ORIGINAL PAPER

Deep learning discovery of macroscopic governing equations for viscous gravity currents from microscopic simulation data

Junsheng Zeng[1] · Hao Xu[2] · Yuntian Chen[1] · Dongxiao Zhang[3,4]

Abstract

Although deep learning has been successfully applied in a variety of science and engineering problems owing to its strong high-dimensional nonlinear mapping capability, it is of limited use in scientific knowledge discovery. In this work, we propose a deep learning based framework to discover the macroscopic governing equation of an important geophysical process, i.e., viscous gravity current, based on high-resolution microscopic simulation data without the need for prior knowledge of underlying terms. For two typical scenarios with different viscosity ratios, the deep learning based equations exactly capture the same dominant terms as the theoretically derived equations for describing long-term asymptotic behaviors, which validates the proposed framework. Unknown macroscopic equations are then obtained for describing short-term behaviors, and additional deep-learned compensation terms are eventually discovered. Comparison of posterior tests shows that the deep learning based PDEs actually perform better than the theoretically derived PDEs in predicting evolving viscous gravity currents for both long-term and short-term regimes. Moreover, the proposed framework is proven to be very robust against non-biased data noise for training, which is up to 20%. Consequently, the presented deep learning framework shows considerable potential for discovering unrevealed intrinsic laws in scientific semantic space from raw experimental or simulation results in data space.

Key points
● Macroscopic governing equation of viscous gravity current is discovered on the basis of high-resolution microscopic simulation data without the need of specifying the potential terms.
● Deep learning framework captures the dominant terms which agree with theoretically derived equations.
● Compensation PDE terms and modifying coefficients are discovered for interpreting the deviation between theoretical predictions and simulation results.

✉ Dongxiao Zhang
zhangdx@sustech.edu.cn

1 Frontier Research Center, Peng Cheng Laboratory, Shenzhen 518000, People's Republic of China

2 College of Engineering, Peking University, Beijing 100871, People's Republic of China

3 Eastern Institute for Advanced Study, Eastern Institute of Technology, Ningbo 315200, People's Republic of China

4 School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, People's Republic of China

1 Introduction

Deep learning algorithms are currently being successfully applied in numerous science and engineering fields, such as physics [1, 2], earth science [3, 4], and computer vision and speech recognition [5]. The advantage of deep learning algorithms is their strong capability of constructing high-dimensional nonlinear mappings. Therefore, a deep learning based surrogate model may be trained to effectively describe a physical process based on collected observation data. However, this type of surrogate model constitutes a "black box" with poor explainability, and fails to deepen our understanding of the essences of the physical process, which greatly decreases its usefulness.

This constitutes a challenging problem, which combines knowledge discovery and explainable machine learning [6]. Humans' understanding of the world is based on semantic space. In semantic space, people are able to understand and construct rules for accurate inferences and analyses. In contrast, pure data-driven surrogates exist in data space. In data space, the expressive capability of surrogates is strong with

infinite possibilities. It is particularly challenging, however, to interpret useful information hidden in amounts of data. Therefore, the key to explainability for a machine-learning based surrogate is to transform high-dimensional mapping from data space into semantic space, and obtain an intuitive and interpretable understanding about the mapping. In modern sciences, the most widely acceptable semantic form or knowledge expression form is represented by partial/ordinary differential equations (PDEs/ODEs), which are usually concise and easy-to-understand.

To realize the transformation of surrogates from data space into semantic space, data-driven discovery of partial differential equations (PDEs) is a feasible solution, and thus has recently attracted increasing attention. Usually, PDEs consist of several complex differential terms, and the candidate library of potential terms can be very large. Determination of how to accurately reproduce the differential terms and find parsimonious PDEs become key objectives. Recently, Xu et al. [7] proposed a deep learning-genetic algorithm (DLGA) framework. In this framework, a deep neural network is utilized to calculate derivatives and generate meta-data, and the genetic algorithm is employed to discover the form of PDEs without the need of including the true terms in the initial guess. The deep neural network has been utilized in PDE discovery in previous works [8–10], because the derivatives calculated by automatic differentiation are more accurate and robust to noise. In the process of the genetic algorithm, genomes are composed of several basic genes, which can be adjusted according to the situation of the discovery process, which markedly increases both flexibility and practicability. Compared to other sparse regression methods, including LASSO [11], SINDY [12, 13], sequential threshold ridge regression (STRidge) [14, 15], and sparse Bayesian regression [16], the genetic algorithm does not need a complete candidate library beforehand [17], which may be impossible for many real-world applications. Meanwhile, the integral form is an efficient approach to facilitate the PDE discovery process and increase the accuracy of discovered PDEs [18].

In this work, we attempt to employ the deep learning based PDE discovery framework [18] to discover unknown macroscopic equations for a real geophysical process, i.e., viscous gravity current. Viscous gravity current is an important natural phenomenon in geophysics [19–21], and typical scenarios include displacement flows in oil reservoirs [22], sea water intrusion [23, 24], water injection in geothermal reservoirs [25], pollutant dispersion through groundwater [26], and proppant transport in hydraulic fractures [21, 27]. Investigating its one-dimensional governing equation has wide practical application significance. Essentially, viscous gravity current is a gravity-driven flow constrained in porous media or narrow vertical fractures due to the density difference between intruding and *in-situ* fluid, which forms a remarkable sharp interface

at macro-scale, i.e., the current front. As a consequence, identification of a macroscopic equation for describing the evolution of current front height is a crucial issue. In the existing literature, there exist many theoretically derived macroscopic equations for viscous gravity currents under various conditions, such as different viscosity ratio [19], multi-layer porous media [28], non-Newtonian intruding fluid [29–31], and varying horizontal permeability [32]. However, these works largely focus on long-term asymptotic behaviors of viscous gravity currents, and macroscopic equations for capturing short-term behaviors remain undetermined.

Different from previous works in data-driven PDE discovery, the training data of current front height are directly extracted from high-resolution microscopic simulation results instead of solving already-known model equations, and the long-term theoretical PDEs are only utilized in posterior tests as a reference. Consequently, no prior knowledge about the underlying equations is needed or utilized in the proposed method. With the assistance of the deep learning based PDE discovery method, on the one hand, we are able to validate the method by quantitative comparison of discovered PDEs and theoretical PDEs for long-term behaviors of viscous gravity currents both in data space and scientific semantic space. On the other hand, it is also possible for us to transform raw simulation data for describing short-term behaviors from data space into discovered PDEs in scientific semantic space. Subsequently, by quantitatively comparing discovered PDEs with theoretical PDEs, the hidden mechanisms can be elucidated, which can be expressed as compensation terms for modifying the original theoretically derived PDEs to capture short-term behaviors in viscous gravity currents.

The remainder of this paper is organized as follows. In Section 2, the two-dimensional microscopic governing equation and numerical solution of viscous gravity current related to training data preparation are introduced. Theoretical derivations of one-dimensional governing equations for viscous gravity currents are also briefly reviewed as references for subsequential quantitative comparisons. Then, the proposed deep learning based PDE discovery framework is introduced. In Section 3, PDE discovery results based on the deep learning framework for both long-term and short-term regimes of viscous gravity currents are demonstrated, and the effects of data noise are discussed. Finally, the conclusions are drawn in Section 4.

2 Methodology

2.1 Preparation for training data of viscous gravity current

Generally, for investigating fluid mechanics, such as viscous gravity currents, three fundamental approaches usually exist, i.e., theoretical analysis, physical experiments, and

numerical simulation [33]. Particularly, theoretical analysis offers strong explainability and scientific consistency. Macroscopic equations can be rigorously derived based on fundamental laws and rational assumptions, and conveniently extended for investigating similar physical processes. The prediction accuracy of theoretical analysis, however, depends greatly on the validity assumption. On the other hand, although physical experiments and microscopic numerical simulations can provide a large amount of observation data and accurate prediction results, they are both time-consuming and possess poor transferring capability. Different from conventional theoretical analysis for obtaining macroscopic equations, we aim to find one-dimensional macroscopic PDEs in scientific semantic space to describe the evolving behaviors of front height based on raw observation data in data space. In this work, the raw observation data are extracted from refined numerical simulation, also referred to as microscopic simulation.

Particularly, we consider a viscous gravity current process in a two-dimensional rectangle vertical fracture with non-penetration boundaries, as shown in Fig. 1(a). The permeability and porosity of the fracture are considered as constant, and fluid leak-off, as well as other source terms, are ignored for simplicity. The domain is initially vertically divided by two types of fluid with different density, i.e., ρ_1 and ρ_2, and different viscosity, i.e., μ_1 and μ_2. Then, determination of how the current front height $h(t, x)$ evolves at macro-scale constitutes the primary aim.

In fact, the fluid motion in this two-dimensional system is governed by Darcy's equation and the continuity equation:

$$\vec{u} = -(k/\mu)(\nabla P - \rho \vec{g}), \tag{1}$$

$$\nabla \cdot \vec{u} = 0, \tag{2}$$

where k is the permeability of fracture; ρ and μ are the fluid viscosity and density, respectively; P is the pressure; $\vec{g}$ is the gravity acceleration; and $\vec{u}$ is the two-dimensional velocity of fluid.

The interface between two fluids can be tracked with a conservative level-set equation [34]:

$$\frac{\partial C}{\partial t} + \nabla \cdot (C\vec{u}) = 0 \tag{3}$$

$$\frac{\partial C}{\partial \tau} + \nabla \cdot \left[C(1-C)\frac{\nabla C}{\|\nabla C\|} \right] = \varepsilon \Delta C \tag{4}$$

where C is the level set function or color function; $\vec{u}$ is the two-dimensional velocity of fluid; τ is the artificial time; and ε is the artificial viscosity. Note that Eq. (3) and (4) have been written in a conservative form by utilizing Eq. (2), and the artificial viscosity ε is introduced in consideration of numerical stability. The above equations indicate the advection step and the artificial compression step, respectively.

Furthermore, by substituting Eq. (1) into the Eq. (2), a Poisson equation for pressure independent of fluid velocity can be obtained:

$$\nabla^2 P = \frac{g\mu}{k} \frac{\partial \rho}{\partial z} \tag{5}$$

Fig. 1 Illustrations of viscosity gravity currents. (**a**) Sketch of viscous gravity current in a rectangle vertical fracture ($\rho_1 > \rho_2$). The green dashed vertical line indicates the initial interface between the two fluids, and the green solid curve $h(t, x)$ indicates the spatio-temporal evolving current front height. (**b**) Evolution of viscous gravity current for case I, where viscosity of two fluids satisfy $\mu_1 = \mu_2$. The yellow region (level-set function $C = 1$) indicates heavy fluid, and the blue region ($C = 0$) indicates light fluid. (**c**) Evolution of viscous gravity current for case II, where viscosity of two fluids satisfy $\mu_1 \gg \mu_2$
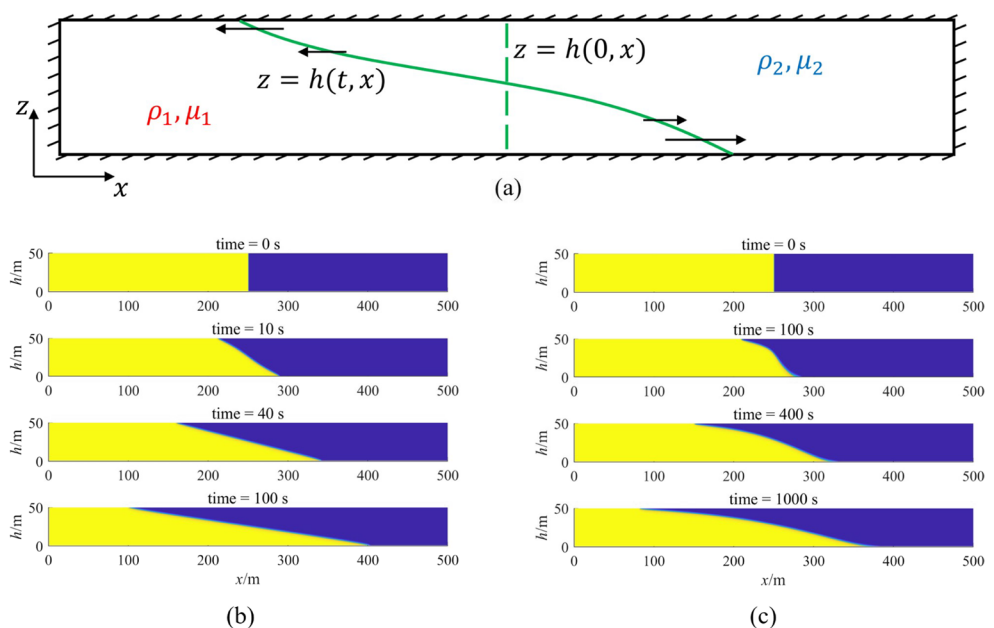
Based on the domain setting as illustrated in Fig. 1(a), the initial and boundary conditions for the above two-dimensional problem can be defined as follows:

$$C(t=0,x,z) = \begin{cases} 1 & x \le L/2 \\ 0 & x > L/2 \end{cases}, P(t=0,x,z) = \left[C\rho_1 + (1-C)\rho_2\right]g(H-z); \quad (6)$$

$$\vec{n} \cdot \nabla C = 0, \vec{n} \cdot \left(\nabla P - \rho \vec{g}\right) = 0, \forall (x,z) \in \partial\Omega; \quad (7)$$

where L and H are the length and height of simulation domain Ω, respectively; $\vec{n}$ indicates the normal direction of four non-penetration domain boundaries; $\partial\Omega$ indicates the non-penetration boundaries.

To obtain the solution of the above two-dimensional system, for each time step, fluid pressure and velocity are first determined by solving the pressure equation, and color function C is then updated according to Eq. (3) and (4). Pressure calculation and evolution of color function are sequentially solved through the simulation timeline. Numerically, the pressure equation is discretized with the center difference scheme, and the AGMG solver based on the multi-grid technique [35, 36] is adopted to solve the discretized linear system. For the advection step (3), the third-order Runge–Kutta TVD scheme [36] and fifth-order WENO scheme [37, 38] are applied for temporal and spatial discretizations, respectively. For the artificial compression step (4), a center difference scheme is adopted, and after several artificial time steps, the color function converges to maintain a constant interface width. The computational bottleneck of the above procedure is solving the pressure equation. During each timestep, a large sparse linear system needs to be solved. The computational complexity of the multi-grid method is of $O(N\log(N))$, where N is the 2D mesh number. Besides, memory requirement is another computational constraint if full matrix storage is adopted instead of sparse matrix storage.

In this work, two typical scenarios of viscous gravity currents are investigated. For case I, viscosity of the heavy fluid is assumed to be equal to that of light fluid, i.e., $\mu_1 = \mu_2$; whereas, for case II, it is assumed that $\mu_1 >> \mu_2$. Figure 1(b) and (c) illustrate the short-term evolving history of current fronts in the two cases. It is observed that the current front tends to be linear-shaped in case I, while a curve-shaped front is seen in case II, during the late time of simulation. By tracking the contour line of level set function $C = 0.5$ in the microscopic simulation results, the spatio-temporal evolving history of current front height $h(t,x)$ is then obtained.

Without loss of generality, related variables, including front height h, horizontal location x, and time t, are normalized through non-dimensional analysis prior to subsequential deep learning based PDE discovery, and the dimensionless variables are defined as:

$$h^* = \frac{h}{H}; x^* = \frac{x}{H}; t^* = \frac{t\Delta\rho gk}{\mu_1 H} \quad (8)$$

where $\Delta\rho = \rho_1 - \rho_2$ is the density difference between the two fluids. It is worth noting that, through the nondimensionalization formula above, changing related physical parameters, including $\Delta\rho$, g, k, μ_1, and H, does not affect the spatio-temporal evolution of dimensionless front height $h^*(t^*, x^*)$. In other words, the obtained PDEs by the deep learning framework are indeed universal forms for similar processes if simulation data are nondimensionalized by Eq. (8).

2.2 Theoretically derived PDEs for viscous gravity currents

Although the theoretically derived PDEs are obtained based on various assumptions and their applications are usually limited, they can still be regarded as baselines, or asymptotic solutions for the to-be-investigated physical process. In this part, we briefly review the details of theoretically derived PDEs for viscous gravity currents following Gardner et al.'s work [19], which describes the long-term asymptotic behavior of current front height.

Basically, it is assumed that the vertical flow is sufficiently smaller than the horizontal flow, i.e., $u_z << k\rho g/\mu$, where u_z indicates the vertical component of fluid velocity. Under this condition, the pressure gradient in the vertical direction can be simplified as $\partial P/\partial z = -\rho g$. Let P_1 and P_2 denote the fluid pressure in fluid 1 and 2 across a vertical line inside the domain, respectively. Then, the pressure can be analytically expressed as:

$$P_2 = \pi + \rho_2 g(H-z), z > h(x) \quad (9)$$

$$P_1 = \pi + \rho_2 g(H-h) + \rho_1 g(h-z), z \le h(x) \quad (10)$$

where $h(x)$ indicates the horizontally varying current front height; H is the total domain height; and π is the reference pressure in the horizontal direction.

Then, by utilizing Darcy's law and the mass conservation equation along vertical lines, it is not difficult to find the velocity solution for one fluid:

$$u_{x,1} = k\Delta\rho g \frac{\partial h}{\partial x} / \left(\mu_1 + \mu_2 \frac{h}{H-h}\right) \quad (11)$$

Note that, for phase 1, we also have the continuity equation:

$$\frac{\partial h}{\partial t} + \frac{\partial\left(u_{x,1}h\right)}{\partial x} = 0 \quad (12)$$

By substituting Eq. (11) into (12), the PDE of current height can finally be determined as:

$$\frac{\partial h}{\partial t} - k\Delta\rho g \frac{\partial}{\partial x}\left(\frac{h(H-h)}{\mu_1(H-h)+\mu_2 h}\frac{\partial h}{\partial x}\right) = 0 \qquad (13)$$

Considering the two cases in this work, i.e., $\mu_1 = \mu_2$ and $\mu_1 \gg \mu_2$, and nondimensionalizing the above equation based on Eq. (8), the theoretically derived dimensionless PDEs can be written as:

$$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(h^*\frac{\partial h^*}{\partial x^*} - h^{*2}\frac{\partial h^*}{\partial x^*}\right) \qquad (14)$$

$$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(h^*\frac{\partial h^*}{\partial x^*}\right) \qquad (15)$$

To numerically solve Eq. (14) and (15), initial and boundary conditions need to be specified. As illustrated in Fig. 1(a), the initial condition for dimensionless front height h^* can be expressed as:

$$h^*(t^*=0, x^*) = \begin{cases} 1 & x^* \le L^*/2 \\ 0 & x^* > L^*/2 \end{cases} \qquad (16)$$

where $L^* = L/H$ is the dimensionless domain length. As for the left and right non-penetration solid boundaries, it is clear that $u_{x,1} = 0$ should be satisfied, which implies that:

$$\frac{\partial h^*}{\partial x^*} = 0, x^* = 0 \text{ or } 1 \qquad (17)$$

In this work, two regimes are investigated for both cases, i.e., long-term and short-term regimes. As mentioned previously, long-term regimes imply that nondimensional vertical velocity is sufficiently small, i.e., $u_z^* = u_z\mu_1/(\Delta\rho g k) \ll 1$ [19]. Here, we distinguish these two regimes through another more intuitive indicator, i.e., the dimensionless horizontal spanning distance $\Delta x^* = \|x|_{h^*\to 0} - x|_{h^*\to 1}\|$ of the current front interface. The long-term regime is defined as $\Delta x^* > 6$, and the short-term regime as $0 \le \Delta x^* \le 6$. Under this condition, for both cases, the calculated average nondimensional vertical velocities of long-term regimes are less than 0.05, which is acceptably small. Furthermore, for subsequential deep learning, the amount of training data (spatio-temporal points of front height) of the two regimes is set as the same size.

2.3 Deep learning based PDE discovery

In fact, if ignoring source/sink terms and inhomogeneity of vertical permeability, one-dimensional macroscopic equations for all viscous gravity currents can be expressed in a unified conservative form, i.e., $\partial h^*/\partial t^* + \partial F^*/\partial x^* = 0$, where F^* indicates dimensionless flux. Then, for a to-be-investigated viscous gravity current process, the goal of discovering a concise PDE for front height evolution can be converted into a sparse regression problem of approximating flux F^* through basic algebraic combinations of height h^* and its arbitrary-order derivatives, such as $\partial h^*/\partial x^*$ and $\partial^2 h^*/\partial x^{*2}$. Above all, prior to regression, flux and various derivatives should be calculated in advance based on raw training data. Particularly, flux F^* can be calculated through integration $F^*(t^*, x^*) = \int_{x_0^*}^{x^*}(-\partial h^*/\partial t^*)dx^*$.

Based on this insight, a natural approach is to calculate the derivatives and integration. First, one may fit the data with smoothing techniques, such as Gaussian process, and then the derivatives and integration can be calculated using conventional numerical difference and integration schemes. However, it has been proven in previous literature [39] that these conventional approaches are sensitive to data noise. Usually, there is no guarantee about the quality of collected data in reality. Therefore, when investigating a physical process without any prior knowledge, the PDE discovery approach is expected to be robust enough while accuracy of data reconstruction, as well as derivative and integral calculation, can also be maintained. To solve the mentioned issues, here we apply a well-balanced framework based on deep learning for discovering macroscopic PDEs of viscous gravity currents, as illustrated in Fig. 2. Basically, it consists of the following three main steps:

(1) Data reconstruction or surrogate training. As previously mentioned, in practical scenarios, raw training data can be noisy, or deviate from ground truth, and randomly distributed in the spatio-temporal domain, which will influence derivative/integral calculations and PDE discovery. Therefore, a deep neural network $h^* = NN(t^*, x^*, \theta)$, where θ is the training parameters in the DNN, is first trained by available data to reconstruct the spatio-temporal evolution of front height. Note that, in the DNN, the nondimensional time and location are the only two input variables in the first layer (or input layer), and the nondimensional height is the only output variable in the last layer (or output layer). Theoretically, the neural network can fit any complex mapping relationship. Compared to other reconstruction methods, neural-network based data reconstruction shows good anti-noise and global fitness capability [10]. In addition, the neural network can effectively generate large amounts of meta-data, which is essential for subsequent integral calculation [18].

(2) Semantic fragments construction or derivative/integral calculation. As the neural-network based surrogate is built up, derivatives of various orders are calculated by automatic differentiation if a smooth activation function is adopted for neural network training. Moreover, sufficient meta-data are generated on a regular lattice, and numerical integration is accomplished based on
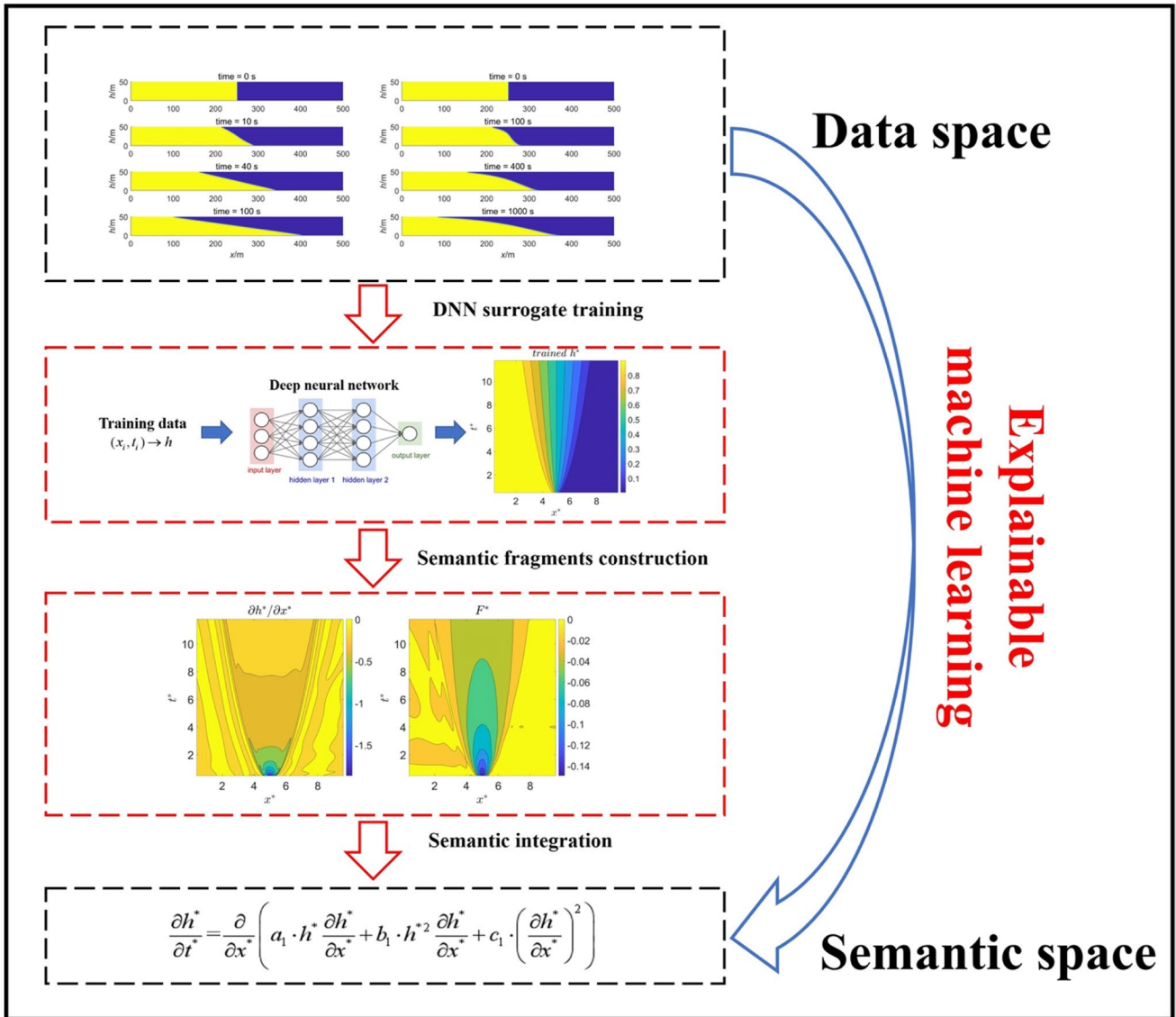
Fig. 2 Flowchart of the deep learning based PDE discovery framework. In order to transform raw simulation data in data space into explainable physical laws (PDEs) in semantic space, the following three steps are necessary. (1) Data reconstruction: training the surrogate based on the deep neural network (DNN); (2) constructing semantic fragments: calculating spatial derivatives and integrated flux; and (3) semantic integration: finding a parsimonious PDE based on sparse regression methods

conventional approaches, such as Gauss–Legendre quadrature. Within this step, flux F^*, original variable h^* and its various spatial derivatives, such as $\partial h^*/\partial x^*$, $\partial^2 h^*/\partial x^{*2}$ and $\partial^3 h^*/\partial x^{*3}$, can be obtained simultaneously, and these variables are considered as semantic fragments ready for the subsequent step.

(3) Semantic integration or sparse regression. Within this step, the possible semantic fragments are discovered and form the eventual explainable governing equation. Generally, regression methods with parsimonious constraints, such as LASSO and genetic algorithm, can be employed to discover the possible PDE terms and corresponding coefficients. Particularly, the genetic algo-

rithm (GA) [7] is adopted in this work. Compared to other sparse regression methods [9, 14, 40] that must specify a large number of potential terms and their combinations in the candidate library, the main advantage of GA is that the scale of the candidate library can be much smaller, which is composed of several basic genes, and numerous combinations can be achieved by automatic generation evolution. For example, a high-order derivative combination of $(h^*)^3(\partial h^*/\partial x^*)^2$ can be recursively generated through mutation and cross-over operations during the GA process with only two basis genes h^* and $\partial h^*/\partial x^*$. As a consequence, the GA-based deep learning framework does not need to include the

true terms of the underlying equation in the initial candidate library, which greatly increases computation efficiency and sparse regression accuracy.

It is worth noting that if only consider the first step "data reconstruction", it is non-explainable because DNN is a black-box surrogate. However, through two subsequential steps, i.e., semantic fragment construction and semantic integration, the eventual results are scientifically explainable because we can obtain a partial differential equation, in which the effect of each term can be rigorously analyzed with mathematical tools, such as advection and diffusion effects. Therefore, from the viewpoint of eventual results, the proposed framework is explainable. Here, it is necessary to clarify the function of DNN in this framework. The first step of building a DNN surrogate is mainly designed as a global smoother for anti-noise of raw training data, and furthermore DNN allows convenient calculation of possible derivative and integral terms for the next step based on its features of automatic differentiation and providing sufficient meta data. These two functions are crucial, and considerably increase the robustness and efficiency for preparation work of PDE discovery compared to other conventional approach such as Gaussian process based smoother. Another feature of this work is the adoption of GA for sparse regression. It allows an incomplete candidate library, and complex terms can be automatically generated through genome mutation and cross-over operations. Moreover, during the optimized process, the parsimonious constraint can be also maintained to a large extent.

Above all, a highly compatible and robust PDE discovery method should be introduced for solving a practical problem without any prior knowledge on the potential equation form. And in this work, the DLGA-PDE framework is adopted based on our previous studies.

3 Results and discussion

3.1 Validation for long-term regimes

First, to validate the proposed deep learning framework for discovering macroscopic PDEs, microscopic simulation data of long-term regimes are utilized for investigation. Training details and parameter settings used in this work during the three steps are listed as follows:

(1) For the data reconstruction step, based on our experiences in previous works for finding various model equations, a five-layer fully-connected neural network with 100 neurons per layer, and soft-plus activation function, are sufficiently accurate for training. The learning rate is set as 0.001 and Adam optimizer is adopted. Besides, the "early stopping" technique is used to avoid over-fitting. For each case, a total of 80% of 500×500 spatial–temporal points are chosen as the training dataset, while the rest 20% points are regarded as the validation set. Along with the DNN training process, the trend of the mean square error (MSE) estimated from the training set keeps decreasing, while the trend from the validation set shall first decrease and then increase. Such a turning point is considered as a criterion to diagnose over-fitting. After training approximately 100,000 epochs, the mean square error (MSE) between raw data and prediction results decreases to a minimal level of 10^{-7}, which is considered to be sufficiently converged.

(2) For derivative/integration calculation, 300×300 lattice points of the inner spatial–temporal domain are selected, and a five-point Gauss–Legendre quadrature scheme is adopted here to calculate flux. In order to improve the regression performance, a large amount of trivial samples out of the mentioned 90,000 lattice points, e.g., $h^* \approx 1$ and $h^* \approx 0$, need to be omitted because flux is also trivial in these regions, i.e., $F^* \approx 0$.

(3) For the sparse regression step, four basic genes are chosen as the candidate library, i.e., $\left(h^*, \partial h^*/\partial x^*, \partial^2 h^*/\partial x^{*2}, \partial^3 h^*/\partial x^{*3}\right)$. Population size is set as 200 per generation, and genes are randomly combined with each other as the initial generation. Mutation rate and cross-over rate are set as 20% and 80%, respectively. The maximum evolving epoch number is set as 100. Fitness function is defined as Fitness $= -\sum \left|F^*_{ref} - F^*_{ga}\right|/N - \lambda \sum a_i/|a_i|$, where N is the number of sample data; and a_i is the coefficient of the i-th derivative combination. The first term indicates data fitness between reference flux value F^*_{ref} and estimated value F^*_{ga}, and the last term indicates the parsimony constraint, i.e., the total number of terms, which is controlled by hyper-parameter λ. In this work, λ is set as 0.002 for most cases.

It is worth noting that there exist several hyper-parameters in the deep learning framework, such as number of hidden layers, choosing strategy of candidate library, and parsimony constraint coefficient. Most of these parameters are determined by experience via an ad hoc procedure. For instance, the number of neuron layers in the neural networks during the data reconstruction step is determined when further adding the number of layers does not significantly affect the eventual data loss. Sensitivity studies with different high-order derivatives are carried out in constituting the candidate

library, and it is found that there is no necessity to intro-
duce derivatives over third-order in the candidate library.
The parsimony constraint coefficient is so determined that
the regression quality is achieved while not sacrificing the
parsimony (with as small a number of PDE terms as pos-
sible). Above all, the primary objective of training the deep
learning framework is to discover an accurate yet parsimoni-
ous PDE, so that the prediction based on the PDE through
posterior tests can well match the raw training data and the
learned PDE can reveal the physical insight.

Eventually, for describing long-term behaviors, the
obtained deep learning based PDEs (DL-PDEs) of the two
cases are expressed as follows:

$$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(0.872 \cdot h^* \frac{\partial h^*}{\partial x^*} - 0.881 \cdot h^{*2}\frac{\partial h^*}{\partial x^*}\right) \quad (18)$$

$$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(0.988 \cdot h^* \frac{\partial h^*}{\partial x^*} + 0.0325 \cdot h^{*2}\right) \quad (19)$$

To provide insight into how GA works to find
the optimized solution, Table 1 lists the GA solu-
tion paths for the two cases. For case I, the discovered
best child for flux in the first generation is found to be
a set of $\{h^*, h^{*2}, \partial^2 h^*/\partial x^{*2}\}$, which contains second-
order derivative $\partial^2 h^*/\partial x^{*2}$. However, after only three
generations, the optimized genomes are converged
to $\{h^* \partial h^*/\partial x^*, h^{*2}\partial h^*/\partial x^*\}$. Quantitatively, the fit-
ness loss eventually decreases to 0.0048 from 0.0077
of the first generation. From the solution path, it is
clearly seen that the initial candidate library, i.e.,
$\{h^*, \partial h^*/\partial x^*, \partial^2 h^*/\partial x^{*2}, \partial^3 h^*/\partial x^{*3}\}$, does not need to
be complete to include eventual terms in this approach.
Indeed, to construct a complete candidate library would
usually mean to include an overwhelmingly large number
of possible terms, which may greatly increase the com-
putational cost and diminish the performance of sparse
regression. It is clear that, to describe the long-term
behaviors of viscous gravity currents, the form of flux is
highly concise. Since in each generation, a total of 200
children are generated for evolution, the optimized solu-
tion can be determined without much difficulty.

Table 1 GA solution path of long-term regimes for two cases

Generations for case I	Structure of discov-ered best child for case I	Generations for case II	Structure of dis-covered best child for case II
1	$h^*, h^{*2}, \frac{\partial^2 h^*}{\partial x^{*2}}$	1	$h^*, h^* \frac{\partial h^*}{\partial x^*}$
2	h^*, h^{*2}	4	$h^{*3}, h^* \frac{\partial h^*}{\partial x^*}$
4	$h^* \frac{\partial h^*}{\partial x^*}, h^{*2} \frac{\partial h^*}{\partial x^*}$	5	$h^{*2}, h^* \frac{\partial h^*}{\partial x^*}$
100	$h^* \frac{\partial h^*}{\partial x^*}, h^{*2} \frac{\partial h^*}{\partial x^*}$	100	$h^{*2}, h^* \frac{\partial h^*}{\partial x^*}$

It can be seen that the trained long-term DL-PDEs (18)
and (19) are highly similar to the theoretically derived
Eq. (14) and (15), while coefficients are slightly dissimi-
lar for both cases, and an additional term appears for case
II. Particularly for case I, the corresponding coefficients of
dominant terms $h^* \partial h^*/\partial x^*$ and $h^{*2}\partial h^*/\partial x^*$ deviate approxi-
mately 12% from the theoretical PDE. For case II, the coef-
ficients of dominant term $h^* \partial h^*/\partial x^*$ deviate approximately
only 2% from the theoretical PDE, and the coefficient of the
compensation term is 0.0325, which is also much smaller
compared to the dominant term.

Because the long-term DL-PDEs are directly learned
from the microscopic simulation data, it is difficult to dis-
cern the error sources resulting in the deviations between
DL-PDEs and theoretical-PDEs. The deviations may be
attributed to the training error of deep learning or the inva-
lidity of assumptions for deriving theoretical PDEs. To
clarify this problem, it is necessary to carry out further
quantitative comparisons among microscopic simulation
data, DL-PDEs, and theoretical PDEs. However, Eq. (18)
and (19) are expressed as PDEs in scientific semantic space,
while simulation results are expressed in data space. Since it
is impossible to directly compare two different spaces, one
must solve Eq. (18) and (19) numerically, and compare the
posterior results with microscopic simulation results in data
space. Particularly, to obtain a numerical solution for deep
learning based PDEs, the center difference scheme is first
applied for calculating the flux in the cell center, and then
the WENO scheme is utilized for flux reconstruction similar
to solving Eq. (4).

The posterior results are illustrated in Fig. 3(a) and
(b). It is clear that the characteristic contour lines of the
three data sets are close to each other. Quantitatively, for
case I, the total relative errors are estimated as 0.45% and
0.75% for DL-PDE and theoretical PDE, respectively.
For case II, the errors are 0.47% and 1.1%, respectively.
Therefore, the prediction results of DL-PDEs achieve a
superior match with the microscopic simulation results
compared to those of theoretical PDEs, irrespective of the
training error. This finding implies that modifications are
necessary, including smaller coefficients and secondary
terms, to better describe viscous gravity currents, even
for long-term regimes. More importantly, it is proven that
the proposed deep learning framework is highly capable
and accurate to discover macroscopic equations from raw
microscopic simulation data.

3.2 PDE discovery for short-term regimes

Although the theoretical equations perform well for predict-
ing the long-term asymptotic behaviors, there still exist no
parsimonious theoretical PDEs for describing the early-time
or short-term pre-asymptotic regimes, which are important

Fig. 3 Contour lines of estimated evolving current front height h^* in posterior tests for two cases: (**a**) long-term behaviors of case I; (**b**) long-term behaviors of case II; (**c**) short-term behaviors of case I; (**d**) short-term behaviors of case II. The black solid lines are reference values extracted from two-dimensional microscopic simulation results, the long red dashed lines are posterior results of deep learning based PDEs, and the short blue dashed lines are posterior results of theoretical PDEs

(a) h^* of case I

(c) h^* of case I

(b) h^* of case II

(d) h^* of case II
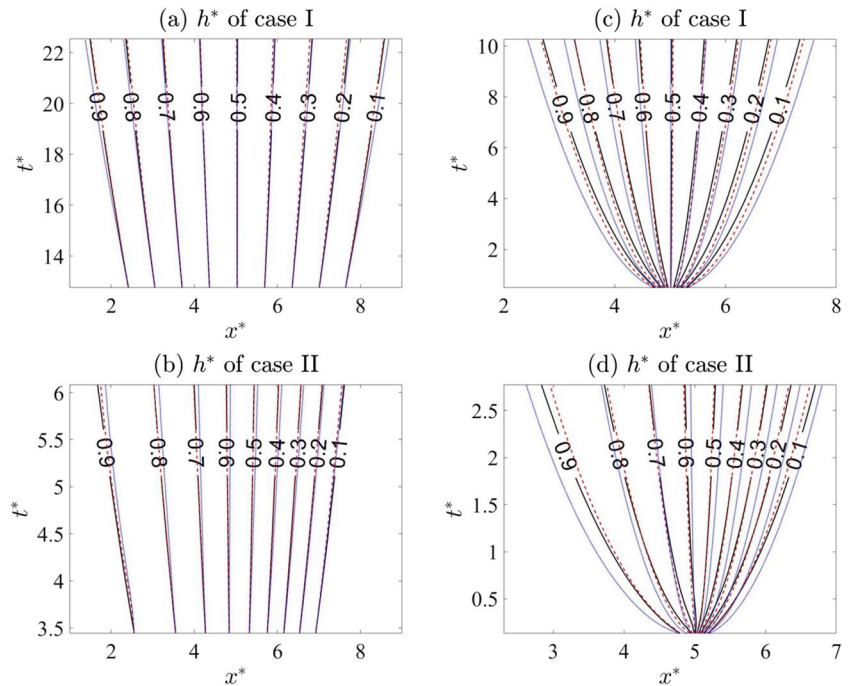
for rapid processes such as proppant transport. Therefore, in this part, we aim to utilize the deep learning framework to discover the macroscopic PDEs for describing short-term behaviors, and elucidate the hidden macroscopic mechanisms to identify how deviation between theoretical assumptions and actual situations influences the macroscopic equations. Here, the training parameters of deep learning are similar to those for long-term DL-PDEs. The eventual learned PDEs for cases I and II are given as follows:

$$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(a_1 \cdot h^* \frac{\partial h^*}{\partial x^*} + b_1 \cdot h^{*2} \frac{\partial h^*}{\partial x^*} + c_1 \cdot \left(\frac{\partial h^*}{\partial x^*}\right)^2 \right) \quad (20)$$

$$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(a_2 \cdot h^* \frac{\partial h^*}{\partial x^*} + b_2 \cdot h^* \left(\frac{\partial h^*}{\partial x^*}\right)^2 + c_2 \cdot h^{*4} \frac{\partial h^*}{\partial x^*} \right) \quad (21)$$

where $a_1 = 0.894$; $b_1 = -0.881$; $c_1 = 0.0757$; and $a_2 = 1.01$; $b_2 = 0.215$; $c_2 = -0.571$.

Compared to the DL-PDEs of long-term regimes, it is evident that flux expressions of Eq. (20) and (21) are more complex, and the GA solution path indicates that many more generations (57 for case I, and 77 for case II) are necessary for convergence, as demonstrated in Table 2. Indeed, at the early stage of viscous gravity currents, the fluid field has not yet been fully developed, and theoretical assumptions are too ideal to reflect the actual scenario, suggesting that a more complex high-dimensional nonlinear mapping relationship is required for describing short-term spatio-temporal evolution of current front height.

By comparing the above deep learning trained PDEs (DL-PDE) with theoretically derived Eq. (14) and (15), the following findings are determined:

Table 2 GA solution paths of short-term regimes for two cases

Generations for case I	Structure of discovered best child for case I	Generations for case II	Structure of discovered best child for case II
1	$\frac{\partial h^*}{\partial x^*}$	1	$h^*, h^* \frac{\partial^2 h^*}{\partial x^{*2}} \frac{\partial^3 h^*}{\partial x^{*3}}, h^{*2}, h^* \frac{\partial h^*}{\partial x^*}, \left(\frac{\partial h^*}{\partial x^*}\right)^2$
2	$h^*, \frac{\partial h^*}{\partial x^*}, h^{*2}$	3	$\frac{\partial^3 h^*}{\partial x^{*3}}, h^*, h^* \frac{\partial h^*}{\partial x^*}, h^{*3}$
3	$h^{*2}\left(\frac{\partial h^*}{\partial x^*}\right)^3, h^* \frac{\partial h^*}{\partial x^*}, h^{*2} \frac{\partial h^*}{\partial x^*}$	4	$h^*, h^* \frac{\partial h^*}{\partial x^*}, h^{*2}$
20	$h^* \frac{\partial h^*}{\partial x^*}, h^{*2} \frac{\partial h^*}{\partial x^*}$	6	$h^*, h^* \frac{\partial h^*}{\partial x^*}, h^{*2}, h^{*2}\left(\frac{\partial h^*}{\partial x^*}\right)^2$
57	$h^* \frac{\partial h^*}{\partial x^*}, h^{*2} \frac{\partial h^*}{\partial x^*}, \left(\frac{\partial h^*}{\partial x^*}\right)^2$	18	$h^{*3} \frac{\partial h^*}{\partial x^*}, h^* \frac{\partial h^*}{\partial x^*}, h^*\left(\frac{\partial h^*}{\partial x^*}\right)^2$
100	$h^* \frac{\partial h^*}{\partial x^*}, h^{*2} \frac{\partial h^*}{\partial x^*}, \left(\frac{\partial h^*}{\partial x^*}\right)^2$	77 ~ 100	$h^* \frac{\partial h^*}{\partial x^*}, h^*\left(\frac{\partial h^*}{\partial x^*}\right)^2, h^{*4} \frac{\partial h^*}{\partial x^*}$

(1) For both cases, it is clear from Eq. (20) and (21) that high-order derivatives, i.e., $\partial^2 h^*/\partial x^{*2}$ and $\partial^3 h^*/\partial x^{*3}$, make no contribution to the flux construction. It seems that flux can be considered as a bivariate function of h and $\partial h^*/\partial x^*$. As mentioned previously, high-order derivatives temporally appear in the GA solution path. However, through fitness selection, these terms are automatically eliminated from the optimized genomes. This fact is also verified from DL-PDEs (18) and (19) of the long-term regime, as well as theoretical PDEs (14) and (15). In fact, in viscous gravity current processes, these two terms have strong physical meanings according to theoretical analysis, as illustrated in Section 2.2. Specifically, h represents vertical length allowing flux transfer, and $\partial h^*/\partial x^*$ is proportional to average horizontal velocity.

(2) Flux terms in theoretical PDEs are exactly captured in DL-PDEs, regardless of slightly different coefficients. Particularly for case I, coefficients of $h^* \partial h^*/\partial x^*$ and $h^{*2} \partial h^*/\partial x^*$ are 0.894 and -0.881, while being 1 and -1 in theoretical PDEs, respectively. Obviously, the characteristic of mutual additive inverse for these two coefficients is not an accidental consequence, but instead reflects the symmetry characteristic of the solution, i.e., $F(h^*, \partial h^*/\partial x^*) = F(1 - h^*, \partial h^*/\partial x^*)$. It is worth noting that this relationship is not satisfied in case II. For case II, the coefficient of leading term $h^* \partial h^*/\partial x^*$ is 1.01, which is very close to the theoretical value of 1.

(3) Excluding the different coefficients of leading terms, it is observed that DL-PDEs provide additional terms for two cases, i.e., $0.0757 \cdot (\partial h^*/\partial x^*)^2$ for case I and $0.215 \cdot h^*(\partial h^*/\partial x^*)^2 - 0.571 \cdot h^{*4} \partial h^*/\partial x^*$ for case II. It should be emphasized that theoretical PDEs are derived for describing long-term behaviors of viscous gravity currents; whereas, in this work, we extend the training data to the whole range, including short-term behaviors. Therefore, we aim to find a global solution for viscous gravity currents, which implies that the additional terms in DL-PDEs and modified coefficients of leading terms may be considered as compensation terms for describing short-term range behaviors.

It is worth noting that it is not guaranteed that the deep learning based PDEs are exactly numerically stable because of training error, and the negative diffusion term in Eq. (20) and (21) at the early time can dominate the process at the early time, where $-\partial h^*/\partial x^*$ is very large and leads to computational divergence, which is a consequence of the training error tolerance. To solve this issue, necessary artificial viscosity is introduced to counter the negative diffusion under these circumstances. To illustrate this, let us take Eq. (20) as an example. Equation (20) can be rewritten as follows:

$$\frac{\partial h^*}{\partial t^*} = (a_1 + 2b_1 h^*)\frac{\partial h^*}{\partial x^*}\frac{\partial h^*}{\partial x^*} + \left(a_1 h^* + b_1 h^{*2} + 2c_1 \frac{\partial h^*}{\partial x^*}\right)\frac{\partial^2 h^*}{\partial x^{*2}} \tag{22}$$

The two terms on the right-hand side can be regarded as parametric convection and diffusion terms, respectively. In this case, if the coefficient of the diffusion term is positive, i.e., $a_1 h^* + b_1 h^{*2} + 2c_1 \partial h^*/\partial x^* > 0$, then the system is numerically stable. If the coefficient is negative, however, the system will break down due to anti-diffusion. According to our numerical experiments, it is found that the system is more likely to collapse where $-\partial h^*/\partial x^*$ is very large, which corresponds to the early time in the viscous gravity current.

To suppress the non-physical anti-diffusion terms, artificial viscosity η is introduced for deep learning based PDEs when needed, which is written as:

$$\frac{\partial h^*}{\partial t^*} - \frac{\partial}{\partial x^*}\left(a_1 \cdot h^*\frac{\partial h^*}{\partial x^*} + b_1 \cdot h^{*2}\frac{\partial h^*}{\partial x^*} + c_1 \cdot \left(\frac{\partial h^*}{\partial x^*}\right)^2\right) = \eta\frac{\partial^2 h^*}{\partial x^{*2}} \tag{23}$$

where $\eta = -min(0, a_1 h^* + b_1 h^{*2} + 2c_1 \partial h^*/\partial x^*)$. Similar remedies are adopted for other DL-PDEs when needed during posterior tests, such as in Eq. (21).

For quantitative comparisons, posterior tests are also carried out, and the evolving front height of two-dimensional simulation results, DL-PDEs, and theoretical PDEs are illustrated in Fig. 3(c) and (d). Clearly, when describing the short-term or early-time pre-asymptotic behaviors, there exist remarkable deviation between the prediction of theoretical derived PDEs and reference data. From the characteristic lines of current height in the $t - x$ domain, it is inferred that wave structures of current height in the short-term can be considered as rarefaction waves with time-dependent characteristic velocities. The comparison results demonstrate that prediction results of DL-PDEs match the microscopic simulation results better than those of theoretical PDEs. Quantitatively, for case I, the total relative errors are estimated as 1.06% and 2.75% for DL-PDE and theoretical PDE, respectively. For case II, however, the errors are 0.85% and 2.72%, respectively.

As previously mentioned, deviation between reference data and theoretical results is mainly attributed to the assumptions when deriving theoretical PDEs. In theoretical analysis, it is assumed that vertical velocity is zero for the whole domain, while microscopic simulation results infer that this condition is not fulfilled, particularly near the interface and during the short-term period. It is seen from Fig. 4 that, over the whole simulation period, the absolute relative magnitude of vertical velocity ranges markedly from 0.5 to 1. Moreover, the vertical velocity is fairly comparable to the horizontal velocity, which is particularly significant near two interface tails. Therefore, the assumption of zero vertical velocity is not valid, and theoretical PDEs cannot precisely reproduce the evolving process of viscous gravity

current. As a consequence, the transverse spreading length in the $t - x$ domain predicted by theoretical PDEs is larger than ground truth. Modified coefficients and compensation terms are necessary to be introduced for improvement. On the other hand, because DL-PDEs are discovered directly from observation data, once training is converged, the posterior performance is reasonably better than that of theoretical PDEs.

From the above discussions, the following key points can be inferred for the mechanisms hidden in DL-PDEs:

(1) Smaller coefficients in leading terms for case I weaken the strength of the rarefaction wave and adjust theoretical PDEs towards ground truth.
(2) Compensation terms for cases I and II essentially represent negative-diffusion mechanisms, which also suppress the over-spreading trend for theoretical PDEs. For cases I and II, the compensation diffusion terms are $0.151 \cdot \partial h^*/\partial x^* \left(\partial^2 h^*/\partial x^{*2}\right)$ and $0.43 \cdot h^* \partial h^*/\partial x^* \left(\partial^2 h^*/\partial x^{*2}\right) - 0.571 \cdot h^{*4} \partial^2 h^*/\partial x^{*2}$, respectively. Furthermore, since in this work we have $0 \le h^* \le 1$ and $\partial h^*/\partial x^* \le 0$, the coefficients of the compensation diffusion terms are negative for both cases.

3.3 Effects of data noise

Usually, data noise is a nonnegligible factor when collecting raw training data, especially in physical experiments. Excessive data noise can dramatically affect the validity of the discovered PDEs. To elucidate the robustness of the proposed deep learning framework, in this part raw simulation data with appended random noise are utilized to discover the PDEs for short-term regimes. Here, case II is taken as an example, and Fig. 5(a) and (c) illustrate the 3D surfaces of raw training data with 10% and 20% non-biased noise. Although a large number of burrs due to data noise are observed, the basic evolving trend of dimensionless current front height can be still maintained in the raw data to some

extent. Figure 5(b) and (d) illustrate the DNN surrogate constructed from simulation data with 10% and 20% non-biased noise, respectively. In this part, the training parameters of DNN are all the same as those mentioned in Section 3.1. Due to the data noise, the training loss finally keeps a level of 10^{-4} after training 100,000 epochs. Clearly, burrs due to data noise no longer exist in the surrogate, which proves that DNN is actually a powerful global smoother to construct a smooth-enough surrogate for subsequential derivative/integral calculations.

Table 3 lists the discovered PDEs based on the deep learning framework for these cases. Essentially, as the magnitude of appended noise increases, the fitness loss in the genetic algorithm monotonically increases. Even so, for case I with 10% data noise, the form of discovered PDE is exactly the same as that discovered from clean simulation data. The relative error of the coefficients of leading terms for the clean case and the 10% noise case is smaller than 2%, and the relative error of compensation terms is approximately 15%. On the other hand, for case I with 20% noise, it seems that the proposed deep learning framework fails to find the aforementioned compensation terms. Instead, the coefficients of leading terms are much smaller than those of the clean data case. As mentioned previously, smaller coefficients and compensation terms both contributed to weaker rarefaction mechanisms. Therefore, even with 20% data noise, the discovered PDE also demonstrates the necessity of leading terms and modifications in the effective direction.

For case II, it is clear that despite data noise, all of the discovered PDEs possess consistent forms. The coefficient of the leading term is always approximately 1, and the relative errors of compensation terms are less than 20%. It is also clear that, although the training loss in this part is much higher than the results with no data noise, the basic structure of the obtained DL-PDEs can still be maintained. Above all, through the numerical tests in this section, it is proven that the proposed DLGA framework is quite robust to discover the explainable PDEs even

Fig. 4 Contour of relative magnitude of vertical velocity component $u_z/\sqrt{u_x^2 + u_z^2}$ at various time instances for two cases: (**a**) case I; (**b**) case II
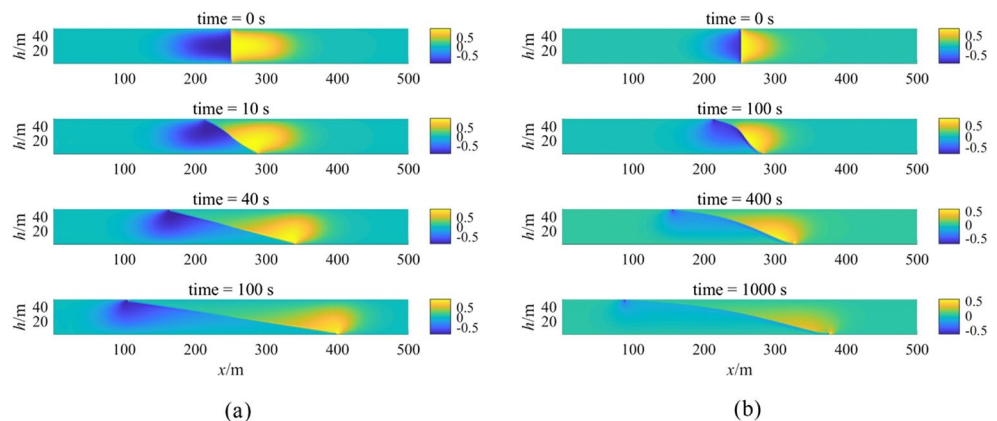
(a)

(b)

Fig. 5 3D illustrations of raw training data ((**a**) and (**c**)) and DNN surrogate ((**b**) and (**d**)). (**a**) and (**b**): 10% data noise; (**c**) and (**d**): 20% data noise
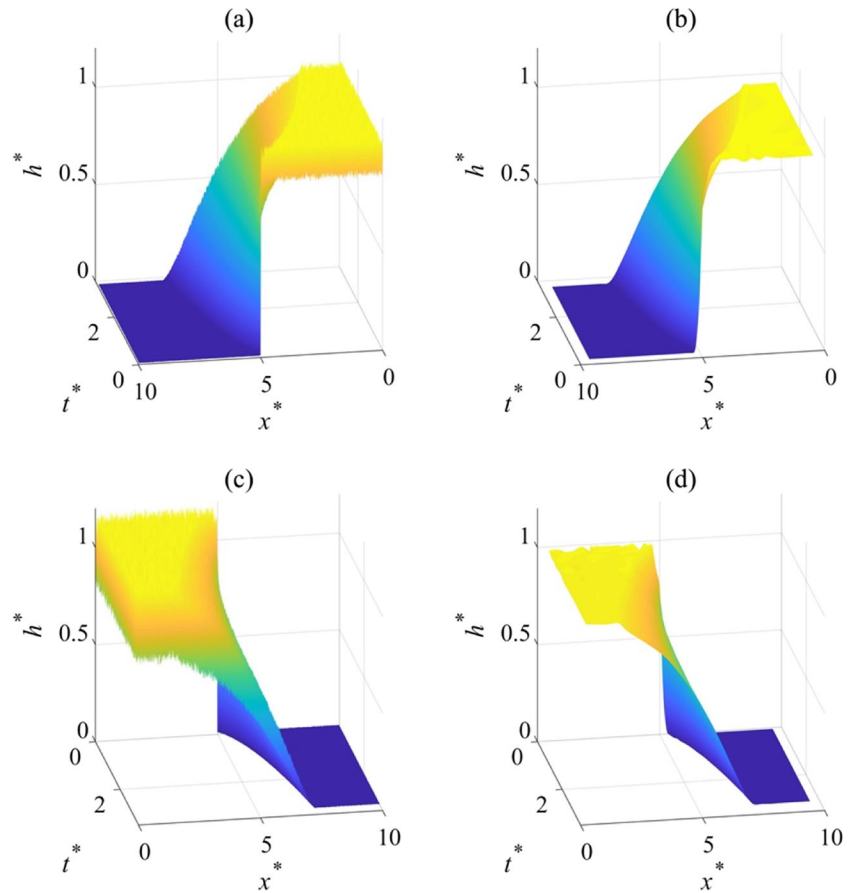

Table 3 DL-PDEs for viscous gravity current (short-term regimes) at different noise levels

Noise level	Case I	Fitness loss	Case II	Fitness loss
0%	$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(0.894h^*\frac{\partial h^*}{\partial x^*} - 0.881h^{*2}\frac{\partial h^*}{\partial x^*} + 0.0757\left(\frac{\partial h^*}{\partial x^*}\right)^2\right)$	0.00716	$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(1.006h^*\frac{\partial h^*}{\partial x^*} + 0.215h^*\left(\frac{\partial h^*}{\partial x^*}\right)^2 - 0.571h^{*4}\frac{\partial h^*}{\partial x^*}\right)$	0.0133
10%	$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(0.912h^*\frac{\partial h^*}{\partial x^*} - 0.900h^{*2}\frac{\partial h^*}{\partial x^*} + 0.0876\left(\frac{\partial h^*}{\partial x^*}\right)^2\right)$	0.00883	$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(1.031h^*\frac{\partial h^*}{\partial x^*} + 0.245h^*\left(\frac{\partial h^*}{\partial x^*}\right)^2 - 0.600h^{*4}\frac{\partial h^*}{\partial x^*}\right)$	0.0162
20%	$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(0.737h^*\frac{\partial h^*}{\partial x^*} - 0.741h^{*2}\frac{\partial h^*}{\partial x^*}\right)$	0.0159	$\frac{\partial h^*}{\partial t^*} = \frac{\partial}{\partial x^*}\left(1.046h^*\frac{\partial h^*}{\partial x^*} + 0.233h^*\left(\frac{\partial h^*}{\partial x^*}\right)^2 - 0.600h^{*4}\frac{\partial h^*}{\partial x^*}\right)$	0.0550

within 20% data noise for the viscous gravity current processes. From the authors' point of view, the robustness of the proposed framework is to a great extent attributable to the strong data fitting and smoothing capability of the neural network. It should be noted that although the data noise is assumed to be non-biased in this part, advanced approaches, such as the 'robust DLGA' framework [41], can be utilized to solve more complex problems that contain biased data noise.

4 Conclusion

In summary, we proposed a deep learning framework to elucidate the macroscopic equations of viscous gravity currents based on microscopic simulation data. The problem of discovering macroscopic PDEs is converted to a sparse regression problem of representing flux with primitive variables and corresponding spatial derivatives. With the assistance of the deep neural network and

genetic algorithm, PDEs with parsimonious forms for both long-term and short-term behaviors for two typical viscous gravity current processes are obtained. By quantitative comparison between theoretical PDEs and DL-PDEs for describing long-term behaviors, it is proven that the proposed framework is highly capable and accurate for discovering physical laws from raw data. The short-term simulation data are then utilized for discovering unknown macroscopic equations.

We show that the proposed deep learning based PDE discovery framework is essentially an explainable machine-learning approach. Microscopic simulation results in data space are eventually transformed into parsimonious PDEs in scientific semantic space through the PDE discovery method. With this framework, explainable compensation terms can be introduced to capture the short-term behaviors more accurately. It is found that, in posterior tests, DL-PDEs perform better than the theoretically derived equations. Moreover, the framework is proven to be very robust against approximately 20% non-biased data noise. Consequently, the proposed method is demonstrated to be highly beneficial to compensate for the drawbacks of ideal assumptions during theoretical derivation and construct more accurate governing equations, which fits observation data better for practical physical problems.

Overall, for a certain physical process, obtained experimental or simulation data contain all kinds of information in data space. However, one needs to extract the most valuable and significant content, and ideally determine intrinsic laws in scientific semantic space. This work demonstrates that the proposed deep learning based PDE discovery framework constitutes a feasible and efficient approach to solve this problem. It is also worth noting that the proposed framework follows a pure data-driven approach, and the hyper-parameters are still manually tuned to achieve parsimony principles, which still need further improvements to consider more prior physics and avoid subjective decisions. Physics-informed structures and Lagrangian dual optimization approaches may be good solutions for these problems. In future research, we will continue to consider these issues and extend the framework for learning multi-dimensional equations and for more challenging scenarios.

Data Availability The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code availability The codes used to generate and analyze data are available through Zenodo (https://doi.org/10.5281/zenodo.4587614).

Declarations

Conflict of interest The authors declare that they have no conflicts of interest to report regarding the publication of this paper.

References

1. Baldi, P., Sadowski, P., Whiteson, D.: Searching for exotic particles in high-energy physics with deep learning. Nat. Commun. **5**, 1–9 (2014). https://doi.org/10.1038/ncomms5308
2. Raissi, M., Yazdani, A., Karniadakis, G.E.: Hidden Fluid Mechanics: Learning velocity and pressure fields from flow visualizations. Science **4741**, 1–8 (2020). https://doi.org/10.1126/science.aaw4741
3. Kadow, C., Hall, D.M., Ulbrich, U.: Artificial intelligence reconstructs missing climate information. Nat. Geosci. **13**, 408–413 (2020). https://doi.org/10.1038/s41561-020-0582-5
4. Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N.: Prabhat: Deep learning and process understanding for data-driven Earth system science. Nature **566**, 195–204 (2019). https://doi.org/10.1038/s41586-019-0912-1
5. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015). https://doi.org/10.1038/nature14539
6. Roscher, R., Bohn, B., Duarte, M.F., Garcke, J.: Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access **8**, 42200–42216 (2020). https://doi.org/10.1109/ACCESS.2020.2976199
7. Xu, H., Chang, H., Zhang, D.: DLGA-PDE: Discovery of PDEs with incomplete candidate library via combination of deep learning and genetic algorithm. J. Comput. Phys. **418**, 109584 (2020). https://doi.org/10.1016/j.jcp.2020.109584
8. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. **378**, 686–707 (2019). https://doi.org/10.1016/j.jcp.2018.10.045
9. Berg, J., Nyström, K.: Data-driven discovery of PDEs in complex datasets. J. Comput. Phys. **384**, 239–252 (2019). https://doi.org/10.1016/j.jcp.2019.01.036
10. Xu, H., Chang, H., Zhang, D.: DL-PDE: Deep learning based data-driven discovery of partial differential equations from discrete and noisy data. Phys. Fluids **33**, 037132 (2021)
11. Brunton, S.L., Kutz, J.N.: Methods for data-driven multiscale model discovery for materials. J. Phys.: Mater. **2**, 044002 (2019). https://doi.org/10.1088/2515-7639/ab291e
12. Kaiser, E., Nathan Kutz, J., Brunton, S.L.: Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc. R. Soc. A: Math. Phys. Eng. Sci. **474**, 2219 (2018)
13. Quade, M., Abel, M., Nathan Kutz, J., Brunton, S.L.: Sparse identification of nonlinear dynamics for rapid model recovery. Chaos **28**, 063116 (2018). https://doi.org/10.1063/1.5027470
14. Rudy, S.H., Brunton, S.L., Proctor, J.L.: Kutz: Data-driven discovery of partial differential equations. Sci. Adv. **3**, e1602614 (2017)

15. Rudy, S., Alla, A., Brunton, S.L., Kutz, J.N.: Data-driven identification of parametric partial differential equations. SIAM J. Appl. Dyn. Syst. **18**, 643–660 (2019). https://doi.org/10.1137/18M1191944

16. Zhang, S., Lin, G.: Robust data-driven discovery of governing physical laws with error bars. Proc. R. S. A: Math. Phys. Eng. Sci. **474**, 20180305 (2018). https://doi.org/10.1098/rspa.2018.0305

17. Maslyaev, M., Hvatov, A., Kalyuzhnaya, A.: Data-driven PDE discovery with evolutionary approach (2019). Preprint at https://arxiv.org/abs/1903.08011

18. Xu, H., Zhang, D., Wang, N.: Deep learning based discovery of partial differential equations in integral form from sparse and noisy data. J. Comput. Phys. **445**, 10592 (2020)

19. Gardner, G.H.F., Downie, J., Kendall, H.A.: Gravity segregation of miscible fluids in linear models. SPE J. **185**, 95–104 (1964)

20. Davis, S.H., Hocking, L.M.: Spreading and imbibition of viscous liquid on a porous base. II. Physics of Fluids **12**, 1646–1655 (2000). https://doi.org/10.1063/1.870416

21. Pritchard, D., Woods, A.W., Hogg, A.J.: On the slow draining of a gravity current moving through a layered permeable medium. J. Fluid Mech. **444**, 23–47 (2001). https://doi.org/10.1017/S002211200100516X

22. Lake, L.: Enhanced oil recovery. Prentice Hall, Hoboken (1988)

23. Werner, A.D.: On the classification of seawater intrusion. J. Hydrol. **551**, 619–631 (2017). https://doi.org/10.1016/j.jhydrol.2016.12.012

24. Werner, A.D., Bakker, M., Post, V.E.A., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., Simmons, C.T., Barry, D.A.: Seawater intrusion processes, investigation and management: Recent advances and future challenges. Adv. Water Resour. **51**, 3–26 (2013). https://doi.org/10.1016/j.advwatres.2012.03.004

25. Woods, A.W.: Liquid and vapor flow in superheated rock. Annu. Rev. Fluid Mech. **31**, 171–199 (1999). https://doi.org/10.1146/annurev.fluid.31.1.171

26. Bear, J.: Dynamics of fluids in Porous media. Dover Publications, New York (1972)

27. Pritchard, D., Hogg, A.J.: Draining viscous gravity currents in a vertical fracture. J. Fluid Mech. **459**, 207–216 (2002). https://doi.org/10.1017/S0022112002008327

28. Woods, A.W., Mason, R.: The dynamics of two-layer gravity-driven flows in permeable rock. J. Fluid Mech. **421**, 83–114 (2000). https://doi.org/10.1017/S0022112000001567

29. Di Federico, V., Longo, S., King, S.E., Chiapponi, L., Petrolo, D., Ciriello, V.: Gravity-driven flow of Herschel-Bulkley fluid in a fracture and in a 2D porous medium. J. Fluid Mech. **821**, 59–84 (2017). https://doi.org/10.1017/jfm.2017.234

30. Chiapponi, L., Petrolo, D., Lenci, A., di Federico, V., Longo, S.: Dispersion induced by non-Newtonian gravity flow in a layered fracture or formation. J. Fluid Mech. **903**, A14 (2020). https://doi.org/10.1017/jfm.2020.624

31. Longo, S., di Federico, V., Chiapponi, L.: A dipole solution for power-law gravity currents in porous formations. J. Fluid Mech. **778**, 534–551 (2015). https://doi.org/10.1017/jfm.2015.405

32. Zheng, Z., Christov, I.C., Stone, H.A.: Influence of heterogeneity on second-kind self-similar solutions for viscous gravity currents. J. Fluid Mech. **747**, 218–246 (2014). https://doi.org/10.1017/jfm.2014.148

33. Anderson, J.D.: Computational fluid dynamics: the basics with applications. McGraw-Hill, New York (1995)

34. Olsson, E., Kreiss, G.: A conservative level set method for two phase flow. J. Comput. Phys. **210**, 225–246 (2005). https://doi.org/10.1016/j.jcp.2005.04.007

35. Notay, Y.: An aggregation-based algebraic multigrid method. Electron. Trans. Numer. Anal. **37**, 123–146 (2010)

36. Gottlieb, S., Shu, C.-W.: Total variation diminishing Runge-Kutta schemes. Math. Comput. Am. Math. Soc. **67**, 73–85 (1998). https://doi.org/10.1090/s0025-5718-98-00913-2

37. Roostaei, M., Nouri, A., Fattahpour, V., Chan, D.: Numerical simulation of proppant transport in hydraulic fractures. J. Petrol. Sci. Eng. **163**, 119–138 (2018). https://doi.org/10.1016/j.petrol.2017.11.044

38. Jiang, G.S., Shu, C.W.: Efficient implementation of weighted ENO schemes. J. Comput. Phys. **126**, 202–228 (1996)

39. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Machine learning of linear differential equations using Gaussian processes. J. Comput. Phys. **348**, 683–693 (2017). https://doi.org/10.1016/j.jcp.2017.07.050

40. Chang, H., Zhang, D.: Machine learning subsurface flow equations from data. Comput. Geosci. **23**, 895–910 (2019). https://doi.org/10.1007/s10596-019-09847-2

41. Xu, H., Zhang, D.: Robust discovery of partial differential equations in complex situations. Phys. Rev. Res. **3**, 033270 (2021). https://doi.org/10.1103/PhysRevResearch.3.033270