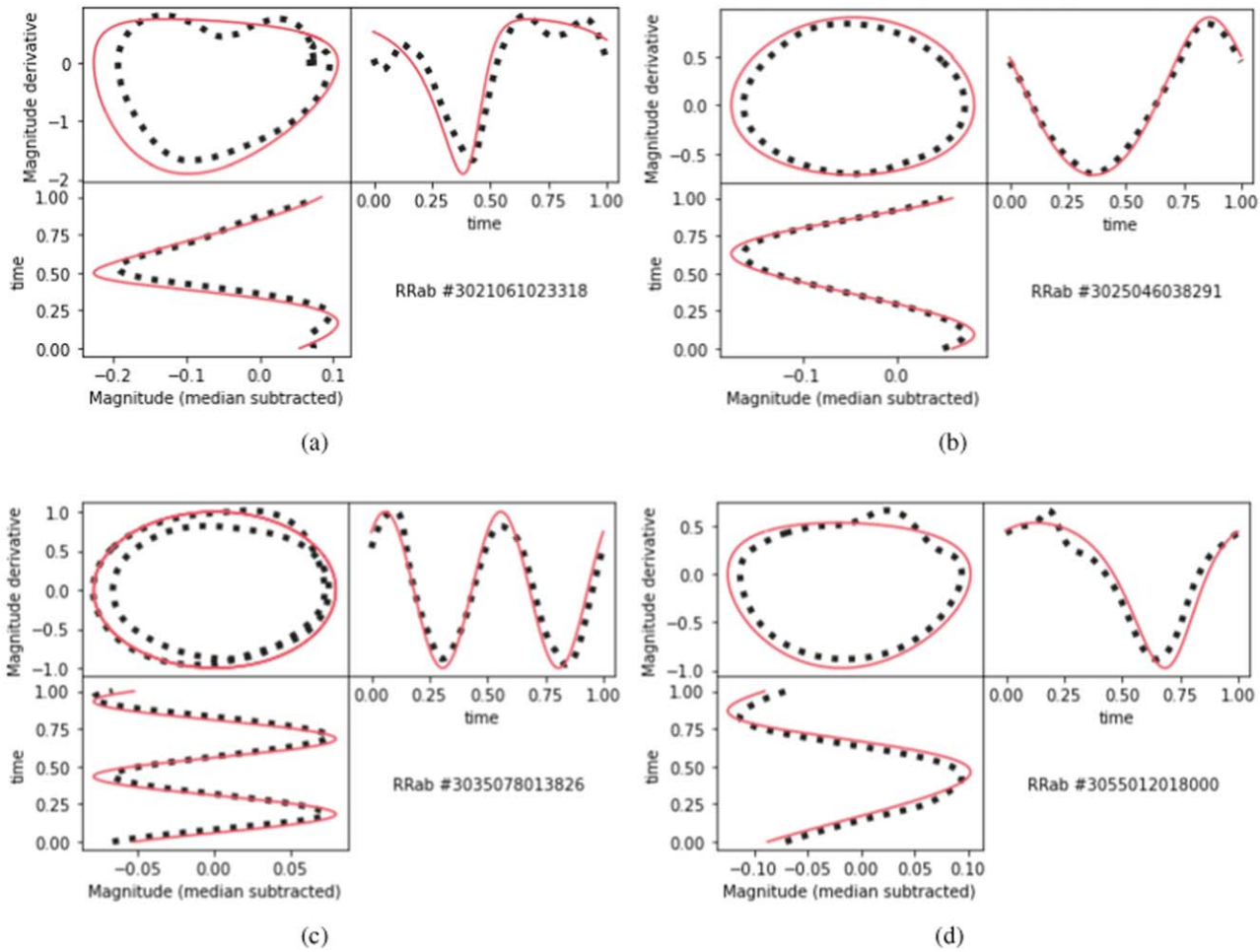

Figure 3. Four light curves for stars classified as RRab (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel tls shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and bottom left panels show the magnitude derivative vs. time and time vs. the magnitude, respectively. The name and type of the star are shown in the bottom right panel.

Table 1
Relevant Parameters for the Stars Whose Light Curves Are Shown in Figure 2

Numerical ID	R.A.	Decl.	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

Note. We report the numerical ID (Col. (1)), R.A. (Col. (2)) and decl. (Col. (3)), period in days (Col. (4)), V-band magnitude (Col. (5)), number of measurements (Col. (6)), and amplitude (Col. (7)).

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 Equation (3) is capable of producing oscillatory behavior at all, it must be a nonconservative oscillator, with the term $a + bu + c\dot{u}$ changing sign at least twice during a period.

The phase plane for the s-∘ form is shown in Figure 7(a) and for the g-∘ form in Figure 7(b), for choices of the coefficients in the range appropriate for reproducing observational light curves. In the following we show that the s-∘ form can be fully solved analytically and find a condition for the g-∘ form to have a stable limit cycle.

4.1. Analytical Solutions for the s-∘ Form

The coefficient c, the only one to be nonzero in the s-∘ form, sets the strength of the $\dot{u}^2$ perturbing term in Equation (3). 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 $\dot{u}^2 \geqslant 0$ endows the s-∘ form with a symmetry whereby mapping $u \rightarrow -u$ 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 the s-∘ form depends only on the square of the magnitude derivative, $\dot{u}$. Therefore, the phase trajectory is

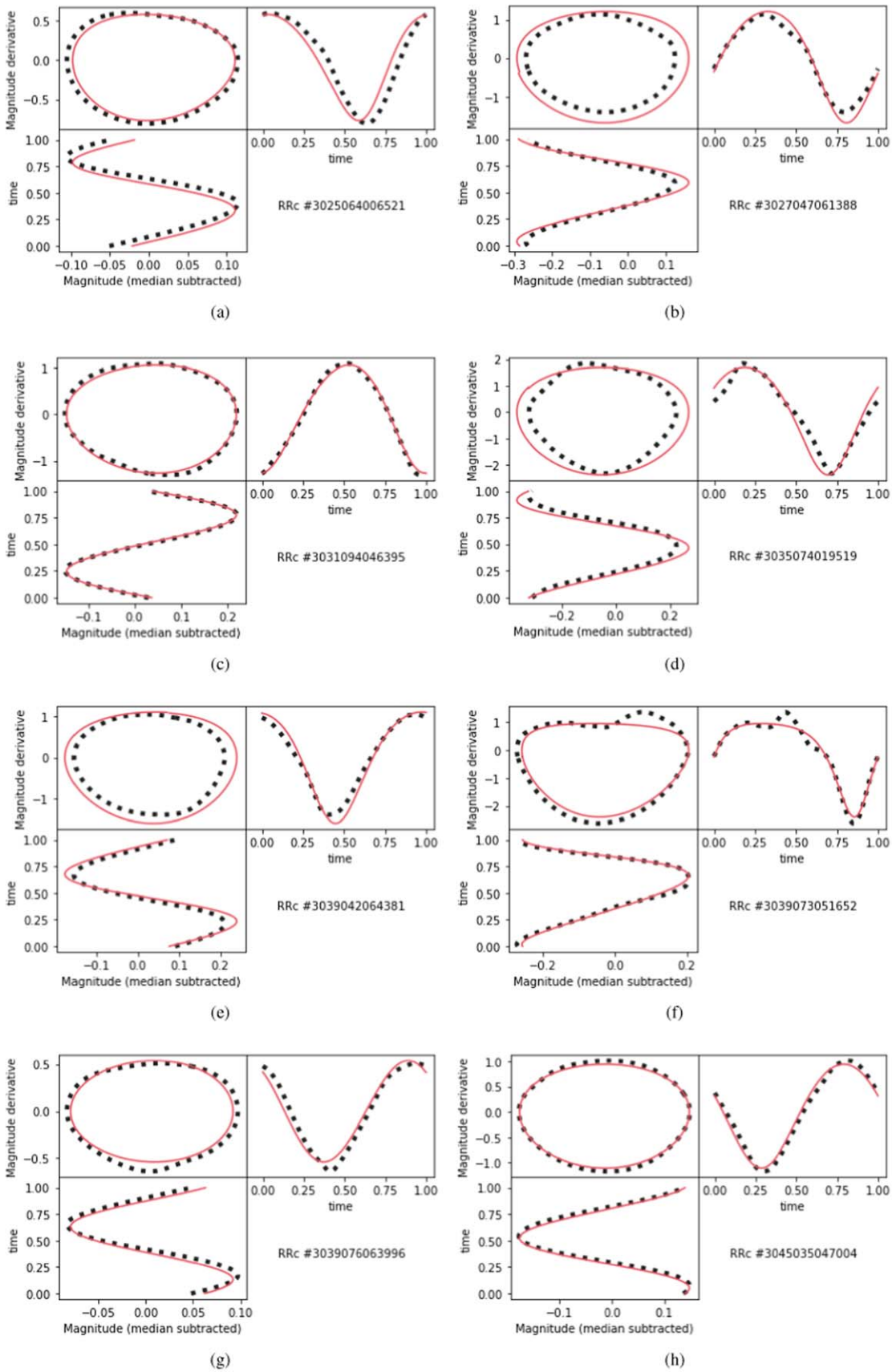

Figure 4. Eight light curves for stars classified as RRc (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and the bottom left panels show the magnitude derivative vs. time and time vs. the magnitude, respectively. The name and type of the star are shown in the bottom right panel.

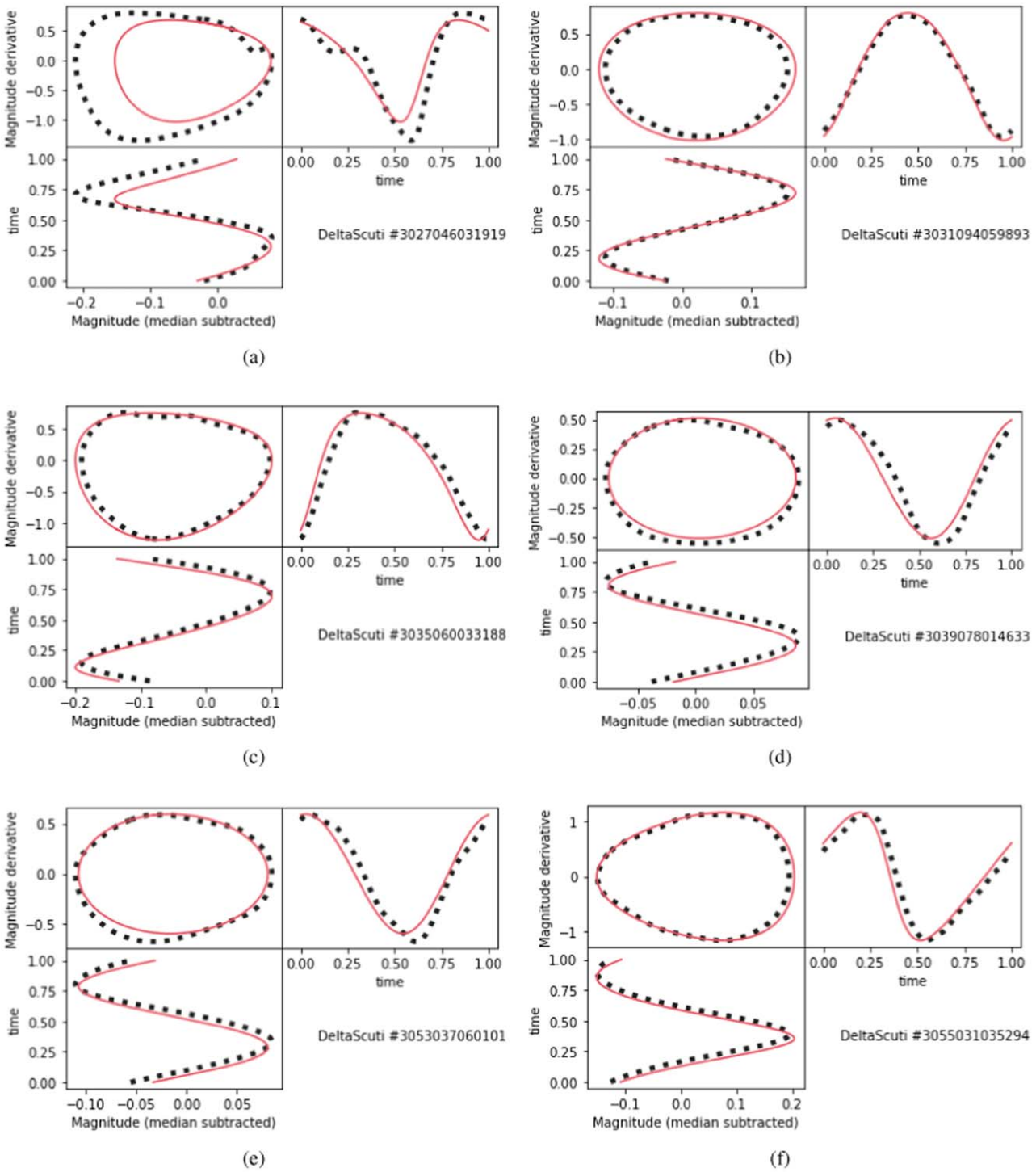

Figure 5. Six light curves for stars classified as δ Scuti (thick dotted black lines) and the respective SINDy solution (thin solid red line). In each plot the top left panel shows the phase plane, where the derivative of the star's magnitude is plotted against its magnitude. A closed curve in this plane represents a periodic evolution, and a perfect ellipse would correspond to a harmonic oscillator. The top right and bottom left panels show the magnitude derivative vs. time and time vs. the magnitude, respectively. The name and type of the star are shown in the bottom right panel.

symmetric with respect to reflection about the u-axis, a property that is not shared by the g-$\circ$ form.

We solve the s-$\circ$ form by introducing the new variable $v = \dot{u}^2$ and observing that $\dot{v} = 2\dot{u}\ddot{u}$; hence, by the chain rule $\ddot{u} = dv/2du$. So the s-$\circ$ form becomes

$$\dot{u} = \pm\sqrt{v} = \pm\sqrt{\frac{1}{2c^2} + ke^{2cu} + \frac{u}{c}}, \tag{6}$$

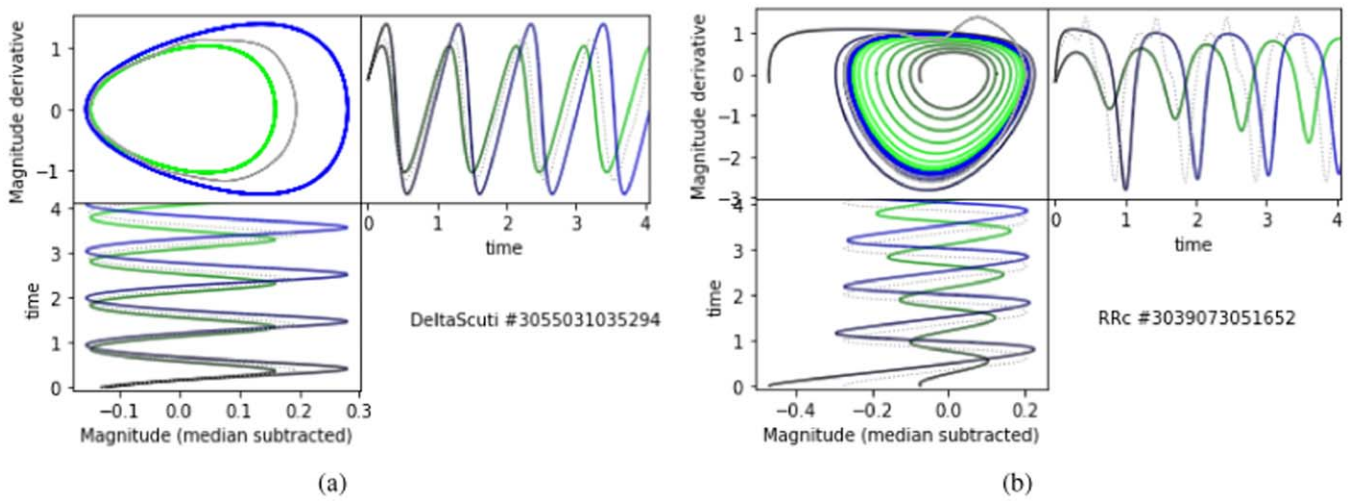

Figure 6. Effects of perturbing the initial conditions of the equation learned by SINDy in two different cases. The left panel shows star 3055031035294, which is a δ Scuti variable, whose learned equation is in the simplified form, with $a = 0$ and $b = 0$; the right panel shows star 3039073051652, which is an RRc variable described by an equation in the full 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 toward the limit cycle corresponding to the unperturbed solution. Unlike in the previous figures, here the top right and bottom left panels show three periods rather than one.

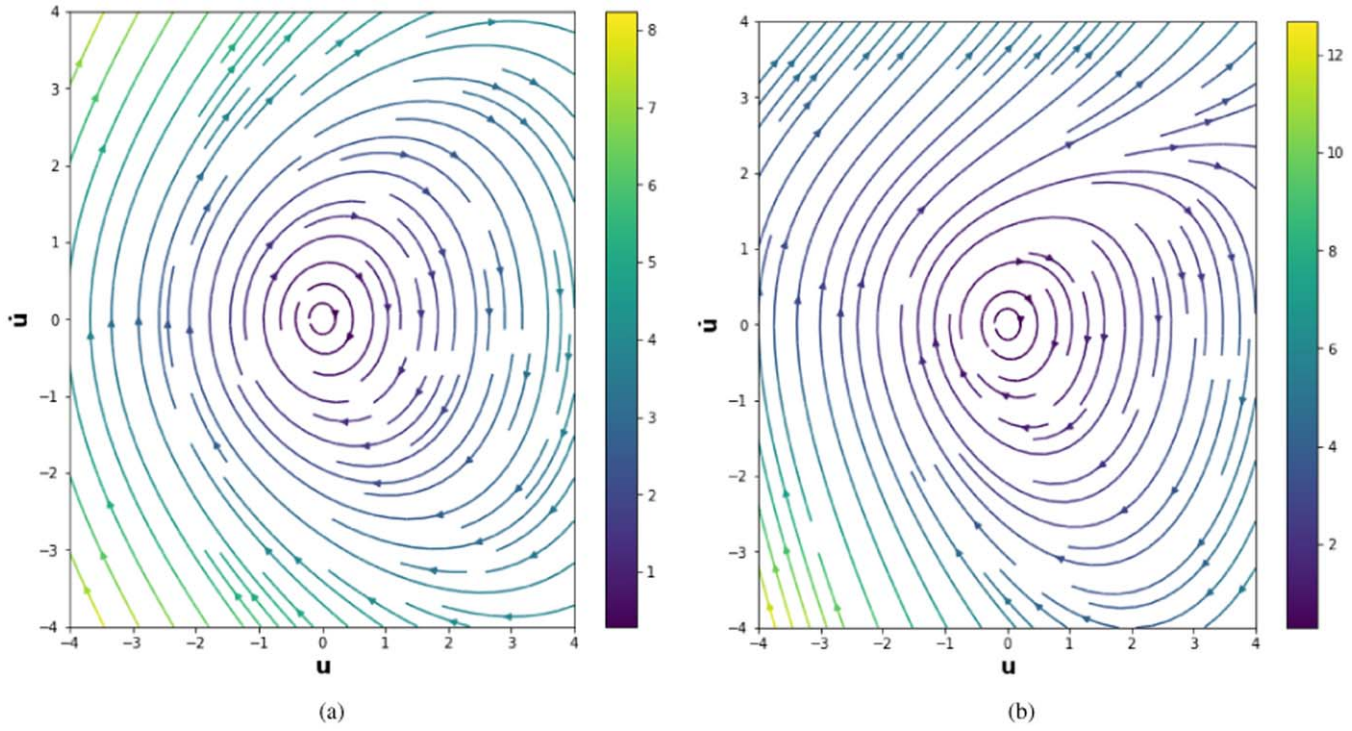

Figure 7. Phase planes for the s-∘ form with $c = 0.2$ (left) and for the g-∘ form with $a = -0.01$, $b = c = 0.25$ (right). Color encodes the modulus of the vector tangent to the phase trajectory in each point.

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. Equation (6) yields two kinds of phase curves: ones exponentially escaping to infinity ("supernova solution") with $k \geqslant 0$, 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 Figure 6, where we alter the initial conditions slightly for a star whose learned governing equation is in the s-∘ form, showing that the resulting solutions are still periodic.

We could not find closed-form solutions for the g-∘ form, as opposed to the s-∘ form. The right panel of Figure 6 shows that, in general, solutions are not periodic. Numerical integration from slightly perturbed initial conditions reveals, however, that

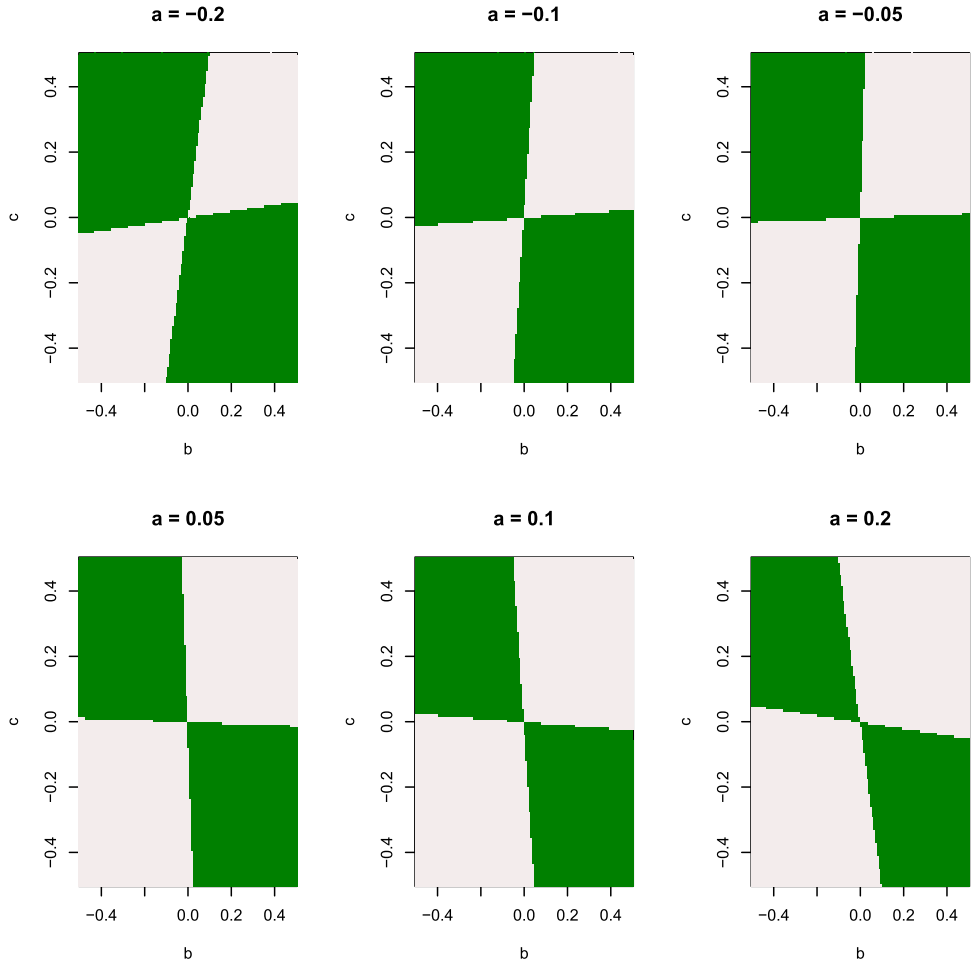

Figure 8. Relation between the a, b, and c coefficients for which $\hat{A} > 0$ (green shaded area), resulting in the existence of a limit cycle for the g-∘ form.

Table 2
Summary of SINDy Results Broken Down by Variable Type (Columns)

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

Note. 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 parentheses.

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 toward 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.

4.2. Limit Cycle Behavior of the g-∘ Form

The g-∘ 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 $\lambda_{1,2}$,

$$J = \begin{bmatrix} 0 & 1 \\ -1 & a \end{bmatrix}, \qquad \lambda_{1,2} = \frac{a}{2} \pm \frac{\mathrm{i}\sqrt{4 - a^2}}{2}. \tag{7}$$

The results from Equation (7) 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 the g-∘ form can be written in the

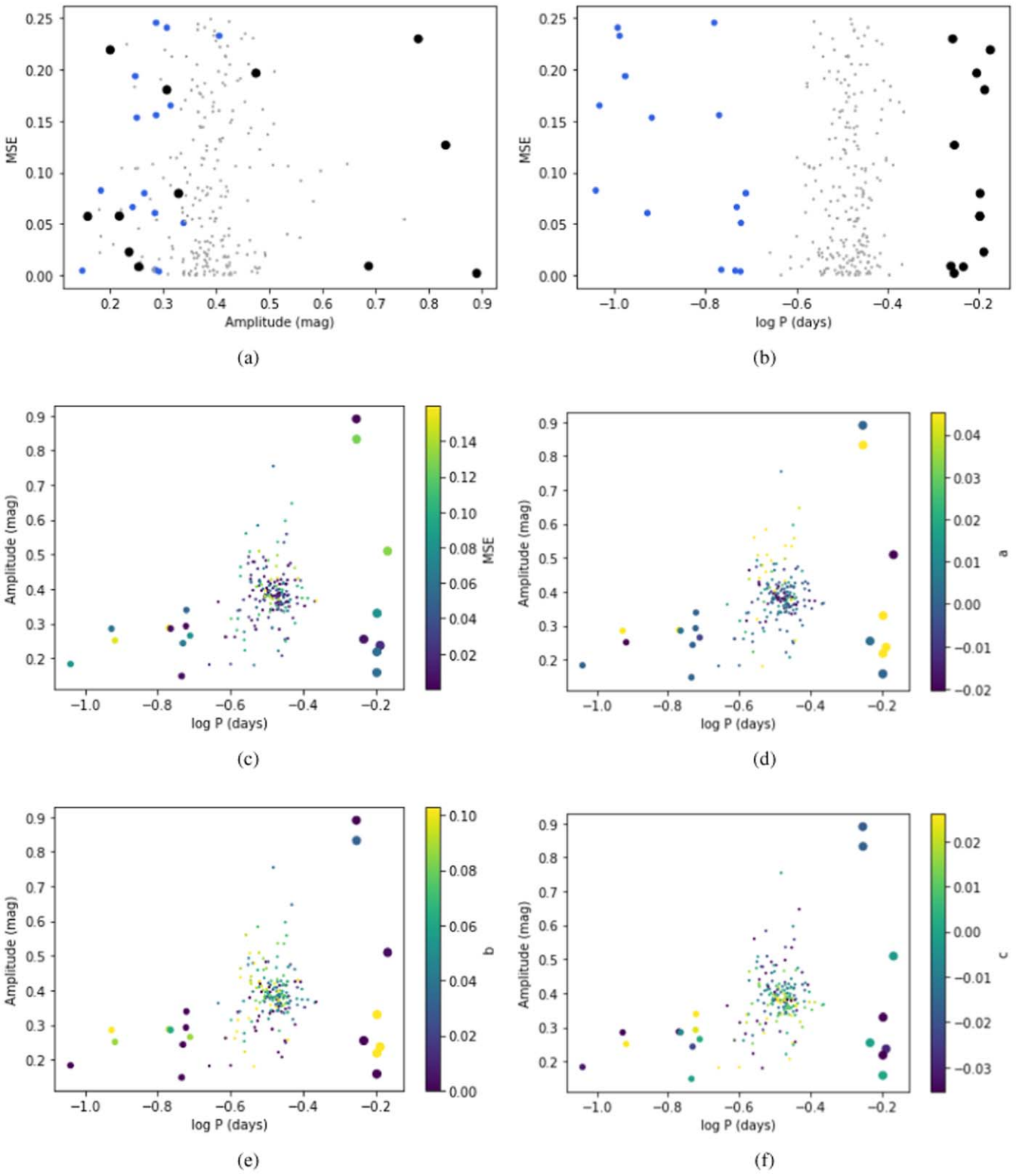

Figure 9. Normalized MSE as a function of the light-curve amplitude (panel (a)) and period (panel (b)). Big black circles represent RRab variables, small gray circles represent RRc variables, and blue circles represent δ Scuti variables. Panels (c) to (f) show, in order, normalized MSE, a, b, and c color-coded 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.

Cartesian Hopf normal form

$$\dot{y} = \alpha y - \omega v + f(y, v),$$
$$\dot{v} = \omega y + \alpha v + g(y, v), \quad (8)$$

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-∘ form equation.

From the linearization eigenvalues, we construct a transformation matrix

$$\Phi = \begin{bmatrix} \frac{a}{2} & \frac{\sqrt{4-a^2}}{2} \\ 1 & 0 \end{bmatrix}, \tag{9}$$

so that the matrix product $\Phi^{-1}J\Phi$ produces the correct coefficients in Equation (8) to first order in (y, v), thereby ensuring the linear relationships necessary between $(u, \dot{u})$ and (y, v) to change system g-∘ form into the normal Hopf form

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

$$\begin{bmatrix} u \\ \dot{u} \end{bmatrix} = J^{-1}\Phi \begin{bmatrix} y \\ v \end{bmatrix} = \begin{bmatrix} \frac{y(a^2-2) + av\sqrt{4-a^2}}{2} \\ \frac{ay + v\sqrt{4-a^2}}{2} \end{bmatrix}. \tag{10}$$

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

$$\begin{aligned} f(y, v) &= \frac{1}{4}a^2[(a^2-2)b + ac]y^2 + 2\sqrt{4-a^2} \\ &\quad \times [(a^2-1)b + ac]yv - (a^2-4)(ab + c)v^2, \\ g(y, v) &= -\frac{a(a^2-2)[(a^2-2)b + ac]y^2}{4\sqrt{4-a^2}} \\ &\quad - \frac{1}{2}(a^2-2)[(a^2-1)b + ac]yv \\ &\quad + \frac{(a^4 - 6a^2 + 8)(ab + c)v^2}{4\sqrt{4-a^2}}. \end{aligned} \tag{11}$$

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[15] $\hat{A}$ and period $\hat{T}$ of these solutions, including their stability, in (y,v) space provided that 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 Equation (8) are 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)

$$\begin{aligned} \hat{A} &\sim \frac{-1}{16\omega}((f_{yyy} + f_{yvv} + g_{yyv} + g_{vvv})\omega + f_{yv}(f_{yy} + f_{vv}) \\ &\quad - g_{yv}(g_{yy} + g_{vv}) - f_{yy}g_{yy} + f_{vv}g_{vv})|_{(y,v)=(0,0)} \\ &= \frac{(ab + 2c)(b + ac)}{4(a^2-4)}, \\ \hat{T} &\sim \frac{2\pi}{\omega} = \frac{4\pi}{\sqrt{4-a^2}}, \end{aligned} \tag{12}$$

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 Figure 8.

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

4.3. 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 Table 2. 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 Figure 9 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 Figure 9 we show the Bailey diagram, with the values of MSE, a, b, and c color-coded. Again, no obvious patterns are apparent.

In Figure 10 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.

5. Conclusions

Variable stars play a key 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 behavior by means of concise sets of equations is as old as science itself, but only recently has it become possible to learn such governing equations directly from the data using a machine-learning approach.

In this paper we have shown that simple governing equations learned from variable starlight 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. 2019, 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 behavior of the two types of variables.

At any rate, when modeling is successful, the learned equations take on a universal form that is essentially a perturbed harmonic oscillator. The perturbing nonlinear terms,

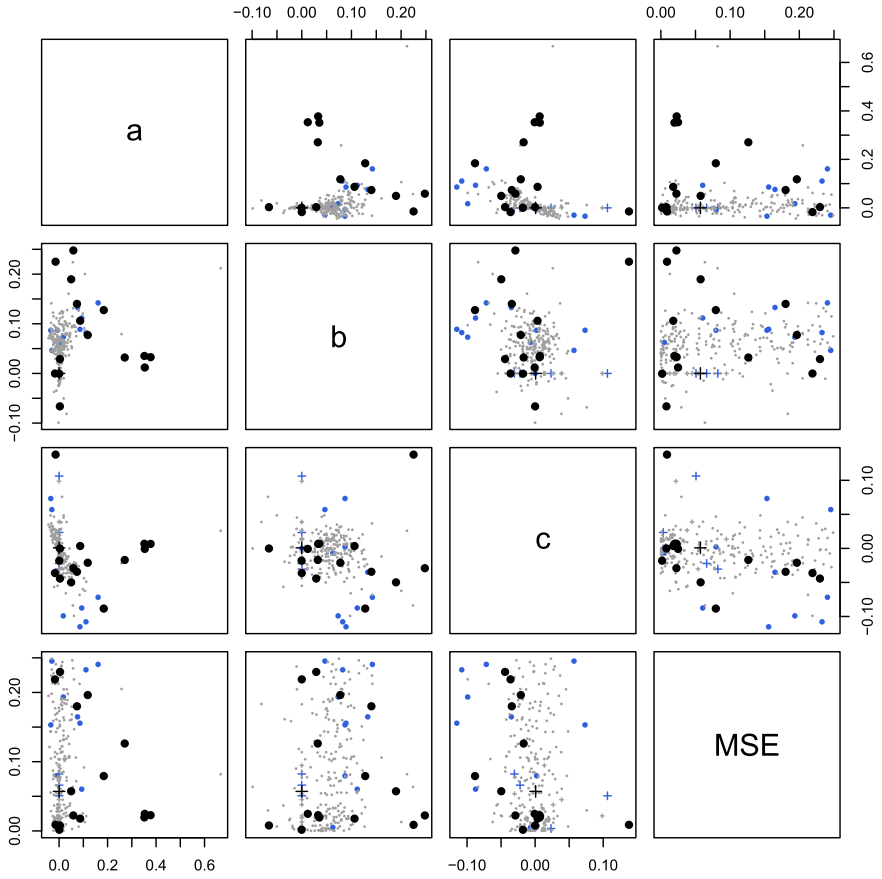

Figure 10. Scatter plots of the *a*, *b*, and *c* coefficients vs. each other and vs. normalized MSE. Big black symbols represent RRab variables, small gray symbols represent RRc variables, and blue symbols represent δ Scuti variables. Filled circles correspond to g-ω form stars and plus signs to s-ω form stars.

however, are nontrivial: 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 machine-learning model (such as a decision tree, e.g., Askar et al. 2019; or a support-vector machine, e.g., Peruzzi et al. 2021), 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 the Blazhko effect.

This material is based on work supported by Tamkeen under the NYU Abu Dhabi Research Institute grant CAP^3. M.P. acknowledges financial support from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 896248. We thank Prof. Eliot Fried and Prof. Hshyar Abdullah for helpful discussion. A.A.T. acknowledges support from JSPS KAKENHI grant Nos. 17H06360, 19K03907, and 21K13914. M.P. wishes to thank Anna Petrivna for the strawberries.

ORCID iDs

Mario Pasquato https://orcid.org/0000-0003-3784-5245
Mohamad Abbas https://orcid.org/0000-0001-9146-0421
Alessandro A. Trani https://orcid.org/0000-0001-5371-3432
Matteo Nori https://orcid.org/0000-0002-5447-0337
Piero Trevisan https://orcid.org/0000-0001-9511-4649
Vittorio F. Braga https://orcid.org/0000-0001-7511-2830
Giuseppe Bono https://orcid.org/0000-0002-4896-8841
Andrea V. Macciò https://orcid.org/0000-0002-8171-6507

References

Abbas, M., Grebel, E. K., Martin, N. F., et al. 2014, AJ, 148, 8
Abdullah, H. K. 2017, in AIP Conf. Ser. 1872, Mathematical Methods and Computational Techniques in Science and Engineering (Melville, CA: AIP), 020006
Addo-Asah, W., Akpati, H. C., & Mickens, R. E. 1995, JSV, 179, 733
Alcock, C., Allsman, R. A., Alves, D. R., et al. 2000, ApJ, 536, 798
Arzani, A., & Dawson, S. T. M. 2020, arXiv:2010.00131
Askar, A., Askar, A., Pasquato, M., & Giersz, M. 2019, MNRAS, 485, 5345
Baker, N. H., Moore, D. W., & Spiegel, E. A. 1966, AJ, 71, 844
Bates, D., Chambers, J., & Hastie, T. 1992, Nonlinear Models (Boca Raton, FL: CRC Press), 421
Bono, G., Caputo, F., Castellani, V., & Marconi, M. 1997, A&AS, 121, 327

Bono, G., & Stellingwerf, R. F. 1994, ApJS, 93, 233
Braga, V. F., Bono, G., Fiorentino, G., et al. 2020, A&A, 644, A95
Brunton, S. L., Proctor, J. L., & Kutz, J. N. 2016, PNAS, 113, 3932
Buchler, J. R. 1993, Ap&SS, 210, 9
de Silva, B., Champion, K., Quade, M., et al. 2020, JOSS, 5, 2104
Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013a, ApJ, 763, 32
Drake, A. J., Catelan, M., Djorgovski, S. G., et al. 2013b, ApJ, 765, 154
Drake, A. J., Djorgovski, S. G., Catelan, M., et al. 2017, MNRAS, 469, 3688
Drake, A. J., Graham, M. J., Djorgovski, S. G., et al. 2014, ApJS, 213, 9
Gillet, D., Mauclaire, B., Lemoult, T., et al. 2019, A&A, 623, A109
Glendinning, P. 1994, Stability, Instability and Chaos: an Introduction to the Theory of Nonlinear Differential Equations (Cambridge: Cambridge Univ. Press)
Guan, Y., Brunton, S. L., & Novosselov, I. 2021, RSOS, 8, 202367
Horrocks, J., & Bauch, C. T. 2020, NatSR, 10, 7061
Icke, V., Frank, A., & Heske, A. 1992, A&A, 258, 341
Iorio, G., Belokurov, V., Erkal, D., et al. 2018, MNRAS, 474, 2142
Ivezić, Ž, Vivas, A. K., Lupton, R. H., & Zinn, R. 2005, AJ, 129, 1096
Kinemuchi, K., Smith, H. A., Woźniak, P. R., McKay, T. A. & ROTSE Collaboration 2006, AJ, 132, 1202
Kunder, A., Tilton, A., Maertens, D., et al. 2019, ApJL, 877, L17
Lomb, N. R. 1976, Ap&SS, 39, 447
Moehlis, J., Josic, K., & Shea-Brown, E. T. 2006, SchpJ, 1, 1358
Moore, D. W., & Spiegel, E. A. 1966, ApJ, 143, 871
Munteanu, A., Bono, G., José, J., García-Berro, E., & Stellingwerf, R. F. 2005, ApJ, 627, 454
Nagy, M., & Petrovay, K. 2013, AN, 334, 964
Neeley, J. R., Marengo, M., Freedman, W. L., et al. 2019, MNRAS, 490, 4254
Nemec, J., & Mateo, M. 1990, in ASP Conf. Ser. 11, Confrontation Between Stellar Pulsation and Evolution, ed. C. Cacciari & G. Clementini (San Francisco, CA: ASP), 64
Peruzzi, T., Pasquato, M., Ciroi, S., et al. 2021, A&A, 652, A19
Pietrukowicz, P., Kozłowski, S., Skowron, J., et al. 2015, ApJ, 811, 113
Prudil, Z., Hanke, M., Lemasle, B., et al. 2021, A&A, 648, A78
R Core Team 2019, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/
Scargle, J. D. 1982, ApJ, 263, 835
Schwarzenberg-Czerny, A. 1989, MNRAS, 241, 153
Sesar, B., Ivezić, Ž, Grammer, S. H., et al. 2010, ApJ, 708, 717
Simion, I. T., Belokurov, V., Irwin, M., & Koposov, S. E. 2014, MNRAS, 440, 161
Stellingwerf, R. F. 1972, A&A, 21, 91
Stellingwerf, R. F. 1986, ApJ, 303, 119
Tanaka, Y., Seya, K., & Takeuti, M. 1990, in ASP Conf. Ser. 11, Confrontation Between Stellar Pulsation and Evolution, ed. C. Cacciari & G. Clementini (San Francisco, CA: ASP), 145
Tibshirani, R. 1996, J. R. Stat. Soc. Series B (Methodol.), 58, 267
Torrealba, G., Catelan, M., Drake, A. J., et al. 2015, MNRAS, 446, 2251
Van der Pol, B. 1934, Proc. IRE, 22, 1051
Welch, D. L., & Stetson, P. B. 1993, AJ, 105, 1813