PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Tang T, Taylor PH, Chen Y, Adcock TAA. 2025 Solitary wave interactions modelled using conserved and empirically conserved quantities discovered with machine learning. *Proc. R. Soc. A* **481**: 20240904. https://doi.org/10.1098/rspa.2024.0904

Received: 25 November 2024
Accepted: 14 March 2025

Subject Areas:
wave motion, fluid mechanics

Keywords:
solitary wave, symbolic regression, conserved quantity, partial differential equations

Author for correspondence:
Tianning Tang
e-mail: tianning.tang@eng.ox.ac.uk

THE ROYAL SOCIETY
PUBLISHING

Solitary wave interactions modelled using conserved and empirically conserved quantities discovered with machine learning

Tianning Tang[1,2], Paul H. Taylor[3], Yuntian Chen[4] and Thomas A. A. Adcock[1]

[1]Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK
[2]Department of Mechanical and Aerospace Engineering, The University of Manchester, Manchester, M13 9PL, UK
[3]School of Earth and Oceans, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia
[4]Eastern Institute for Advanced Study, Zhejiang 315000, People's Republic of China

TT, 0000-0002-6365-9342; PHT, 0000-0002-3680-5543; YC, 0000-0003-4566-8197; TAAA, 0000-0001-7556-1193

Conservation laws play an important role in analysing physical systems and solving partial differential equations (PDEs). In this paper, we use associated conserved quantities (CQs) to construct a simple and robust way to approximate the interaction between surface solitary waves—a common nonlinear phenomenon in many research fields and engineering applications. One major constraint of such an approach is how many CQs are available from the governing equation. In this study, we start with the Korteweg–De Vries (KdV) equation, which has an infinite number of CQs. We show that these CQs can be used to predict soliton interaction accurately. The regularized long wave (RLW) equation—an equivalent reduction form to the KdV equation, however, has insufficient CQs to apply the approach. To address this problem, we use a new symbolic-based machine learning (ML) approach to search for any higher-order empirically CQs (e-CQs), which allow the proposed approach to be applied directly

to the RLW equations and leverages the constraints on the number of CQs available. This opens a new opportunity for more complex systems governed by a variety range of PDEs to be modelled with such a CQ-based approach.

1. Introduction

The interaction between solitary waves in a nonlinear dispersive medium is a common phenomenon that appears in a wide range of research fields, such as fluid mechanics [1], water waves [2,3], nonlinear plasma [4–7] and fibre optics [8,9], and can be governed by many nonlinear equations [10]. The Korteweg–de Vries (KdV) equation [11] and its regularized version—the regularized long wave (RLW) equation also known as the Peregrine or Benjamin–Bona–Mahony equation [12] are the classical equations describing such nonlinear interactions between unidirectional finite-amplitude waves in a dispersive system. Of particular interest are solitary wave solutions to these.

Various direct methods have been developed to obtain the analytical solutions for these two equations, including the inverse scattering technique [13] and the Darboux transform method [14] for the KdV equation, the extended tanh method [15], the homogeneous balance method [16], the modified auxiliary equation method [17] and others [18] for the RLW equation. In addition, many numerical approaches have also been used to solve each equation for solitary wave interactions [19–22]. These solitary wave interaction problems also attract significant attention from the practical point of view, such as plasma collisions [4], surface wave collisions [23] and for the behaviour of shallow water waves [24,25]. As such, obtaining fast, robust and accurate solutions to the solitary wave interaction problem is important for a better understanding of the nonlinear mechanics, and also hints at faster engineering modelling tools to capture such nonlinear phenomenon.

In this study, we take advantage of invariant integrals or conserved quantities (CQs) to construct a framework for modelling the interaction of two-solitary waves. These CQs are important because they provide known constants preserved throughout the nonlinear interactions. With the number of CQs being larger than the number of parameters used to describe the states of the system, the nonlinear system can be effectively determined at any time instance. This provides a computationally efficient method to examine the characteristics of solitary wave interactions, providing a good approximation throughout the evolution without the necessity of using more complicated analytical results or numerical time-marching techniques. Similar approaches have been used to investigate the nonlinear changes during the evolution of a Gaussian wave group for waves on deep water [26] and for solitary interactions governed by KdV and RLW equations [27–29].

A major step forward for a wider application of such CQ models is to relax the limits on how many CQs are available for the governing equation. Traditionally, CQs for a given equation can be mathematically derived. For example, the classical KdV equation has an infinite number of CQs [30]. This was first proved as part of the work leading to the discovery of the inverse scattering transform [13], which required significant analytical and computational input. Other equations, however, have been proved to have only a finite number of CQs, such as the RLW equation with only three CQs [31,32]. Recently, computational algorithms [33] and machine learning (ML) [34–36] approaches have been developed as alternative approaches to finding conservation laws. These methods, however, are mostly focused on re-discovering the physically interpretable lower-order CQs, whereas for the proposed CQ models, the higher-order CQs are the critical ones for increasing the total CQ numbers.

In this study, we take a more radical approach and develop a new symbolic ML CQ discovery approach, where the new method allows the discovery of complex higher-order CQ forms that are conserved numerically to sufficient accuracy to be used in analysis. This provides additional apparent CQs to be incorporated into modelling and extends the capability of the CQ models for

modelling more complex nonlinear systems. In addition, for those equations with an insufficient number of exact CQs for modelling, this novel scheme brings a new concept of integral quantities that are almost conserved. These discovered integrals stay close enough to a constant value throughout all stages of the simulation and can be used to provide additional information for modelling—we name these empirically CQs (e-CQs). A series of e-CQs may allow an extension of the existing CQ-based modelling for complex equations where higher-order CQs are required. However, we note the possibility that such e-CQs might not exist or, even if they do, they might not be revealed in our search.

In this study, we aim to develop a non-time-step modelling of the classic two-solitary wave interaction problem, which can be achieved through two main steps:

(i) Finding sufficient CQs for the analysis. Where these are not available in the literature, we develop a symbolic regression-based CQ and e-CQ discovery scheme that finds sufficient CQs and e-CQs allowing for characterizing the complete interaction between two-solitary waves (§2a).
(ii) A reduced-order parameterization that fully uses the CQs and e-CQs of the governing equation to approximate the two-solitary wave interaction without time-stepping (§2b).

We first examine the reduced-order parameterization approach for the KdV equation in §4, where it is well known that there are an infinite number of CQs. We then move on to analyse the RLW (or Peregrine equation), where a single additional higher e-CQ is required to fully characterize the system, arising from the current parameterization of the solitary wave interactions. We identify two additional e-CQs through a new symbolic regression ML scheme and validate these against high-accuracy numerical simulations (§5).

2. Methodology

(a) An ML-based discovery of CQs and e-CQs

We develop a new symbolic-based ML approach to discover CQs and e-CQs. Following the definition of the CQ, we have

$$\int T\,\mathrm{d}x = \int [F(u) - I(u)]\,\mathrm{d}x = \text{constant for } t \in R, \tag{2.1}$$

where T is the CQ, and its spatial integral does not change during the entire evolution of the system, $F(u)$ is the leading-order term, which is usually written in the form of $F(u) = (1/k)u^k$ following the Miura forms [30] for the kth-order CQ for both KdV and the RLW equations, and $I(u)$ are the following terms that satisfy the condition. In this paper, we take the one-dimensional form for CQs in equation (2.1). More generally, this could be done over an arbitrary number of dimensions. For instance, in two dimensions:

$$\int\int T(u(x,y))\,\mathrm{d}x\,\mathrm{d}y = \int\int [F(u(x,y)) - I(u(x,y))]\,\mathrm{d}x\,\mathrm{d}y = \text{constant for } t \in R, \tag{2.2}$$

where y is the extra dimension of the system in addition to x, This will naturally allow the methodology below to be extended to nonlinear systems with more than one dimension by including an extra spatial integral. This would allow our method to be potentially applied to high-dimensional systems with different governing equations, for example, the Zakharov–Kuznetsov equation for ion-acoustic waves in a plasma with three known CQs in two dimensions [6] and also for a modified KdV–Zakharov–Kuznetsov equation in a magnetized electron–positron [7] in three dimensions.

In this study, we aim to find the analytical expression of $I(u)$ corresponding to each order of $F(u)$ with minimal restrictions on their free form. To achieve this, we start with the classic symbolic regression approach, where the mathematical expression of $I(u)$ can be expressed by a tree consisting of constants, operators and elementary fields. In principle, this tree-based expression

allows the symbolic regression package (PySR, [37]) to uncover expressions with unbounded complexity. We allow $Op = [\text{'plus'}, \text{'minus'}, \text{'multiply'}]$ as the only three allowable operators to be consistent with the expression of CQs. The elementary fields are the solitary wave surface profile $u(x,t)$ and its spatial derivatives, here up to third order written as u_0, u_1, u_2, u_3, as in the original work on CQs by Miura *et al.* [30]. A genetic algorithm is implemented to explore a large number of candidate expressions and select the best ones, before mutating and adapting them into the next generation. The candidate expressions can be any elementary fields connected with allowable operators, e.g. $-u_0 \times u_1 + u_1 \times u_0 \times u_1 \times u_2$ or $u_1 \times u_1 - u_3 \times u_1$. Over 10 000 possible candidates are generated and searched for each CQ. Finally, the symbolic regression package will assemble a list of candidate expressions of $I(u)$ and present the Pareto-optimal solutions at different complexity. We select the best solution manually and pick the expression with the best parsimony score without stiff coefficients.

One key difference, however, between the existing symbolic regression approaches for partial differential equation (PDE) identification and the current CQ discovery task is the optimization target, where the current symbolic regression minimizes the difference between target $F(u)$ and the identified $I(u)$ at every point in the entire spatial-temporal domain:

$$\text{Find } I(u(x_i,t_j)) \in \mathcal{T}_{Op} \quad \text{that minimizes } \frac{1}{N_i N_j}\sum_i\sum_j [F(u(x_i,t_j)) - I(u(x_i,t_j))]^2, \tag{2.3}$$

where $\mathcal{T}_O$ is the functional space that is consistent with the pre-defined operator set O, and N_i, N_j denotes the number of spatial and temporal grid points, respectively.

The CQ discovery method, however, requires the minimization of the temporal change of the differences between the $F(u)$ and the $I(u)$ within a spatial integral form:

$$\left.\begin{aligned}&\text{Find } I(u(x_i,t_j)) \in \mathcal{T}_{Op} \quad \text{that minimizes}\\ &\frac{1}{N_j}\sum_j \int [F(u(x,t_j)) - I(u(x,t_j))]\mathrm{d}x - \int [F(u(x,t_o)) - I(u(x,t_o))]\mathrm{d}x,\end{aligned}\right\} \tag{2.4}$$

where t_o is a random but different time step to t_j during the evolution. Here, we fully use one of the key characteristics of CQs—the spatial integral remains constant regardless of the temporal sequence. This randomization in time sequence helps better discovery of CQs.

As such, we use a new objective function $\mathcal{L}$ for the CQ discovery with an extra spatial integration step:

$$\mathcal{L} = \sum_{j=l}^{N_j} \frac{\left\| (\partial/\partial t_r)\int_0^l F(u^{(t=t_j)})\,\mathrm{d}x - (\partial/\partial t_r)\int_0^l I(u^{(t=t_j)})\,\mathrm{d}x \right\|}{\int_0^l |F(u^{(t=t_j)})|\mathrm{d}x\,\mathrm{d}t}, \tag{2.5}$$

where l is the length of the assumed periodic numerical domain and t_r is the time without temporal dependence. We achieve this by randomizing the temporal sequence of the input profiles. As such, the new PySR-CQ package can discover free-form mathematical expressions of $I(u)$ based on the given expression of the leading-order term $F(u)$ for any PDE with this new objective function. The accuracy of this new package is further examined in §4, and the choice of $F(u)$ is discussed in §7.

(b) Parametrization of two-soliton interactions with CQs

(i) Single-solitary wave solutions of the KdV and RLW equations

We start with the non-dimensional form for the KdV [11] and RLW [12]

$$\left.\begin{aligned}&\text{KdV: } u_t + (1+u)u_x + u_{xxx} = 0\\ \text{and}\quad &\text{RLW: } u_t + (1+u)u_x - u_{xxt} = 0,\end{aligned}\right\} \tag{2.6}$$

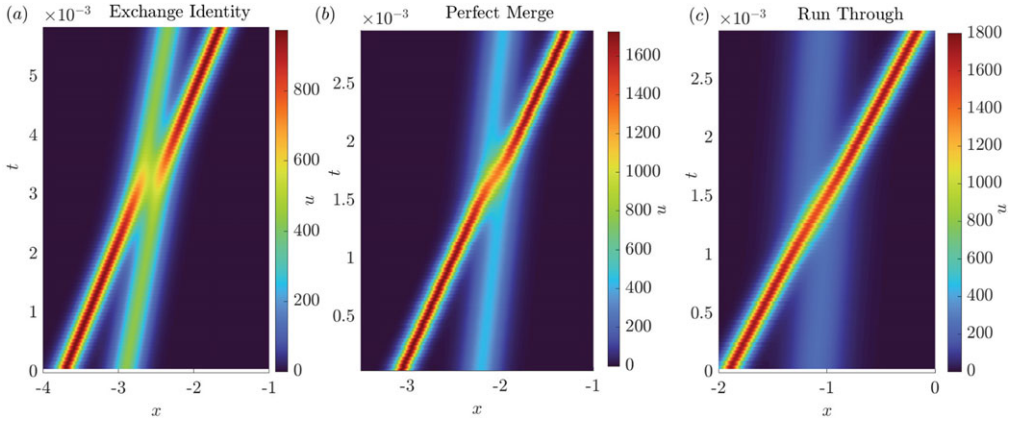

Figure 1. Two-soliton interactions for KdV equation. Panel (*a*) for exchange identity interaction with amplitude $b_1 = 6$ and $b_2 = 9$, panel (*b*) for perfect merge case with amplitudes $b_1 = 6$ and $b_2 = 12$, and panel (*c*) for run-through case with amplitudes $b_1 = 4.5$ and $b_2 = 12$.

where the subscript u_t denotes $\partial u/\partial t$ and the subscript u_x denotes $\partial u/\partial x$. We look for the classic single-solitary solutions in the form of $u = a\,\mathrm{Sech}^2[b(x - ct)]$ for both KdV and RLW equations [38]. The solitary wave solution to the KdV equation is given by

$$u = 12b^2\,\mathrm{sech}^2[b(x - (1 + 4b^2)t)], \tag{2.7}$$

and the solitary wave solution to the RLW equation by

$$u = 12b^2\,\mathrm{sech}^2\left[\frac{b(x - (1 + 4b^2)t)}{\sqrt{1 + 4b^2}}\right], \tag{2.8}$$

where the only parameter b controls both the width and the height of the solitary waves for both the KdV and RLW equations.

(ii) The interaction of two-solitary waves in the KdV and RLW equations

For the case of two-solitary waves, due to the solitary wave with a larger amplitude travelling faster than the smaller solitary wave, interactions take place when these two-solitary waves overlap. Depending on the different combinations of initial amplitudes of these two-solitary waves, three different patterns can be observed at the time when maximally overlapped, which were briefly described in [39].

We present these three patterns in figure 1. Figure 1*a* shows the case of the exchange of identity, where the two-solitary waves never fully overlap. Instead, when the taller wave catches up with the smaller wave, its height reduces as mass is transferred to the smaller wave until both waves are the same in amplitude. This mass transfer continues until the smaller wave becomes the taller one and moves away ahead. Figure 1*b* shows the cases of perfect merge, where at the maximum overlap, the merged shape is also a perfect sech^2 shape. For the KdV equation, this happens when the relative amplitude ratio of the two initial solitary waves is exactly 4:1. The perfectly merged shape then has a relative height of 3. If the amplitude ratio of the well-separated solitary waves is greater than 4:1, the taller wave will apparently run straight over the smaller wave (figure 1*c*), creating a shape at the maximum overlap with a tall and narrow peak and a wider base, like a traffic cone. One common observation for both perfect-merge and run-over cases is that the amplitude of the merged shape will be intermediate between the taller and the smaller solitary waves. For the exchange of identity case, the maximally overlapped waves have the same amplitude (and width), this height again being between the heights of the two-solitary waves when far away.

The two-solitary wave interactions governed by the RLW equation also exhibit a similar three patterns but with two important differences. First, the interaction pattern is not only determined by the relative ratio of the two-solitary waves but also depends on the absolute value. In addition, these pairwise interactions are not exactly reversible. This will lead to tiny trailing wiggles being shed from the interaction, although these wiggles are several orders of magnitude smaller (within 3/10 of one per cent of the original amplitude when observed in [40]) than the solitary wave for such a typical interaction. This is later confirmed in [41], where the small wiggles could not be accounted for by numerical inaccuracies. A numerical example can be found in https://en.wikipedia.org/wiki/Benjamin%E2%80%93Bona%E2%80%93Mahony_equation#cite_note-Bona_Pritchard_Scott-1. Thus, KdV solitary waves are solitons as defined by Zabusky and Kruskal [4], whereas RLW solitary waves are not solitons.

(iii) Parametrization of the interaction of two-solitary waves

An ansatz can be made to describe all three interaction patterns shown in figure 1 as shapes in space without direct dependence on time. In this study, we use two stretched sech^2 forms to approximate the full interaction of two-solitary waves:

$$u = 12{b_1}^2\,\mathrm{sech}^2\left[w_1\left(x-\frac{L}{2}\right)\right] + 12{b_2}^2\,\mathrm{sech}^2\left[w_2\left(x+\frac{L}{2}\right)\right], \tag{2.9}$$

where we have five parameters: b_1 and b_2 describe the amplitude of two fitted sech^2 shapes, w_1 and w_2 control the width of the two fitted sech^2 shapes and L is the separation distance between the two peaks of the sech^2 shapes. It is straightforward to show that $w = b$ for the KdV equation and $w = b/\sqrt{(1+b^2)}$ for the RLW equation when the two solitary waves are sufficiently far apart. However, as the separation distance L reduces, the heights and widths of each hump become functions of L (and implicitly of time). Thus, the entire structure of the interaction is approximated through the hump parameters (b_1, w_1) and (b_2, w_2) as functions of the separation distance L.

(iv) The approximation of the interaction using CQs

Finally, we use the CQs to solve for the parameters involved in the parametrization (i.e. b_1, b_2, w_1, w_2) as functions of the separation distance L. Four unknown parameters require four equations, so the first four CQs are used to construct the equation system. We solve this equation system via a multi-objective optimization, where L is pre-defined to specify the status of the interaction. In this study, we obtain the corresponding value of L for each time instance in the simulation by fitting equation (2.9) to the simulated solitary wave profiles. We use analytically derived formulae to obtain spatial derivatives such as u_1 and second-order numerical integration to evaluate the CQs.

3. Numerical experiments

In this study, we use accurate numerical simulations governed by both KdV and RLW equations as the training dataset for discovering CQs with symbolic regression.

The KdV equation is solved via the Chebfun package [20], with a spatial domain between zero and 20, a total of 800 spatial nodes and a time step of 0.0025. The physical meaning of these parameters is arbitrary in this paper as they are unitless, but a rigorous grid convergence test was performed to ensure these parameters produce reliable results. The maximum percentage variation in the first lowest-order CQs is less than 0.5% and in most cases, the variations are considerably smaller.

For the RLW equation, we numerically solve using a pseudo-spectral approach with Mathematica, where a spatial domain of 0–1000 is used with a minimum of 1000 spatial nodes and an adaptive time step. Similarly, we also performed rigorous grid convergence tests to ensure numerical accuracy. The maximum percentage variation in all three known CQs is less than 0.4%.

We have used two distinct types of initial conditions for our simulation, both of which are different from the two-solitary wave interaction problem. This further helps ensure that the discovered CQs are generally applicable properties of the governing equation and are not dependent on specific initial conditions.

The first type of initial condition provides a situation where a hump will interact and travel past a hole, following the formula

$$u = A \cdot 2c(B - x) \cdot \exp(-c^2(B - x)^2), \tag{3.1}$$

where A controls the envelope of the shape with a range of $5 \leq A \leq 10$, B controls the horizontal position of the curve, which is fixed at 100 as the centre of the numerical domain and c controls the width of the curve and is limited within the range of $0.1 \leq c \leq 1$ to avoid numerically stiff solutions. This initial condition has a deep hole in front of a tall crest, which is very different in form to that occurring during the interaction of two-solitary waves. This difference is deliberate as we seek widely applicable e-CQs if they can be found.

The second initial condition is given by the three-solitary wave interaction problem, as

$$\begin{aligned} u = {} & 12{a_{\text{init}}}^2 \operatorname{sech}^2[a_{\text{init}}(x - x_a - (1 + 4{a_{\text{init}}}^2)t)] + 12{b_{\text{init}}}^2 \operatorname{sech}^2[b_{\text{init}}(x - x_b - (1 + 4{b_{\text{init}}}^2)t)] \\ & + 12{c_{\text{init}}}^2 \operatorname{sech}^2[c_{\text{init}}(x - x_c - (1 + 4{c_{\text{init}}}^2)t)], \end{aligned} \tag{3.2}$$

where $1.7 < a_{\text{init}} < 3.3$, $0.9 < b_{\text{init}} < 1.6$ and $0.5 < c_{\text{init}} < 0.9$ controls the amplitude of the three-solitary waves and x_a, x_b and x_c controls spatial positions of the three-solitary waves, which are adjusted accordingly to ensure full three-solitary wave interactions take place within the simulation time.

4. Results

(a) KdV equation

(i) Re-discovering the CQs for KdV equation

We first examine the performance of the ML-based CQ discovery scheme by re-discovering the first five CQs for KdV, and the first 10 are listed in the remarkable paper by Miura *et al.* [30]. In table 1, we compare our ML-discovered CQs with the mathematically derived CQs, where we report an exact re-discovery of the third-order CQ. From table 1, we see a very close match to the exact ones for lower-order CQs (i.e. third- and fourth-order CQs), leading to almost perfect re-discovery of CQs. For higher-order CQs (i.e. fifth-order CQs), we observe some differences in the re-discovered coefficients when compared with the exact ones, but our ML discovery scheme still managed to re-discover all the correct terms. This deviation in the coefficients is not an intrinsic limitation of the method but is probably due to the noise introduced during numerical differentiation, especially for those higher-order derivatives, which are also amplified when being multiplied together. This can be evidenced by the fact that the re-discovered CQs are actually marginally better conserved than the exact CQs for the training and testing datasets due to the inaccuracy during the numerical differentiation.

We further explore the variation of the CQs and their individual terms at different time instances in figure 2. For both the fourth-order CQs (*a*) and the fifth-order CQs (*b*), the variation of each term is quite significant. For example, the leading term in both cases varies over 25% during the two-solitary wave interactions. The final CQ, however, remains constant throughout with the maximum relative variation less than 0.5%.

(ii) Two-soliton interactions in the KdV equation

We first look into the classic example of the two-soliton interaction governed by the KdV equation. In figure 3, we show the three types of interaction: exchange of identity, perfect merging and run-through. Figure 3*a–c* shows the exchange of identity interaction, where the distance between

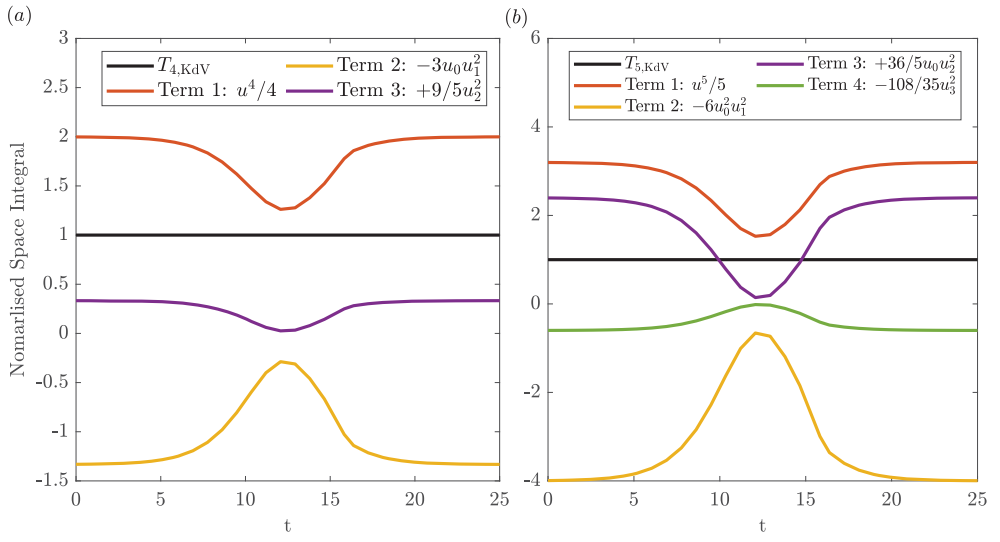

Figure 2. Variation of the CQs and individual terms for (*a*) fourth-order exact CQ and (*b*) fifth-order exact CQ for a KdV solitary wave pair with perfect merging ($b_1 = 6, b_2 = 12$).

Table 1. The comparison of symbolic regression discovered CQs to the mathematically derived to the exact forms in [30], where the subscripts denote the order of spatial derivatives. Other than the first pre-defined entry in each, we convert the fractions in the following terms of the exact CQs into decimals for easier comparison with the discovered CQs.

order	exact CQ	re-discovered CQ
first	$T_1 = u_0$	no $I(u)$ terms
second	$T_2 = \frac{1}{2}{u_0}^2$	no $I(u)$ terms
third	$T_3 = \frac{1}{3}{u_0}^3 - {u_1}^2$	$T_{3,\mathrm{SR}} = \frac{1}{3}{u_0}^3 - 1.00{u_1}^2$
fourth	$T_4 = \frac{1}{4}{u_0}^4 - 3u_0{u_1}^2 + 1.8{u_2}^2$	$T_{4,\mathrm{SR}} = \frac{1}{4}{u_0}^4 - 2.98u_0{u_1}^2 + 1.74{u_2}^2$
fifth	$T_5 = \frac{1}{5}{u_0}^5 - 6{u_1}^2{u_0}^2 + 7.2u_0{u_2}^2 - 3.08{u_3}^2$	$T_{5,\mathrm{SR}} = \frac{1}{5}{u_0}^5 - 5.71{u_1}^2{u_0}^2 + 6.08u_0{u_2}^2 - 2.14{u_3}^2$

two humps will decrease to a non-zero minimum value and two solitons will exchange their properties. Figure 3*d–f* shows the second perfectly merged type, where the distance between the two humps will decrease to zero and the profile of the merged shape is also exactly a sech^2 form (though with lower height to width ratio to be a solitary wave). Figure 3*g–i* shows the last type of interaction, where the larger soliton runs straight through the smaller soliton, where the height of the taller hump drops and the height of the small one increases with an associated width change for both humps. This exchange of mass or energy is returned as the humps separate. From figure 3, the proposed CQ model provides accurate profiles for all three types of interaction, without the necessity of solving the equation in a time-stepped sequence. A slightly larger difference is reported for the exchange identity type at the time instance when the distance between two peaks becomes minimal, this is primarily due to the small mismatch in the pre-specified minimum distance value L, which can be further mitigated by considering L as a tunable parameter during the multi-objective optimization.

Note that, because we aim to model the general interaction of two-solitary waves, our ansatz has five coefficients (equation (2.9)) with one of these, L, as a known control parameter. Hence, we need four CQs in general. Using the symmetry of all forms at the maximal overlap, a smaller number of CQs is required only if the time instances with two-solitary waves an infinite distance apart and at the maximal overlap are needed [27,29]. For the exchange of identity interaction, the

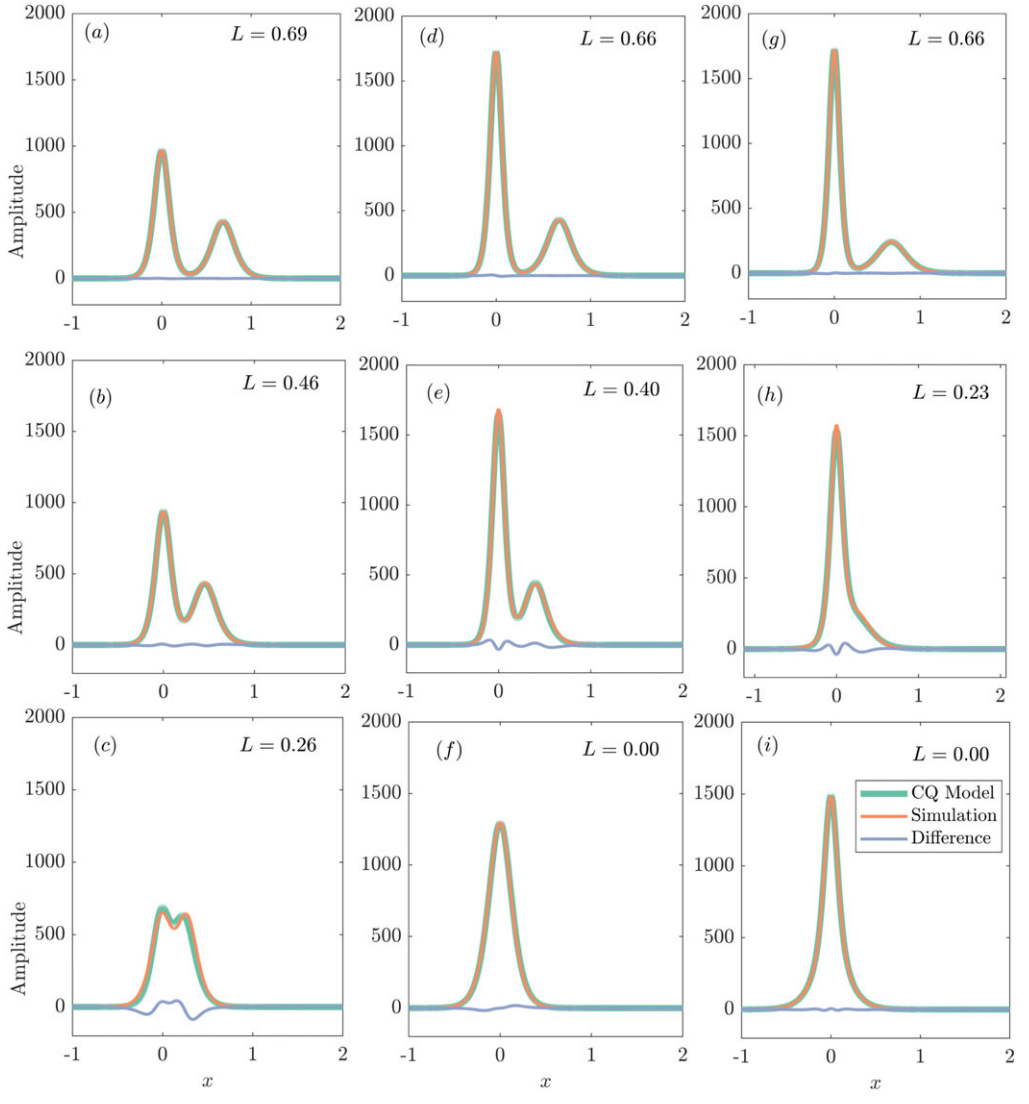

Figure 3. Two-soliton interactions governed by the KdV equation for exchange of identity in (*a–c*) ($b_1 = 6$, $b_2 = 9$), perfect merging in (*d–f*) ($b_1 = 6$, $b_2 = 12$), and run-through in (*g–j*) ($b_1 = 4.5$, $b_2 = 12$). The separation distance L is the distance between the two superposed humps in the approximated form.

three parameters are (b, w, L) since the two waves are identical, at the minimal distances. For the perfect merging case, parameters are b,w only with $L = 0$, since our two sech^2 humps are identical at $L = 0$. For the run-through case, b_1, w_1 and b_2, w_2 are needed with $L = 0$. Hence, the number of CQs required is determined by whether we start with separated solitons and merge them, or we start completely merged and ask what solitons separate out at infinity.

(b) RLW equation

(i) Re-discovering existing two CQs

We start with re-discovering the existing two known CQs for the RLW equation with the ML-based discovery scheme. Table 2 lists a comparison of the exact CQs and the re-discovered CQs for the RLW equation, where we report almost a perfect match between them. Our ML-based

Table 2. Comparison of symbolic regression discovered CQs to the mathematically derived in [31] for RLW, where the subscripts denote the order of spatial derivatives. We convert the fractions in the following terms of the exact CQs into two decimal places for easier comparison with the discovered CQs.

order	exact CQ	re-discovered CQ
first	$T_{1,\mathrm{RLW}} = u_0$	no $I(u)$ terms
second	$T_{2,\mathrm{RLW}} = \left(\frac{1}{2}{u_0}^2 + 0.5{u_1}^2\right)$	$T_{2,\mathrm{SR}} = \left(\frac{1}{2}{u_0}^2 + 0.50{u_1}^2\right)$
third	$T_{3,\mathrm{RLW}} = \left(\frac{1}{3}{u_0}^3 - 1{u_1}^2\right)$	$T_{3,\mathrm{SR}} = \left(\frac{1}{3}{u_0}^3 - 1.00{u_1}^2\right)$

discovery scheme identifies all the terms correctly with very close coefficient values. This confirms that both the databases we simulated for the ML-based discovery scheme and the set-up are appropriate for further discovery of e-CQs.

(ii) New empirically but not exactly CQs

Olver [31] proved that there are only three exact CQs for the RLW equation. Using our ansatz form to approximate the interaction of two-solitary waves at arbitrary separation distance (L) in the RLW equation, however, requires four CQs in total. Hence, we deploy the ML-based discovery scheme to discover two more new e-CQs, which are almost conserved for the RLW equation. The rationale behind this is the known similarity in the behaviour of waves in the KdV and the RLW equation for the solitary wave interaction problem.

We report the new fourth-order $T_{4,\mathrm{SR}}$ and fifth-order $T_{5,\mathrm{SR}}$ e-CQs for the RLW equation:

$$T_{4,\mathrm{SR}} = 1/4u^4 - 6.05u_0{u_1}^2 - 4.00u_0u_2 - \frac{1}{2}{u_0}^2{u_1}^2 + 3.50u_0{u_2}^2 \tag{4.1}$$

and

$$T_{5,\mathrm{SR}} = 1/5u^5 - 2.83{u_0}^3{u_1}^2 - 1.37{u_0}^3u_1u_3 - 8.43{u_1}^2 - 4.07u_1u_3. \tag{4.2}$$

The accuracy of these two e-CQs is shown in figure 4, where we first explore the variation of the discovered e-CQs and their individual terms during a typical two-solitary wave interaction. Similar to the variation pattern for exact CQs for KdV equation shown in figure 2, the variation of each term is quite significant for the fourth-order e-CQ (a) and the fifth-order e-CQ (b) with over 20% change in the leading term. However, the sum of all the discovered terms (i.e. the e-CQ) only shows tiny variations during the entire interaction of two-solitary waves with the maximum percentage variation being less than 1%. The variation of the terms appears to indicate that the discovered e-CQs are not trivial solutions.

We also present the maximum percentage variation of the e-CQs for all of the test cases in figure 4*c*,*d*, where the maximum percentage variation measures the largest percentage fluctuation of a quantity during the entire simulation. We define the maximum percentage variation as

$$\text{Maximum percentage variation} = \frac{\max T_{(i,t\in[0,R])} - \min T_{(i,t\in[0,R])}}{T_{(i,t)}|_{t=0}} \times 100\%, \tag{4.3}$$

where $T_{(i,t)}$ is the ith-order e-CQs at time instance of t, and R denotes the upper boundary of the time domain. Similar to the examples presented in (a, b), the e-CQ shows tiny variations throughout all of the test cases. The fourth-order e-CQ shows better performance and has less variation than the fifth-order e-CQ, which is to be expected since the fifth-order e-CQ involves higher-order derivatives.

(iii) Two-solitary waves interaction example with RLW

Based on the e-CQs, we discovered with the ML-based equation discovery scheme, we now adopt a similar CQ model for solving the two-solitary waves interaction problem governed by the RLW equation. We use all three re-discovered exact CQs and a fourth-order e-CQ in equation (5.1) to solve the equation sets, and present the results in figure 5.

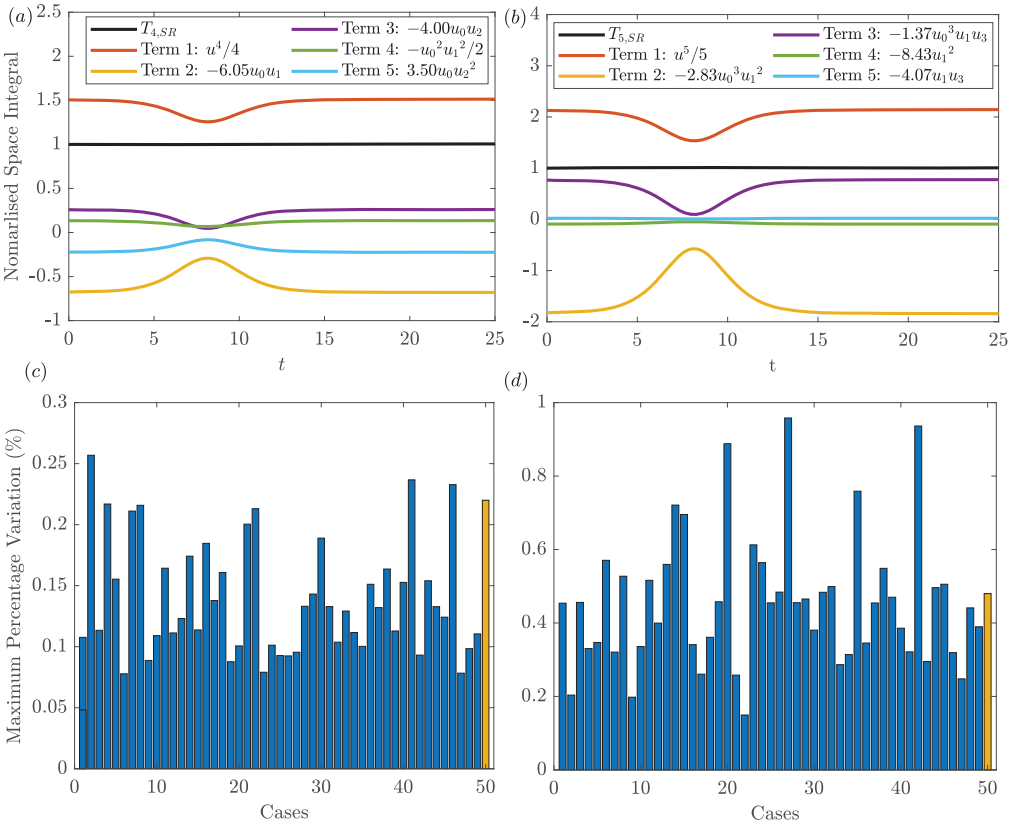

Figure 4. Variation of the CQs and individual terms for (*a*) fourth- and (*b*) fifth-order e-CQ of the case with for the perfect merge with amplitudes $b_1 = \sqrt{2}/2$ and $b_2 = 1/(2\sqrt{3})$. (*c*,*d*) The maximum percentage variation of (*c*) the fourth- and (*d*) fifth-order e-CQs for two-solitary wave interactions governed by RLW. The yellow bars show the results for the case in (*a*) and (*b*).

The two-solitary wave interactions governed by the RLW equation exhibit similar behaviour with those governed by KdV equation, which also has three types: exchange identity, perfect merge and run through. Similarly, figure 5*a–c* shows the exchange identity interaction, figure 5*d–f* shows the perfect merge and figure 5*g–i* shows the run-through type. For all three types of interaction, the CQ model with ML-discovered e-CQs approximates the two-solitary wave interaction phenomenon accurately without the need to solve the RLW equation directly via time-stepping. The largest error appears at panel (*c*) as the exchange identity with minimum distance between two-solitary wave peaks L. This is again probably due to the imperfect estimation of L. For this exchange of identity case at minimal distances, it is interesting to see that we actually need fewer CQs when using symmetry [27,29] and should obtain the same answer. Using the current more generalized ansatz leads to results with slightly larger errors, which is probably due to the slight inaccuracy in the discovered e-CQs and numerical inaccuracy when learning these e-CQs.

We note that the discovered fifth-order e-CQ (equation (5.2)) can also be used in an exchangeable manner to the fourth-order e-CQ in our CQ model, suggesting both e-CQs have a similar predictive power when modelling solitary wave interaction for the RLW equation (details in appendix A). In addition, for the RLW equation, it is now known that the interaction of a pair of solitary waves is not 'clean' [41], the emerging waves are slightly different and very small oscillations are also generated, which might contribute to the difference observed herein.

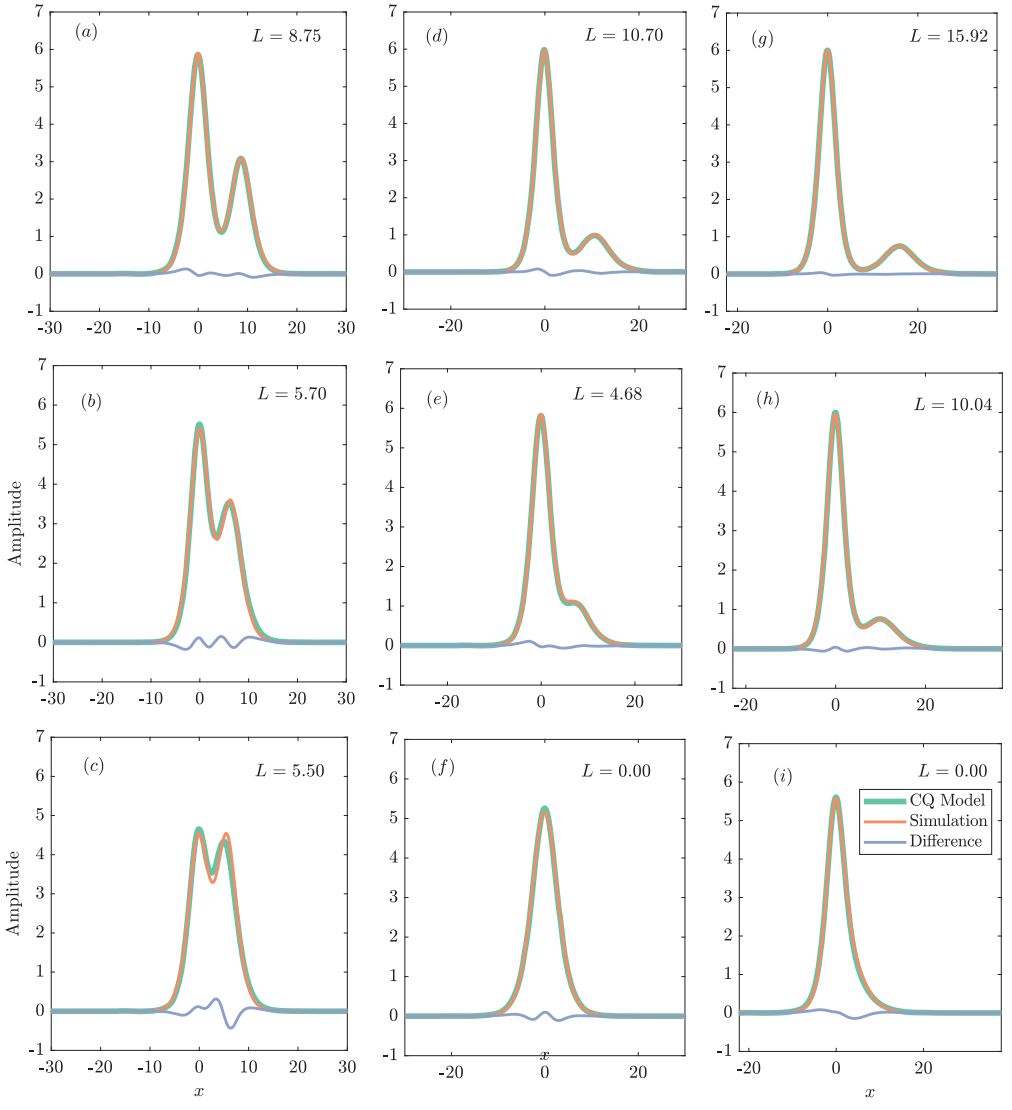

Figure 5. Two-solitary wave interactions governed by the RLW equation for exchange of identity in (*a–c*) ($b_1 = \sqrt{2}/2$, $b_2 = 1$), perfect merging in (*d–f*) ($b_1 = \sqrt{2}/2$, $b_2 = 1/(2\sqrt{3})$) and run through in (*g–j*) ($b_1 = \sqrt{2}/2$, $b_2 = 1/4$). The separation distance L is the distance between the two superposed humps in the approximated form.

5. Evolution of superposed hump amplitudes for the KdV and RLW equations

The new parametrization and the e-CQ discovered by ML enable our CQ model to track and predict the entire structure of the two-solitary wave interactions as two superposed humps for both KdV and RLW equations. We present the evolution of these two hump amplitudes in figure 6, where the continuous lines show the fitted ansatz (equation (2.9)) to the numerical simulations and the square dots show the predicted amplitude using CQ model. The overall interaction patterns for both KdV and RLW equations are similar, and the CQ model can provide accurate estimations for the amplitude of solitary waves via the assumed double hump form when compared to numerical results. The CQ model with e-CQs applied to the RLW equation has slightly larger errors than for the KdV equation, which is probably due to the relative width of the solitary wave governed by RLW equation being larger than the solitons governed by the

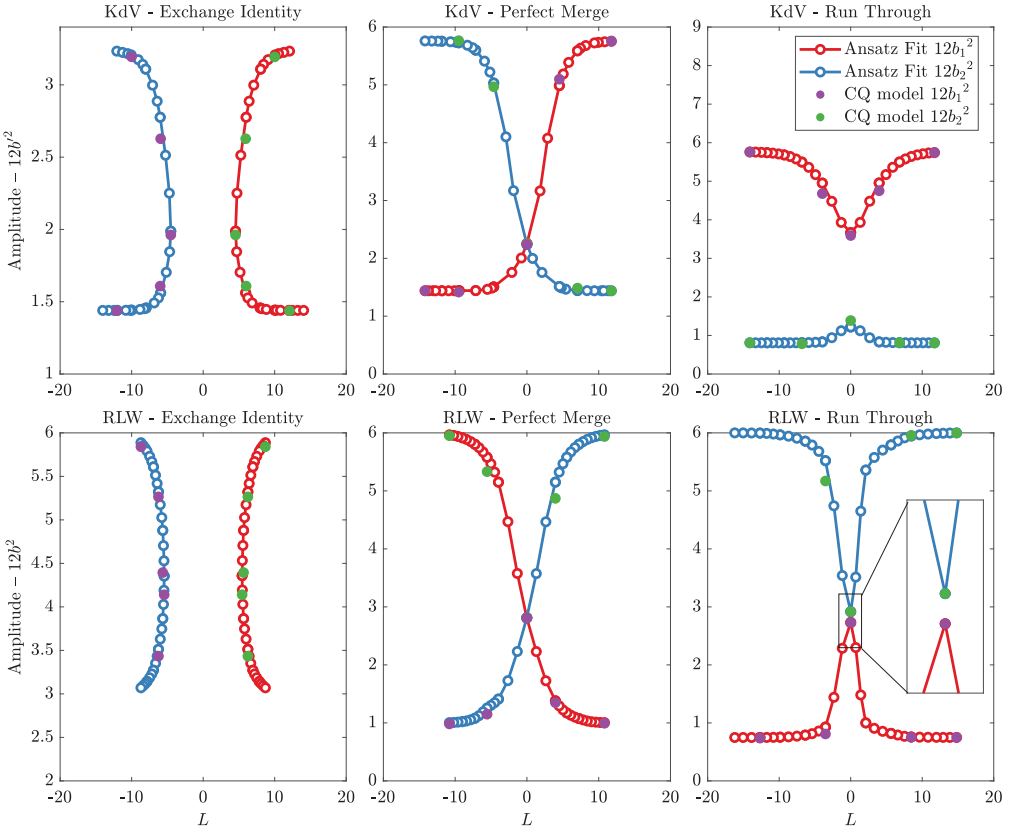

Figure 6. Evolution of superposed hump amplitudes for the KdV and RLW equations for both ansatz fitting and CQ model under different interaction patterns for both KdV and RLW equations. The separation distance L is the distance between the two superposed humps in the approximated form. We scale the KdV solutions by $b' = b/\sqrt{300}$ so within a similar numerical range of the solutions from RLW equation.

KdV equation. This gives rise to extra difficulties in finding the accurate value of L (distance between two superposed humps in the approximated form) when two humps are close to each other.

6. Discussions and mathematical analysis

Finally, we would like to discuss some possible significance of the discovered e-CQs for the RLW equation. We first compare the overall variation of each term for the exact KdV CQs presented in figure 2 and the e-CQs for RLW in figure 4*a*,*b*. During the two-solitary wave interaction, the behaviour of $T_{4,\mathrm{KdV}}$, $T_{5,\mathrm{KdV}}$ are superficially similar to the discovered $T_{4,\mathrm{SR}}$, $T_{5,\mathrm{SR}}$ for the RLW equations. This is expected since the RLW and KdV equations are equivalent reductions of the water wave equations and solutions of the two equations are known to remain 'in the vicinity' of each other in the appropriate limits.

In addition, we have also explored other possibilities of e-CQs in the same order by specifying other possible leading terms inspired by the KdV exact CQs. For example, for the fourth-order e-CQ for RLW $T_{4,\mathrm{SR}}$, we investigated the additional conditions $F(u) = 3u_0{u_1}^2$ and $F(u) = 1.8{u_2}^2$, using these the new PySR-CQ package failed to discover any formulations at any level of complexity.

This seems to suggest that the discovered e-CQs required to have the same leading-order term as the exact CQs for the KdV equation is not a coincidence. We observe that the search

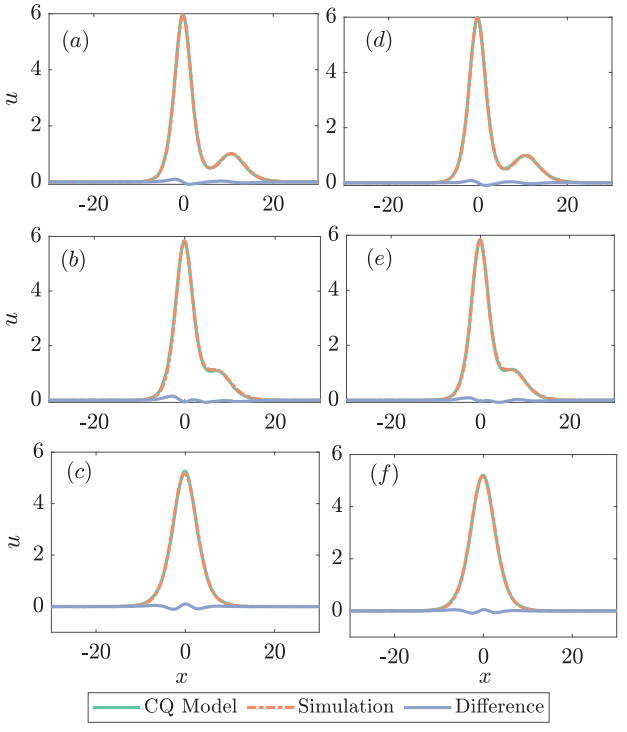

Figure 7. Two-soliton interactions governed by the RLW equation for perfect merging using CQ model with the first three CQs and (*a–c*) the fourth-order e-CQ $T_{4,\mathrm{SR}}$, (*d–f*) the fifth-order e-CQ $T_{5,\mathrm{SR}}$.

for discovered e-CQs must start with the same leading-order term $\frac{1}{k}u^k$ as Miura's KdV exact CQ forms, otherwise no e-CQ form can be identified. On the other hand, the mathematical structures in the following terms $I(u)$ are yet to be fully understood, but we find that all the following terms appearing in the exact CQs for KdV equation drop out during the discovery for the e-CQ for RLW. This might help explain why the additional format of e-CQs with the leading term fixed as other terms in the KdV exact CQ is not possible, and also why the expression for the KdV exact CQ is not transferable across into the RLW equation modelling.

7. Conclusion

In this study, we demonstrate a fast and robust approach to model the solitary wave interaction problem using CQs. This CQ model first requires the choice of a simple ansatz for the shape of interacting solitary waves with arbitrary constants tending to the known solitary wave structure when separated. The CQs of the governing equation are then used to constrain this chosen form, and with sufficient numbers of these CQs, the equation system can be solved. The solved form provides an accurate approximation of the two-solitary wave interactions for all three different patterns: exchange identity, perfect merge and run-through.

We first demonstrate the accuracy of this CQ model via examples of two-solitary wave interaction governed by the KdV equation, where there are an infinite number of CQs. We further extend the CQ model to accommodate an apparently very similar equation with only a limited number of exact CQs. We develop a new ML discovery scheme for finding e-CQs. We use this new scheme to find two new e-CQs for the RLW equation and then use the first e-CQ in our model. The new CQ model performs well in approximating two-solitary wave interactions governed by the RLW equation.

This study illustrates a new way of incorporating ML in solving PDEs. Instead of trying to use conservation laws to constrain ML models, these CQs can be directly used to approximate the states of wave solution to the PDE at all times. The evolution in time of the whole solution to the PDE is then reduced to the evolution of shape and size parameters in terms of the control parameter, i.e. the distance between the humps approximating the solution. Although our examples in this study potentially hint at a brand-new approach, some limitations should be noted:

(i) A low-dimensional symbolic parameterization for the solutions of the PDE is needed for such a CQ model.
(ii) The number of CQs that can be derived or identified from the governing equation must be equal to, or larger than the parameters used during the parameterization.

These two limitations constrain the direct application of this approach at this time due to the difficulties in obtaining such parameterization, and only limited CQs can be derived mathematically. However, such parameterization is becoming increasingly common in many applications [42,43] and with recent advances in symbolic ML approaches [44] for systems where prior knowledge is limited. More importantly, these parameterizations, in various forms, can be incorporated into our CQ approach without any further modifications. In addition, given that ML-discovered e-CQs work very well in these CQ models, we believe the new ML discovery scheme proposed herein can lift the constraint on how many CQs are available for a system, which allows the CQ models to have broader utilities for complex systems with more parameters. Therefore, we foresee this to be a tangible application with widespread explainable ML and other modern data-driven techniques.

Data accessibility. The data for soliton simulations can be found at [45].
Declaration of AI use. We have not used AI-assisted technologies in creating this article.
Authors' contributions. T.T.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, validation, visualization, writing—original draft, writing—review and editing; P.H.T.: conceptualization, data curation, investigation, methodology, supervision, writing—review and editing; Y.C.: methodology, software, supervision, writing—review and editing; T.A.A.A.: conceptualization, data curation, funding acquisition, investigation, methodology, project administration, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.
Conflict of interest declaration. We declare we have no competing interests.
Funding. T.T. is supported by Schmidt Sciences, LLC.

Appendix A. CQ model with fifth e-CQs for RLW equation

We present the CQ model prediction using the first three exact CQs plus the fourth-order e-CQ and one using the first three CQs plus the fifth-order e-CQ in figure 7. We report almost identical results between the left-hand columns (with fifth-order e-CQ) and the right-hand columns (with fourth-order e-CQ), suggesting that the fifth-order e-CQ can be used in an exchangeable way to the fourth-order e-CQ in our model. Similar tests have been performed for both exchange identity and run-through cases, where we arrive at the same conclusions—using either the fourth-order e-CQ or fifth-order e-CQ will give very similar solitary wave profiles. This suggests that both e-CQs have a similar predictive power in our CQ model. One minor difference, however, is that higher-order e-CQs usually result in stiffer equations and hence are more difficult to solve.

References

1. Salupere A, Engelbrecht J, Peterson P. 2003 On the long-time behaviour of soliton ensembles. *Math. Comput. Simul.* **62**, 137–147. (doi:10.1016/S0378-4754(02)00178-7)
2. Zabusky NJ, Galvin CJ. 1971 Shallow-water waves, the Korteweg-deVries equation and solitons. *J. Fluid Mech.* **47**, 811–824. (doi:10.1017/S0022112071001393)
3. Hereman W. 2022 Shallow Water Waves and Solitary Waves. In *Solitons* (ed. MA Helal). Encyclopedia of Complexity and Systems Science Series. New York, NY: Springer. (doi:10.1007/978-1-0716-2457-9_480)
4. Zabusky NJ, Kruskal MD. 1965 Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett.* **15**, 240. (doi:10.1103/PhysRevLett.15.240)
5. Roshid HO, Roshid MM, Rahman N, Pervin MR. 2017 New solitary wave in shallow water, plasma and ion acoustic plasma via the GZK-BBM equation and the RLW equation. *Propuls. Power Res.* **6**, 49–57. (doi:10.1016/j.jppr.2017.02.002)
6. Seadawy AR. 2014 Stability analysis for Zakharov–Kuznetsov equation of weakly nonlinear ion-acoustic waves in a plasma. *Comput. Math. Appl.* **67**, 172–180. (doi:10.1016/j.camwa.2013.11.001)
7. Seadawy AR. 2016 Stability analysis solutions for nonlinear three-dimensional modified Korteweg–de Vries–Zakharov–Kuznetsov equation in a magnetized electron–positron plasma. *Phys. A* **455**, 44–51. (doi:10.1016/j.physa.2016.02.061)
8. Blanco-Redondo A, de Sterke CM, Xu C, Wabnitz S, Turitsyn SK. 2023 The bright prospects of optical solitons after 50 years. *Nat. Photonics* **17**, 937–942. (doi:10.1038/s41566-023-01307-9)
9. Agrawal GP. 2000 Nonlinear fiber optics. In *Nonlinear Science at the Dawn of the 21st Century* (eds P. L. Christiansen, M. P. Sørensen, A. C. Scott), pp. 195–211. New York, NY: Springer.
10. Seadawy AR, Ali KK, Nuruddeen R. 2019 A variety of soliton solutions for the fractional Wazwaz-Benjamin-Bona-Mahony equations. *Results Phys.* **12**, 2234–2241. (doi:10.1016/j.rinp.2019.02.064)
11. Korteweg DJ, De Vries G. 1895 On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *Lond., Edinb., Dublin Phil. Mag. J. Sci.* **39**, 422–443. (doi:10.1080/14786449508620739)
12. Benjamin TB, Bona JL, Mahony JJ. 1972 Model equations for long waves in nonlinear dispersive systems. *Phil. Trans. R. Soc. Lond. A* **272**, 47–78. (doi:10.1098/rsta.1972.0032)
13. Gardner CS, Greene JM, Kruskal MD, Miura RM. 1967 Method for solving the Korteweg–deVries equation. *Phys. Rev. Lett.* **19**, 1095. (doi:10.1103/PhysRevLett.19.1095)
14. Glampedakis K, Johnson AD, Kennefick D. 2017 Darboux transformation in black hole perturbation theory. *Phys. Rev. D* **96**, 024036. (doi:10.1103/PhysRevD.96.024036)
15. Wazwaz AM. 2008 The extended tanh method for new compact and noncompact solutions for the KP–BBM and the ZK–BBM equations. *Chaos, Solitons & Fractals* **38**, 1505–1516. (doi:10.1016/j.chaos.2007.01.135)
16. Salas AH, Frias BA. 2010 New periodic and soliton solutions for the generalized BBM and Burgers–BBM equations. *Appl. Math. Comput.* **217**, 1430–1434. (doi:10.1016/j.amc.2009.05.068)
17. Khater MMA, Attia RAM, Lu D. 2018 Modified auxiliary equation method versus three nonlinear fractional biological models in present explicit wave solutions. *Math. Comput. Appl.* **24**, 1. (doi:10.3390/mca24010001)
18. Ren J, Ilhan OA, Bulut H, Manafian J. 2021 Multiple rogue wave, dark, bright, and solitary wave solutions to the KP–BBM equation. *J. Geom. Phys.* **164**, 104159. (doi:10.1016/j.geomphys.2021.104159)
19. Karakoc SBG, Bhowmik SK. 2019 Galerkin finite element solution for Benjamin–Bona–Mahony–Burgers equation with cubic B-splines. *Comput. Math. Appl.* **77**, 1917–1932. (doi:10.1016/j.camwa.2018.11.023)
20. Montanelli H, Bootland N. 2020 Solving periodic semilinear stiff PDEs in 1D, 2D and 3D with exponential integrators. *Math. Comput. Simul.* **178**, 307–327. (doi:10.1016/j.matcom.2020.06.008)
21. de la Hoz F, Cuesta CM. 2016 A pseudo-spectral method for a non-local KdV–Burgers equation posed on R. *J. Comput. Phys.* **311**, 45–61. (doi:10.1016/j.jcp.2016.01.031)
22. Bai D, Zhang L. 2011 Numerical studies on a novel split-step quadratic B-spline finite element method for the coupled Schrödinger–KdV equations. *Commun. Nonlinear Sci. Numer. Simul.* **16**, 1263–1273. (doi:10.1016/j.cnsns.2010.06.003)

23. Su CH, Mirie RM. 1980 On head-on collisions between two solitary waves. *J. Fluid Mech* **98**, 509–525. (doi:10.1017/S0022112080000262)
24. Craig W, Guyenne P, Hammack J, Henderson D, Sulem C. 2006 Solitary water wave interactions. *Phys. Fluids* **18**, 057106. (doi:10.1063/1.2205916)
25. Kochanov MB, Kudryashov NA, Sinel'Shchikov DI. 2013 Non-linear waves on shallow water under an ice cover. Higher order expansions. *J. Appl. Math. Mech.* **77**, 25–32. (doi:10.1016/j.jappmathmech.2013.04.004)
26. Adcock TAA, Taylor PH. 2009 Focusing of unidirectional wave groups on deep water: an approximate nonlinear Schrödinger equation-based model. *Proc. R. Soc. A* **465**, 3083–3102. (doi:10.1098/rspa.2009.0224)
27. Taylor PH. 2016 A simple approach for shallow-water solitary wave interactions. In *Proc. 20th Australasian Fluid Mechanics Conf, University of Western Australia, 2016*. Australasian Fluid Mechanics Society.
28. You X, Xu H, Sun Q. 2022 Analysis of BBM solitary wave interactions using the conserved quantities. *Chaos, Solitons & Fractals* **155**, 111725. (doi:10.1016/j.chaos.2021.111725)
29. You X, Xu H, Sun Q. 2023 Analysis of soliton interactions of modified Korteweg–de Vries equation using conserved quantities. *Phys. Scr.* **98**, 085224. (doi:10.1088/1402-4896/ace567)
30. Miura RM, Gardner CS, Kruskal MD. 1968 Korteweg–de Vries equation and generalizations. II. Existence of conservation laws and constants of motion. *J. Math. Phys.* **9**, 1204–1209. (doi:10.1063/1.1664701)
31. Olver PJ. 1979 Euler operators and conservation laws of the BBM equation. In *Mathematical Proc. of the Cambridge Philosophical Society*, vol. 85, pp. 143–160. Cambridge University Press.
32. Lewis JC, Tjon J. 1979 Resonant production of solitons in the RLW equation. *Phys. Lett. A* **73**, 275–279. (doi:10.1016/0375-9601(79)90532-2)
33. Hereman W. 2006 Symbolic computation of conservation laws of nonlinear partial differential equations in multi-dimensions. *Int. J. Quantum Chem.* **106**, 278–299. (doi:10.1002/qua.20727)
34. Liu Z, Tegmark M. 2021 Machine learning conservation laws from trajectories. *Phys. Rev. Lett.* **126**, 180604. (doi:10.1103/PhysRevLett.126.180604)
35. Lu PY, Dangovski R, Soljačić M. 2023 Discovering conservation laws using optimal transport and manifold learning. *Nat. Commun.* **14**, 4744. (doi:10.1038/s41467-023-40325-7)
36. Belyshev A, Kovrigin A, Ustyuzhanin A. 2024 Beyond dynamics: learning to discover conservation principles. *Mach. Learn.: Sci. Technol.* **5**, 025055. (doi:10.1088/2632-2153/ad4a20)
37. Cranmer M. 2023 Interpretable machine learning for science with PySR and SymbolicRegression. jl. (http://arxiv.org/abs/2305.01582)
38. Russell JS. 1845 "Report on Waves". Report of the fourteenth meeting of the British Association for the Advancement of Science, York, September 1844 (London 1845), Plates XLVII–LVII, pp. 311–390.
39. Yoneyama T. 1984 The Korteweg–de Vries two-soliton solution as interacting two single solitons. *Prog. Theor. Phys.* **71**, 843–846. (doi:10.1143/PTP.71.843)
40. Eilbeck J, McGuire G. 1975 Numerical study of regularized long-wave equation. 1. Numerical-methods. *J. Comput. Phys.* **19**, 43–57. (doi:10.1016/0021-9991(75)90115-1)
41. Abdulloev KO, Bogolubsky I, Makhankov V. 1976 One more example of inelastic soliton interaction. *Phys. Lett. A* **56**, 427–428. (doi:10.1016/0375-9601(76)90714-3)
42. Grepl MA, Nguyen NC, Veroy K, Patera AT, Liu GR. 2007 Certified rapid solution of partial differential equations for real-time parameter estimation and optimization. In *Real-time PDE-Constrained Optimization*, pp. 199–216. Philadelphia: SIAM. (doi:10.1137/1.9780898718935.ch10)
43. Tang T, Adcock TA. 2022 A reduced order model for space–time wave statistics using probabilistic decomposition–synthesis method. *Ocean Eng.* **259**, 111860. (doi:10.1016/j.oceaneng.2022.111860)
44. Zhang M, Kim S, Lu PY, Soljačić M. 2023 Deep learning and symbolic regression for discovering parametric equations. *IEEE Trans. Neural Netw. Learn. Syst.* **35**, 16 775–16 787. (doi:10.1109/TNNLS.2023.3297978)
45. Tang T, Taylor PH, Chen Y, Adcock TAA. 2025 Solitary wave interactions modelled using conserved and empirically conserved quantities discovered with machine learning. Zenodo. (doi:10.5281/zenodo.15026071)