SCHMIDT FUTURES

DEPARTMENT OF ENGINEERING SCIENCE

UNIVERSITY OF OXFORD

What makes a wave break?

How machine learning can shed light on the underlying physics of breaking waves

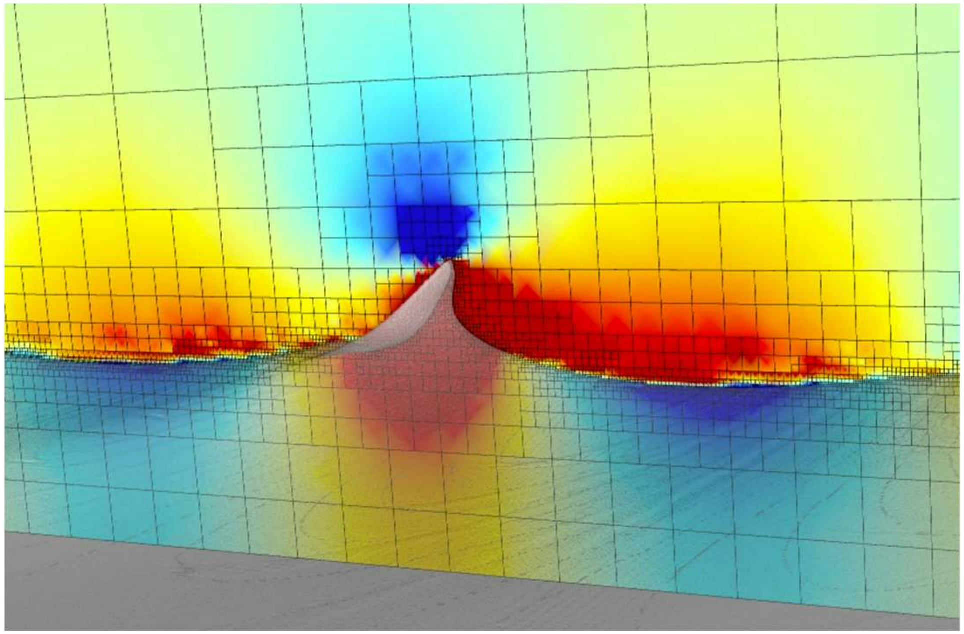
Tianning Tang (Tim);
Schmidt AI in Science Fellow @ University of Oxford
Lecturer @ University of Manchester

30/06/2025

Background

However, modelling these breaking waves requires DNS solving the **Navier-Stokes equation** and are very **computational heavy**.

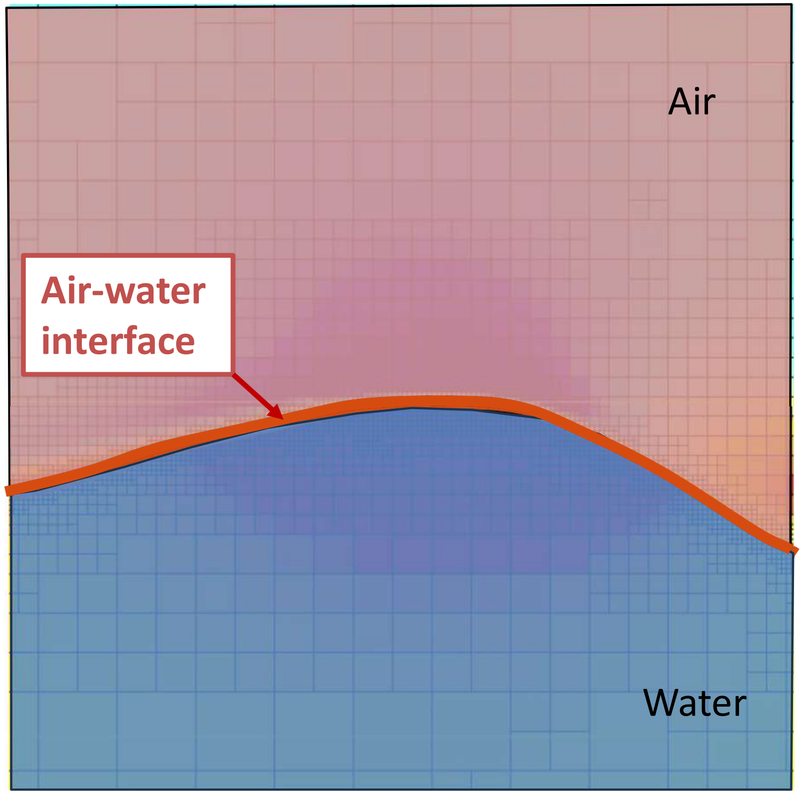
2D breaking wave with Navier Stokes Equations – **3 days on Cluster**



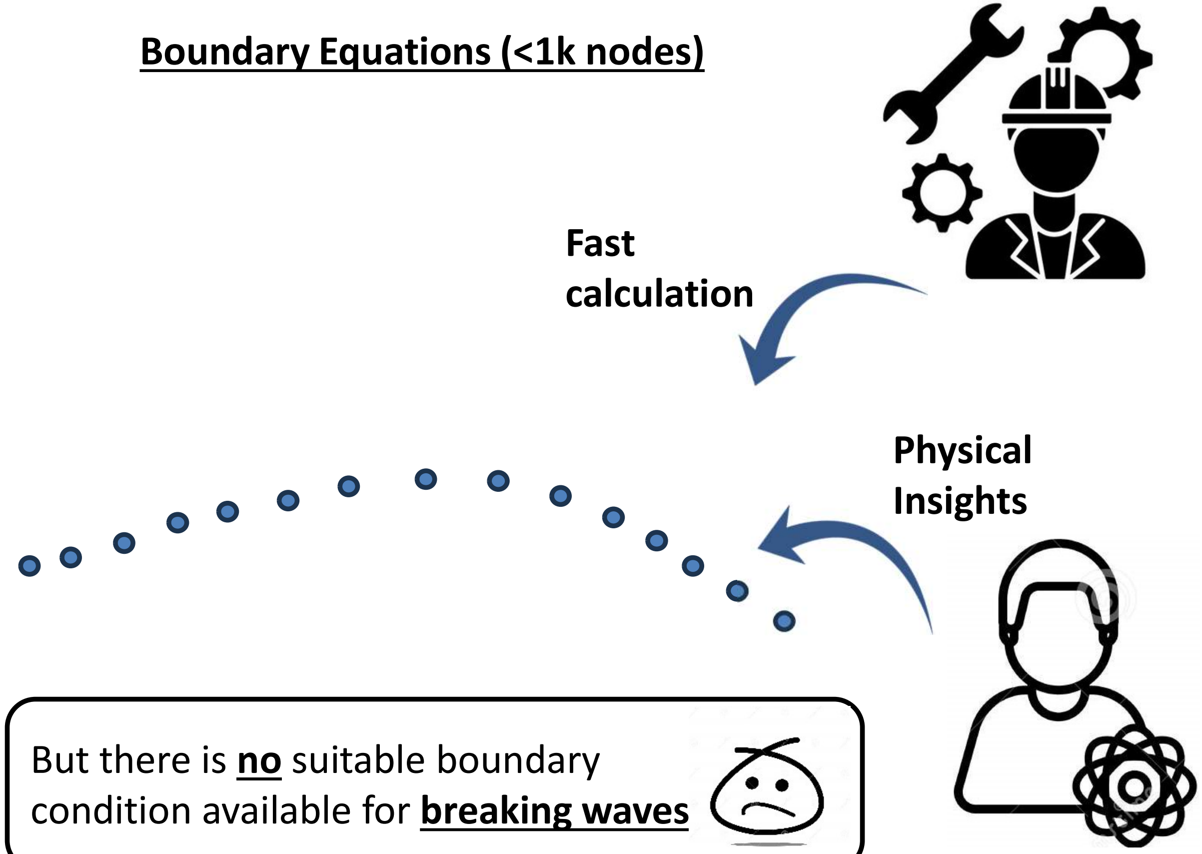
3D breaking wave with Navier Stokes Equations – **3 weeks on Cluster**

Why solving Navier Stokes Equations so slow?

Navier Stokes Equations (500k nodes)



Boundary Equations (<1k nodes)



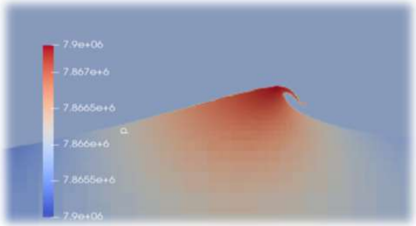
But there is **no** suitable boundary condition available for **breaking waves**

Machine Learning Approach

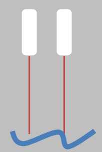
DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF OXFORD

Database Simulation with DNS

$\eta(x,z,t), u(x,z,t)\ldots$

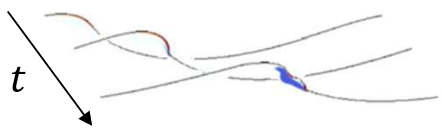


Ray Casting



Effective Description

$\eta(x,t)|_{z=\eta^*}, u(x,t)|_{z=\eta^*}\ \cdots$



PySR

New Equation Describing Data

Breaking Wave Evolution Equation

$$\eta_t = -\frac{g}{\omega_p}\eta_x - [abs(U) * \eta_x]$$

Domain Knowledge

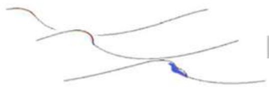
Fully nonlinear boundary conditions :

$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

$$u_t = -g\,\eta_x - u_x u - w_x w - w\,w_x\eta_x^2 - w^2\eta_x\eta_{\{x,x\}}$$

Scientific ML with PySR

Data Collection



$\eta(x,t)|_{z=\eta}$

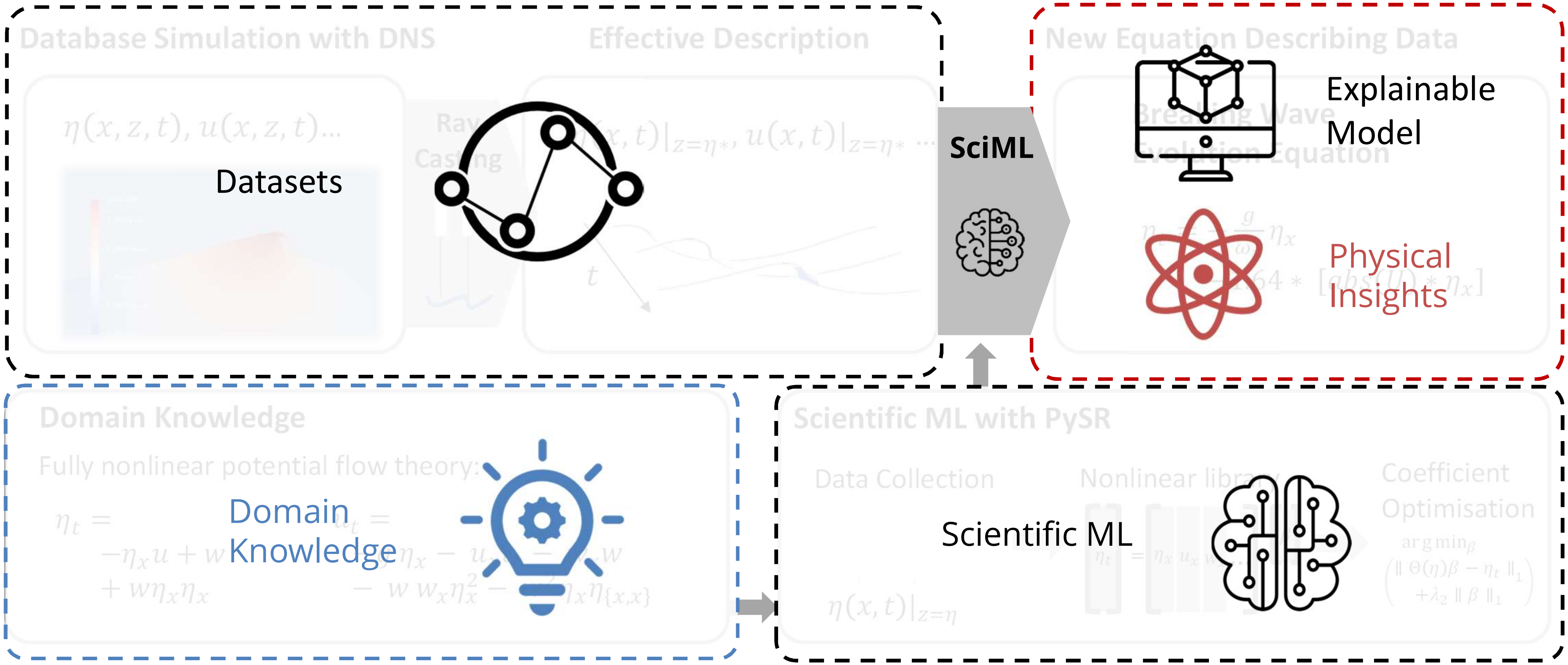
Nonlinear library

$$\begin{bmatrix}\eta_t\end{bmatrix} = \begin{bmatrix}\eta_x & u_x & w_x & \ldots\end{bmatrix}\begin{bmatrix}\beta\end{bmatrix}$$

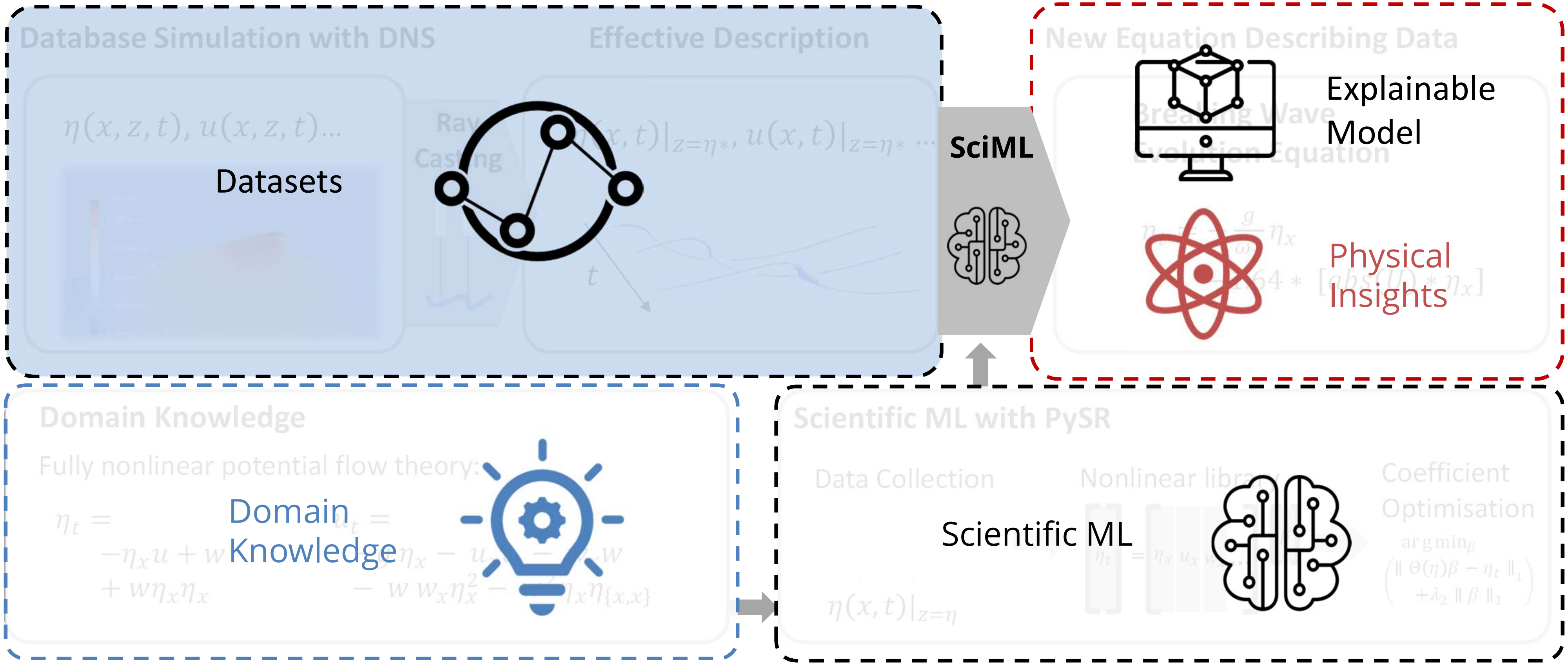
Coefficient Optimisation

$$\arg\min_\beta \left(\| \Theta(\eta)\beta - \eta_t \|_1 + \lambda_2 \| \beta \|_1 \right)$$

Machine Learning Approach



Machine Learning Approach



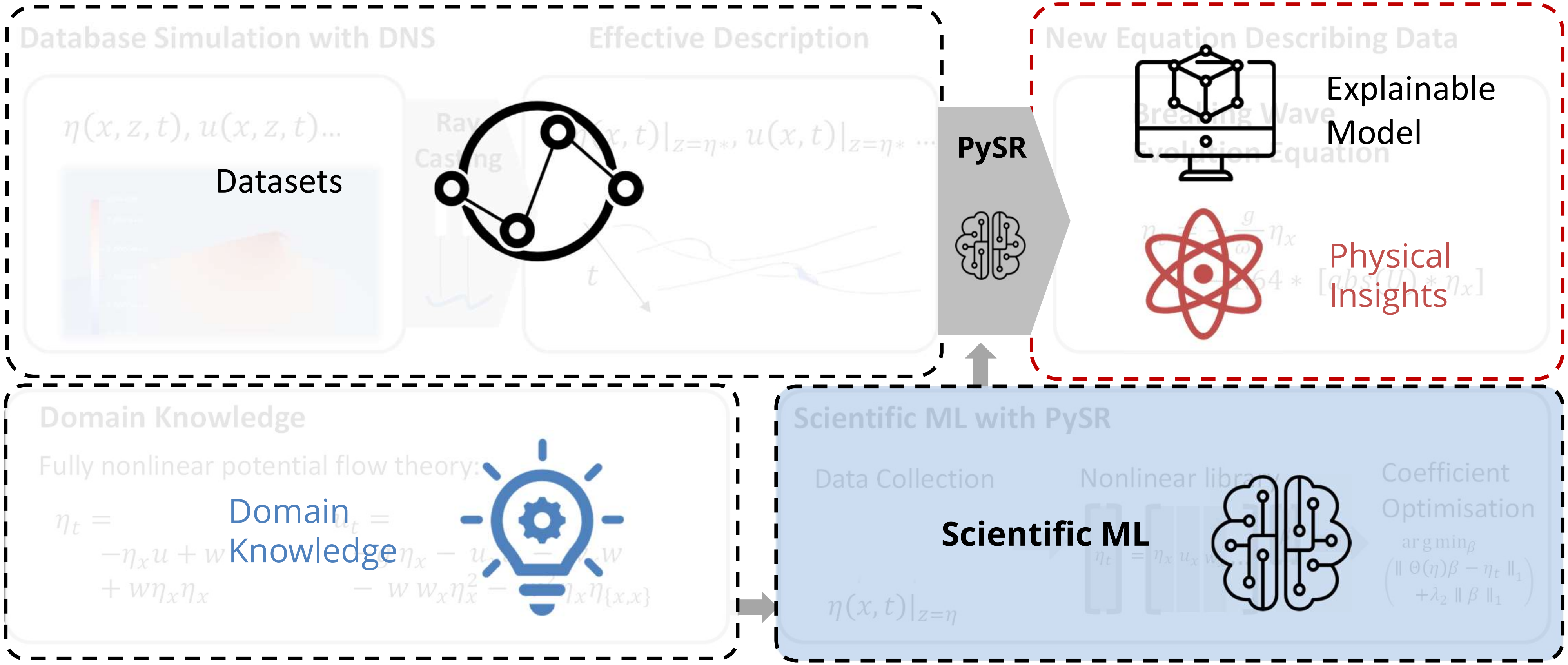
Datasets

Datasets

Over 75 2D breaking wave cases with over 1 million data points

A single case

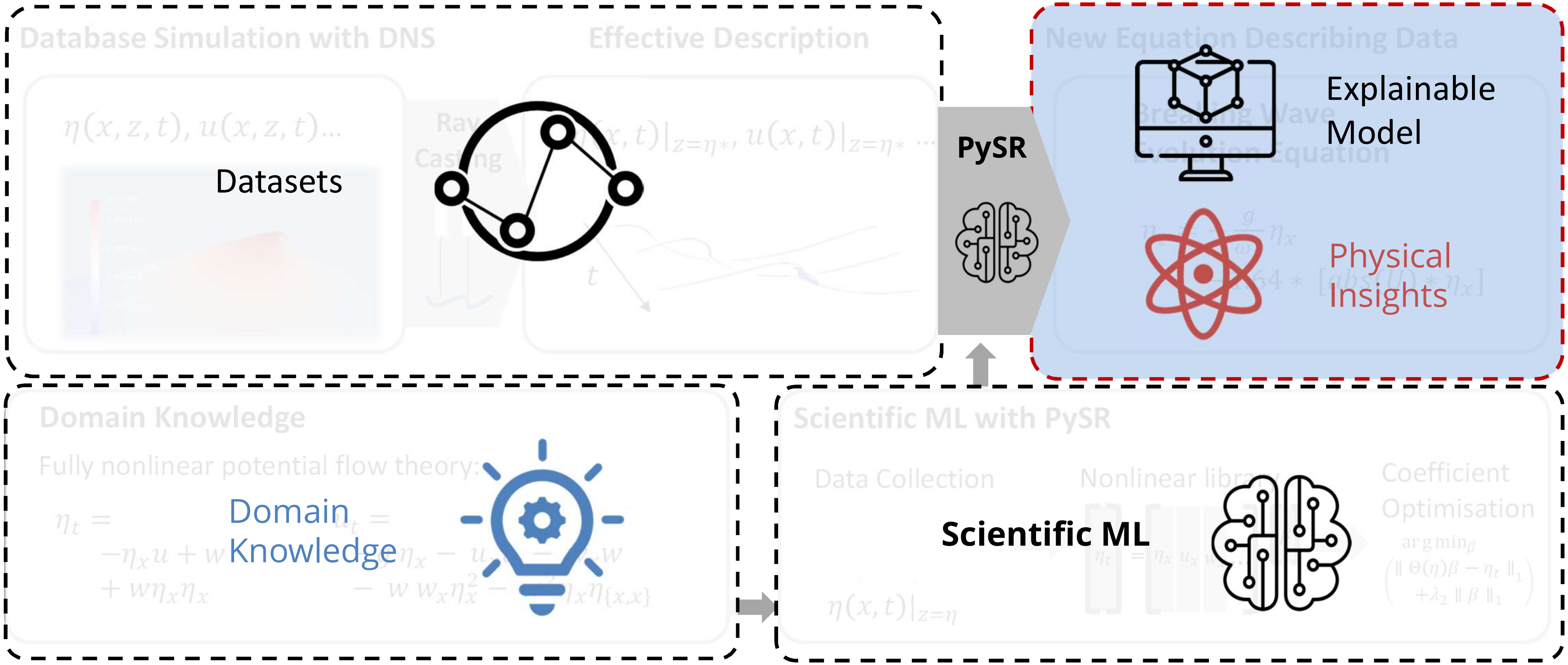
Machine Learning Approach



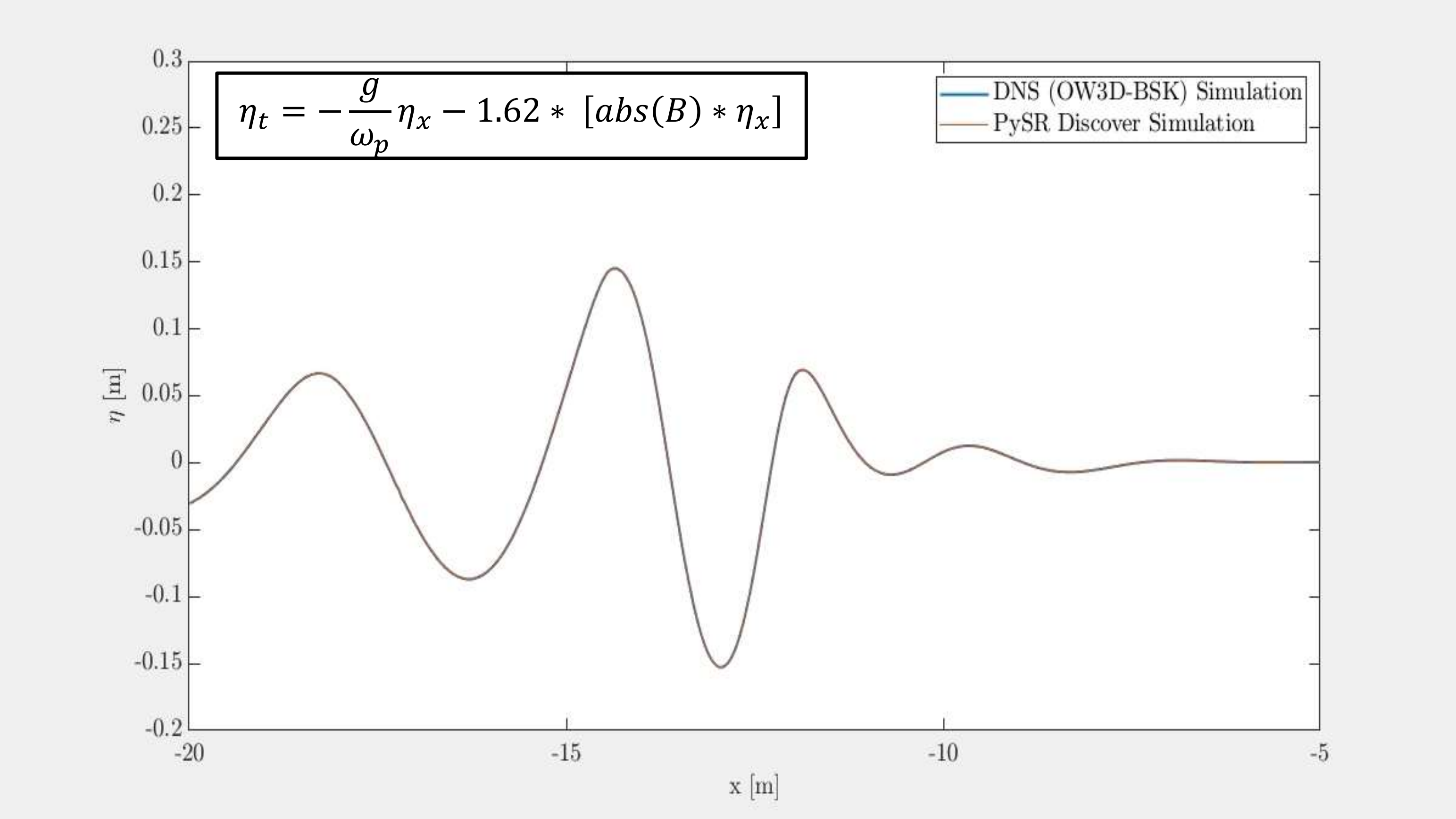
Symbolic Regression

https://github.com/MilesCranmer/PySR

Machine Learning Approach



Preliminary Results



** Only for 2d deep water spilling breakers so far*

We aim to develop a new model discovered by ML (in-progress) that:

- Overlooks bubbles and white cap details
- Equation based **numerical simulation** (*white box*)
- Very **Fast** (*2 minutes* on desktop vs *3250* of core *hours* on supercomputer)
- Mathematically interpretable
- Directly applicable to various scales of the wave

Physical insights

Non-breaking evolution

Breaking evolution

FNBC framework:

$$\eta_t = -1\eta_x u + 1w + O(3)$$

SciML discovered equation:

$$\eta_t = -1.058\eta_x u + 0.98w + O(3)$$

FNBC framework:

$$\eta_t = -1\eta_x u + 1w + O(3)$$

SciML discovered equation:

$$\eta_t = -\frac{g}{\omega_p}\eta_x + 1.6\, abs(U)\eta_x + O(3)$$

Residual is reduced significantly

Physical insights

Non-breaking evolution

FNBC framework:

$$\eta_t = -1\eta_x u + 1w + O(3)$$

SciML discovered equation:

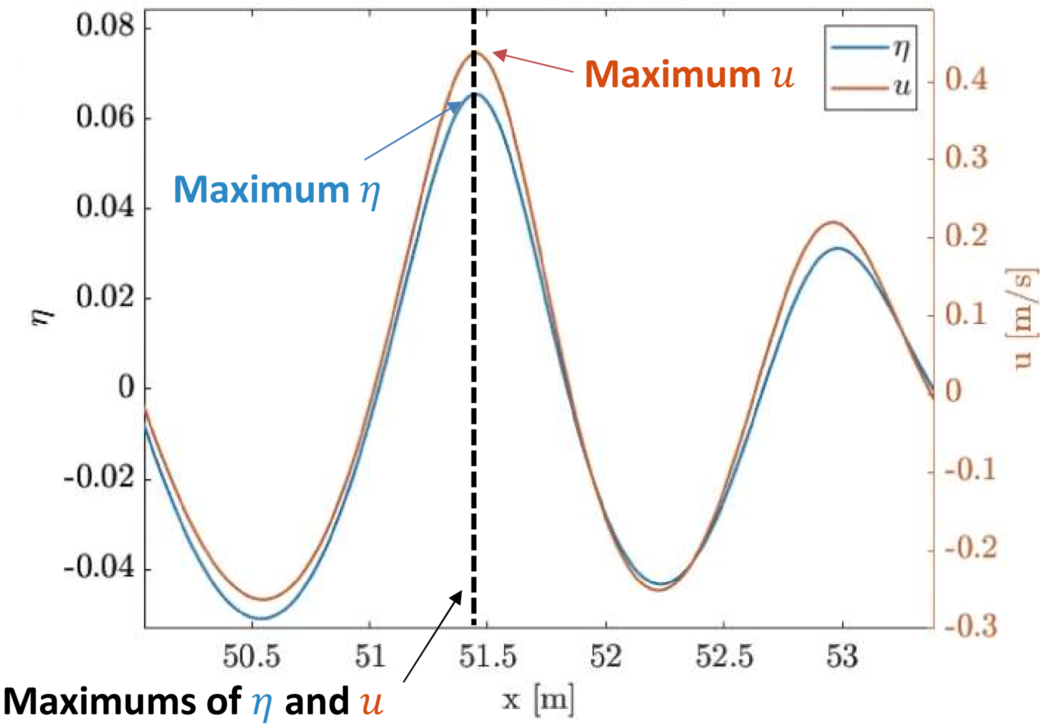
$$\eta_t = -1.058\eta_x u + 0.98w + O(3)$$

Terms with surface elevation

Terms with velocities

FNBC & SciML:

$\eta_t \leftarrow u$ terms & η terms



Maximums of η and u are aligned

Physical insights

Breaking evolution

FNBC framework:

$$\eta_t = 1w - 1\eta_x u + O(3)$$

SciML discovered equation:

$$\eta_t = -\frac{g}{\omega_p}\eta_x + 1.6\, abs(U)\eta_x + O(3)$$

Terms with surface elevation (η) **only**

Why the SciML uses surface elevation terms (η) **only** ?

FNBC:

$\eta_t \leftarrow$ u terms & η terms

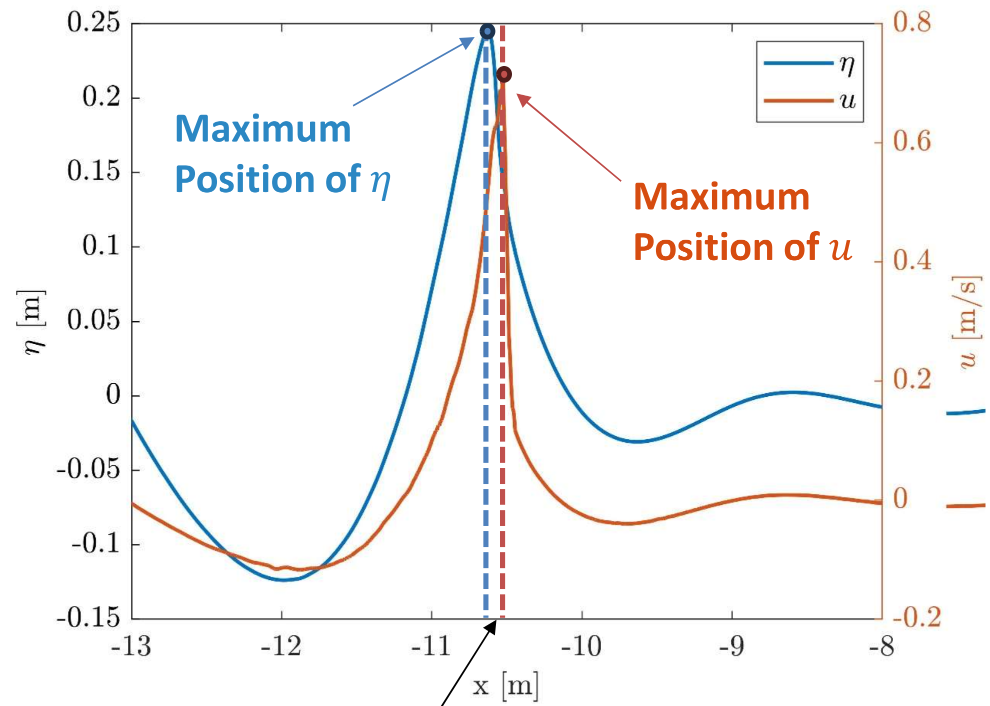
SciML :

$\eta_t \leftarrow$ η terms

$\neq$

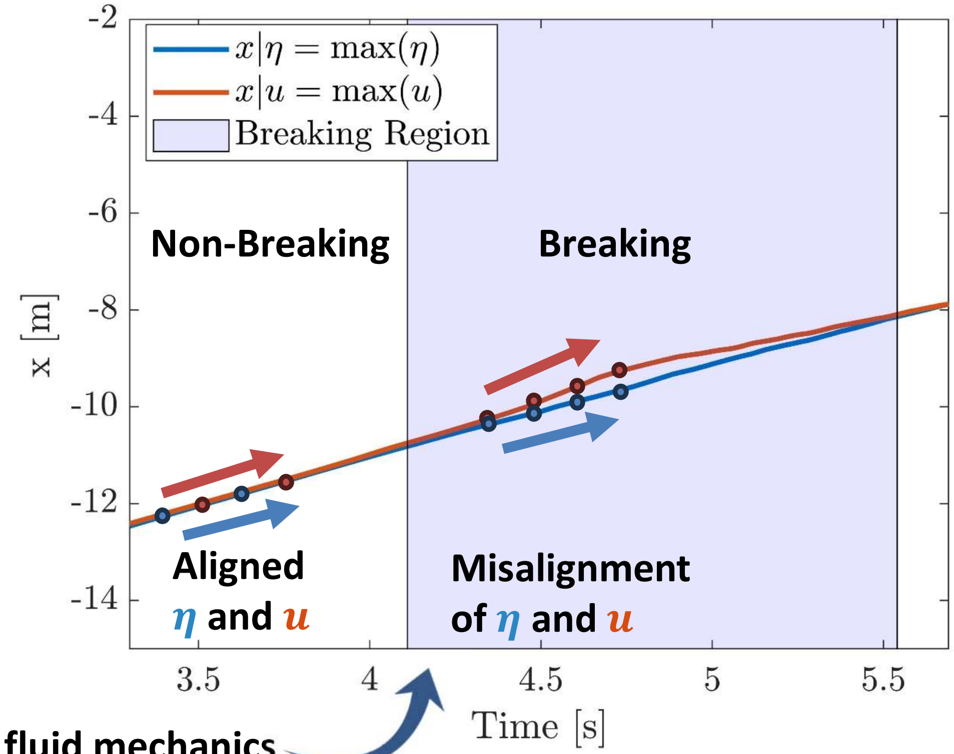
Physical insights

Maximum η and u



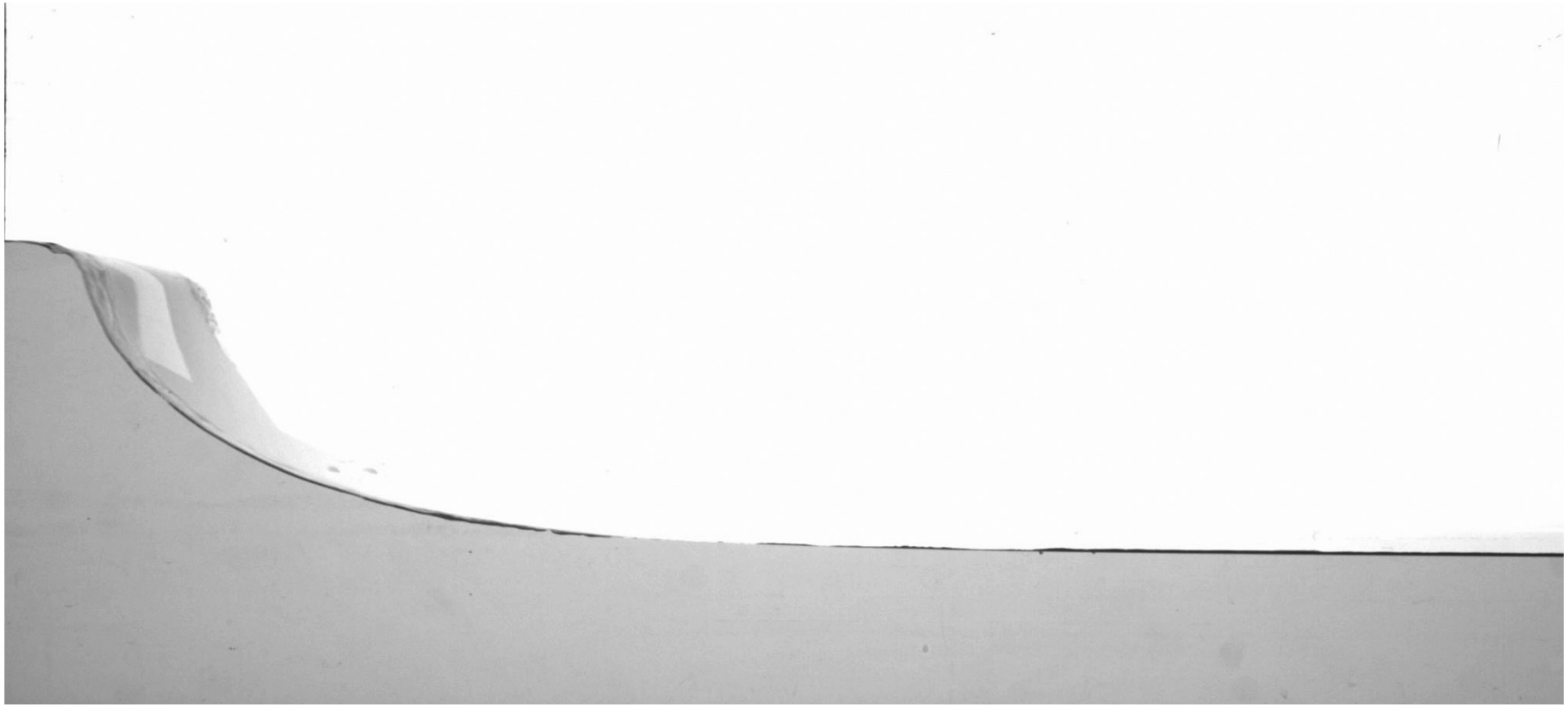
Maximums of η and u are NOT aligned

Evolution of the maximum position η and u

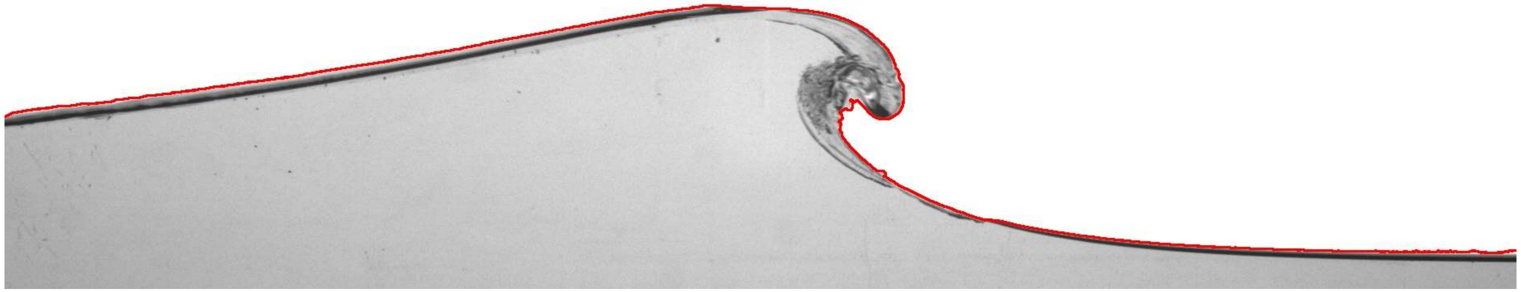


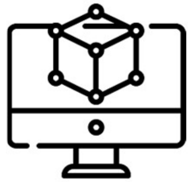
New findings in fluid mechanics

Experiments

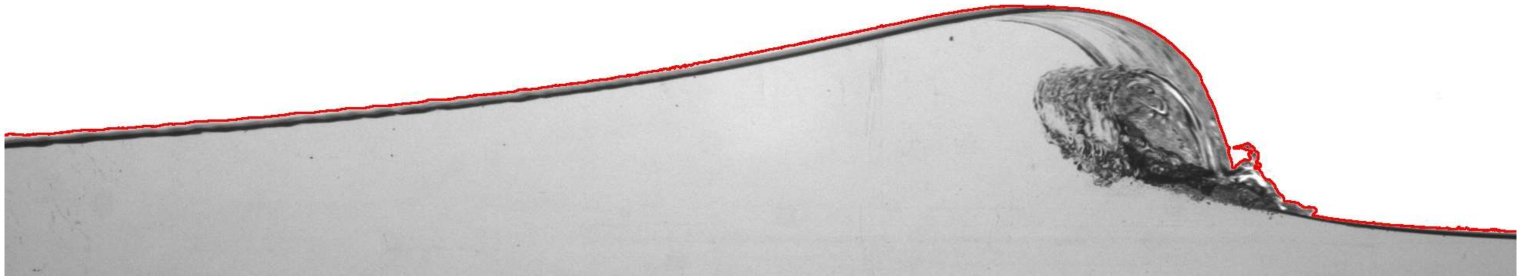


Preliminary Results

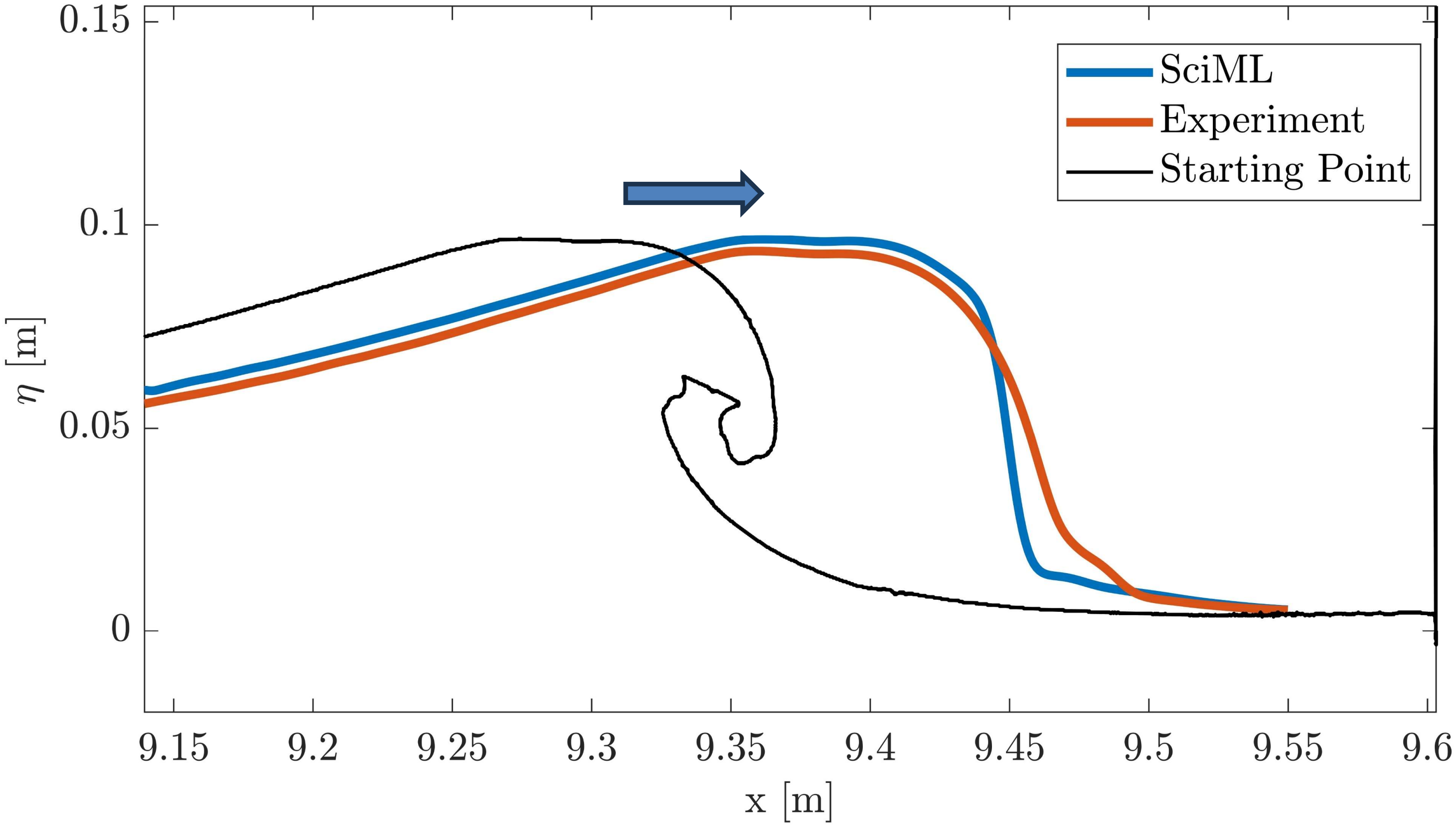




Machine Learning Discovered Equation

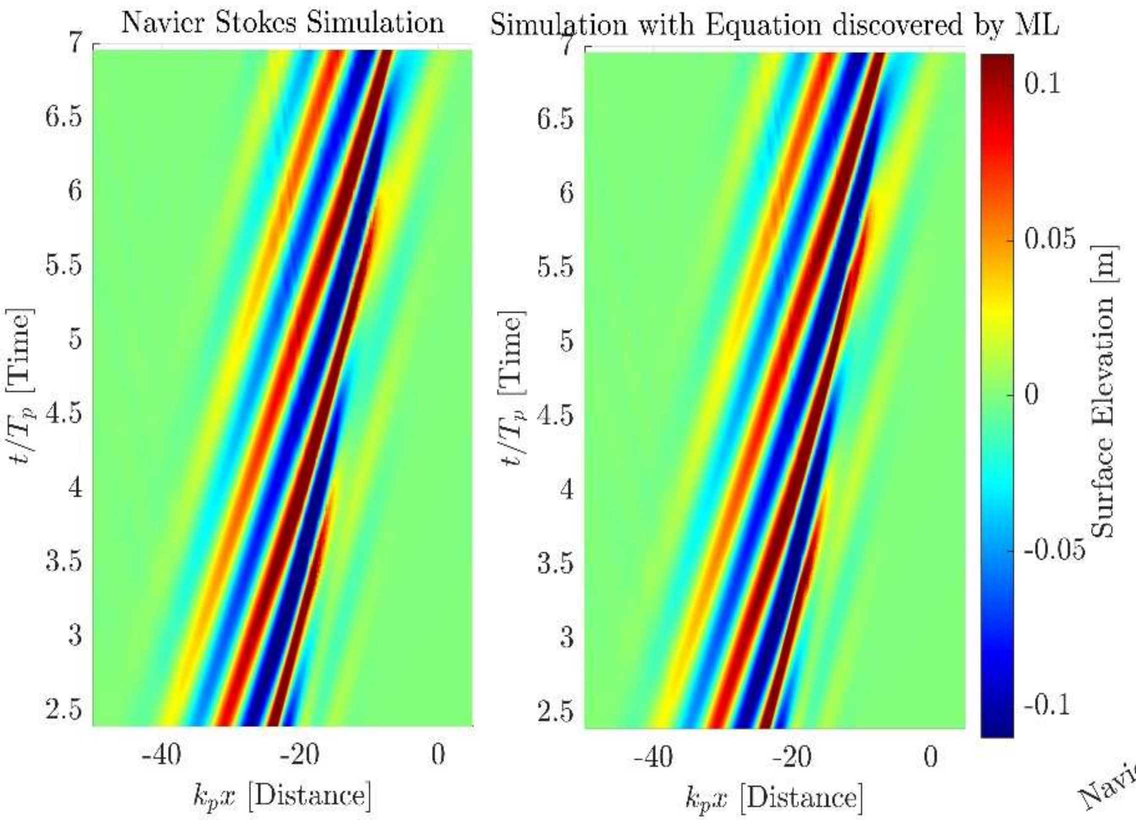


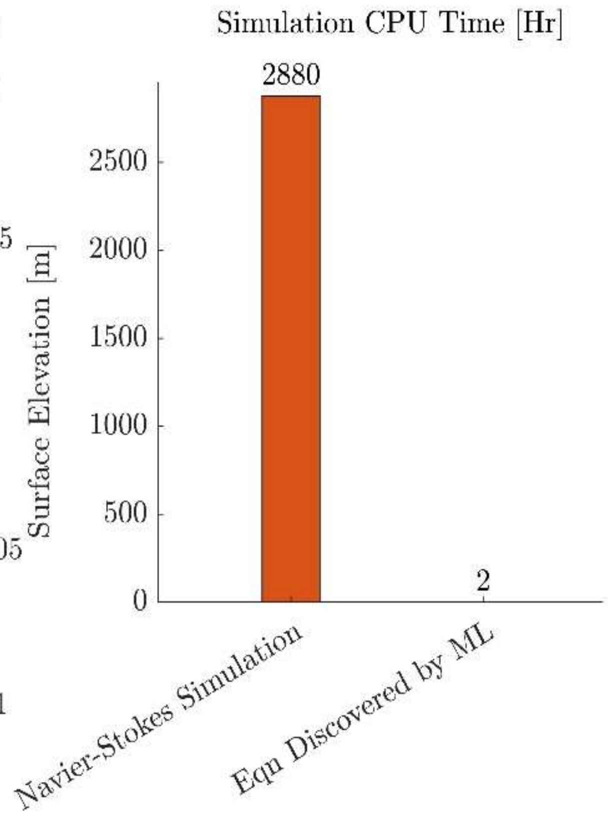
Preliminary Results



Conclusion

- **Accurate** approximation of breaking wave with Equation discovered by SciML
- **Significant** reduction in modelling time
- **New physical insights**





Breaking evolution

Misalignment of η and u

Thank you!

Tianning Tang (Tim);
Email: tianning.tang@eng.ox.ac.uk

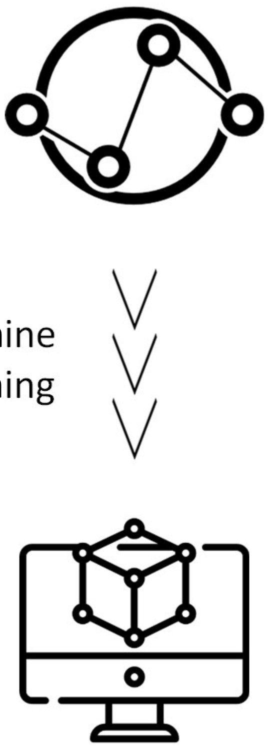
Can we now do better with lagre amount of data?

Machine Learning

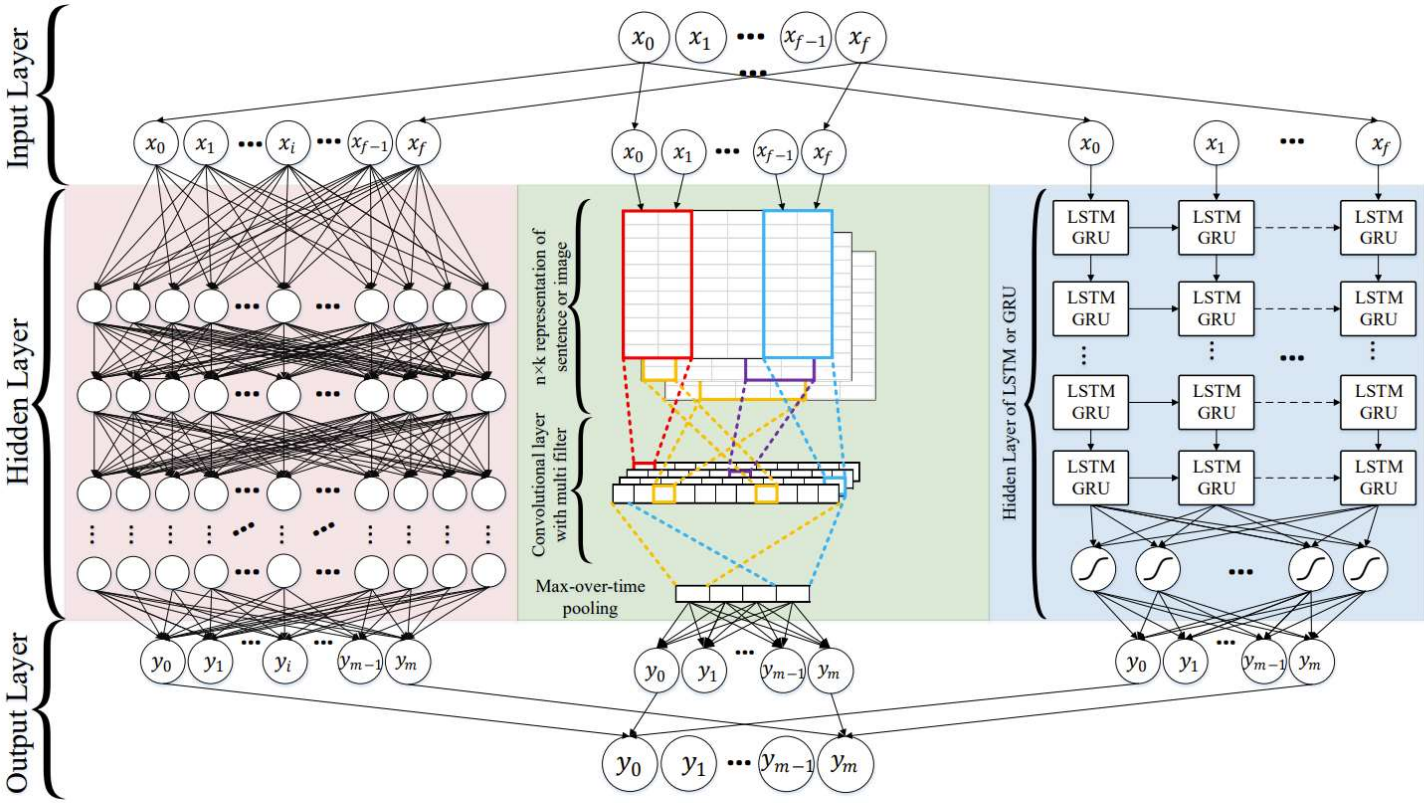
Datasets

Machine Learning

Empirical Model

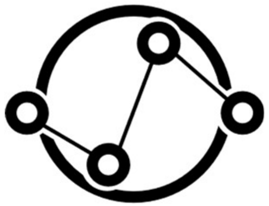


RMDL: Random Multimodel Deep Learning for Classification



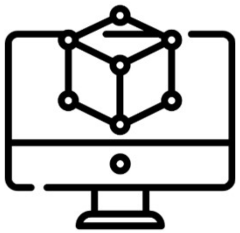
Machine Learning

Datasets

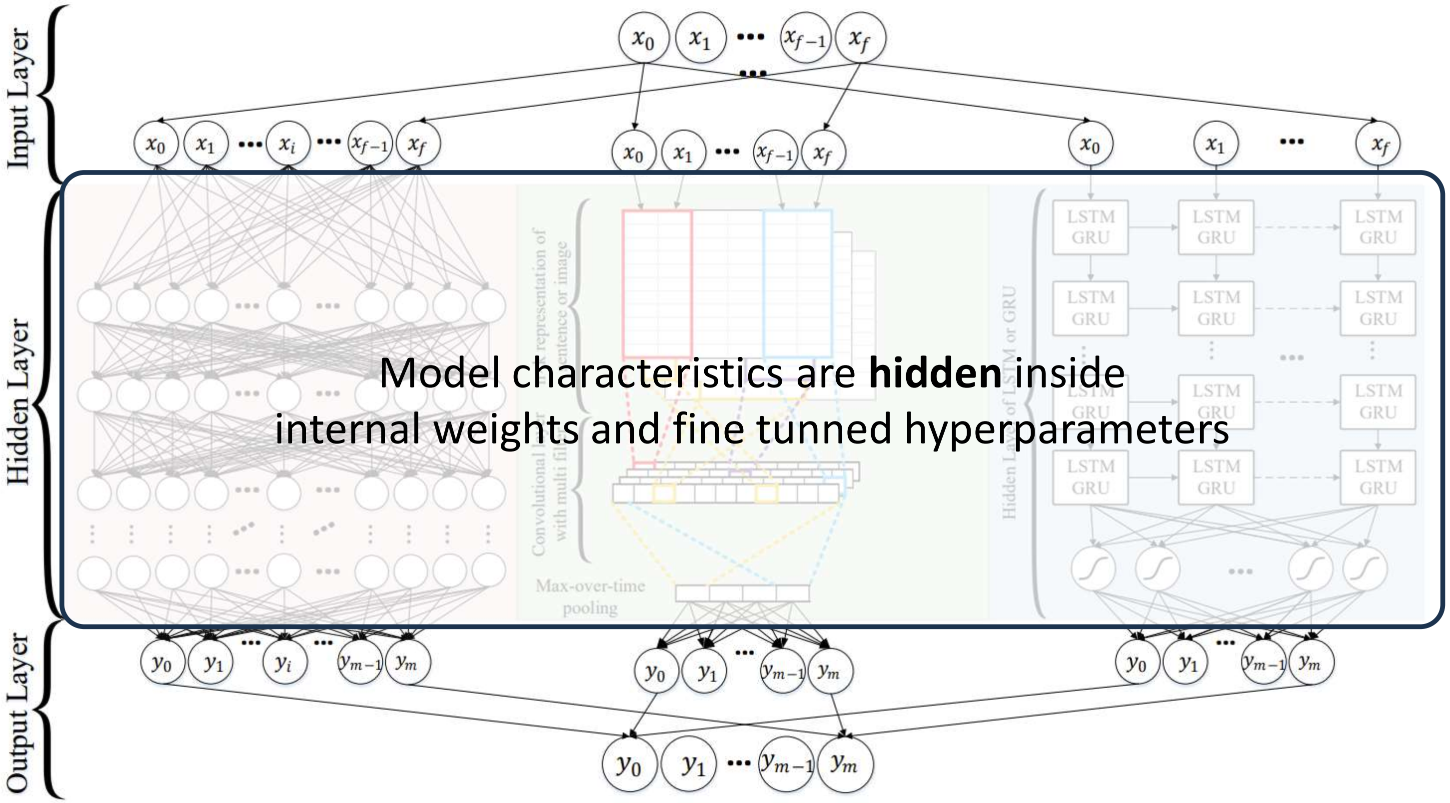


Machine
Learning

Empirical
Model



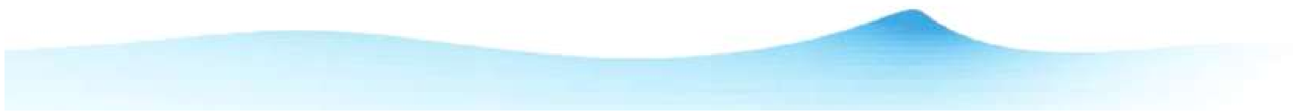
RMDL: Random Multimodel Deep Learning for Classification



Model characteristics are **hidden** inside
internal weights and fine tunned hyperparameters

Domain Knowledge

Non-breaking evolution



Fully Nonlinear Boundary Conditions (FNBC)

Time derivatives of surface elevation

Spatial derivatives of surface elevation and velocities

$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

$$\boxed{\eta_t = \mathrm{FNBC}}$$

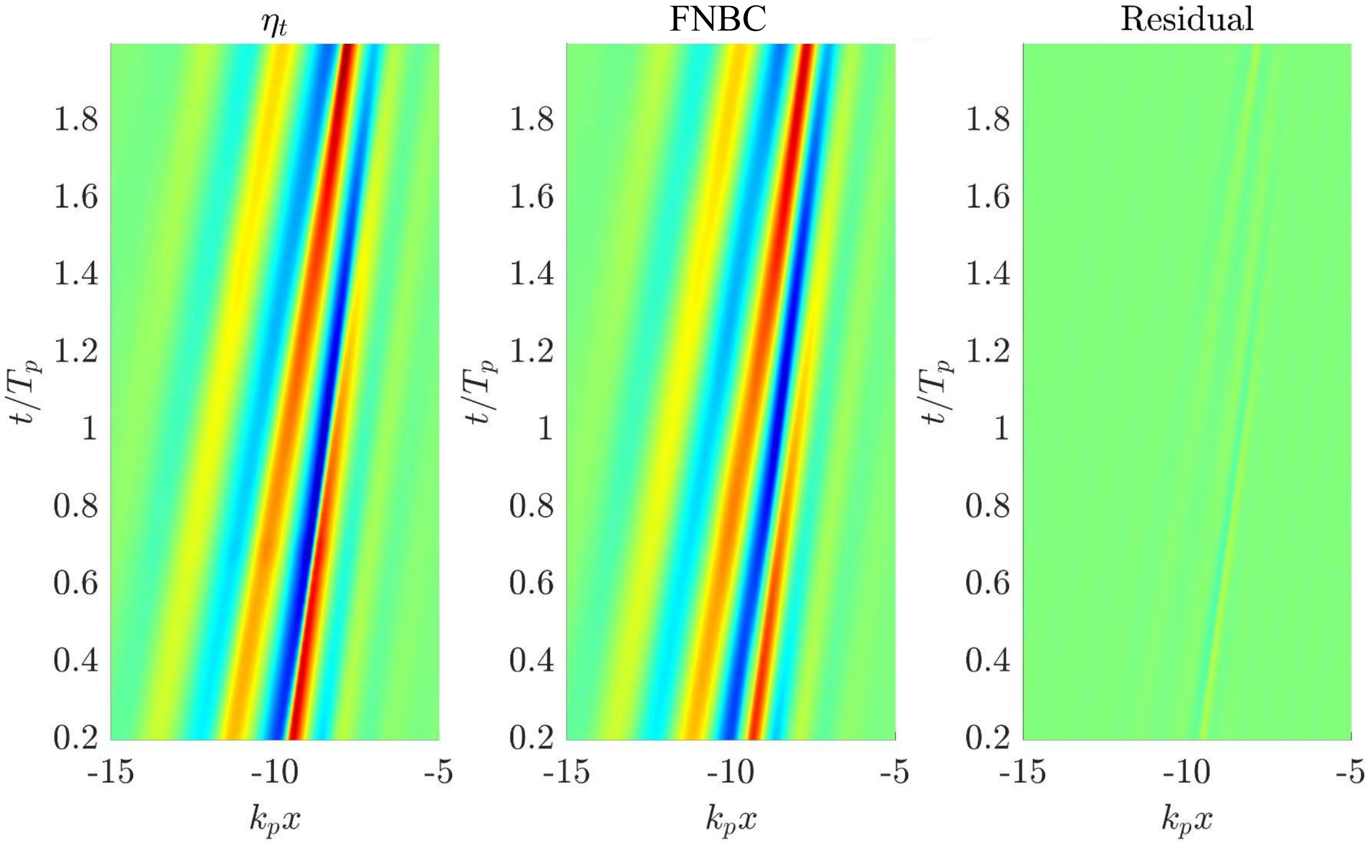
η and u are *coupled* in FNBC framework

*Subscripts denote partial differentiation

Domain Knowledge

$$\boxed{\eta_t = \text{FNBC}}$$

LHS: **RHS:** **Res = LHS - RHS**

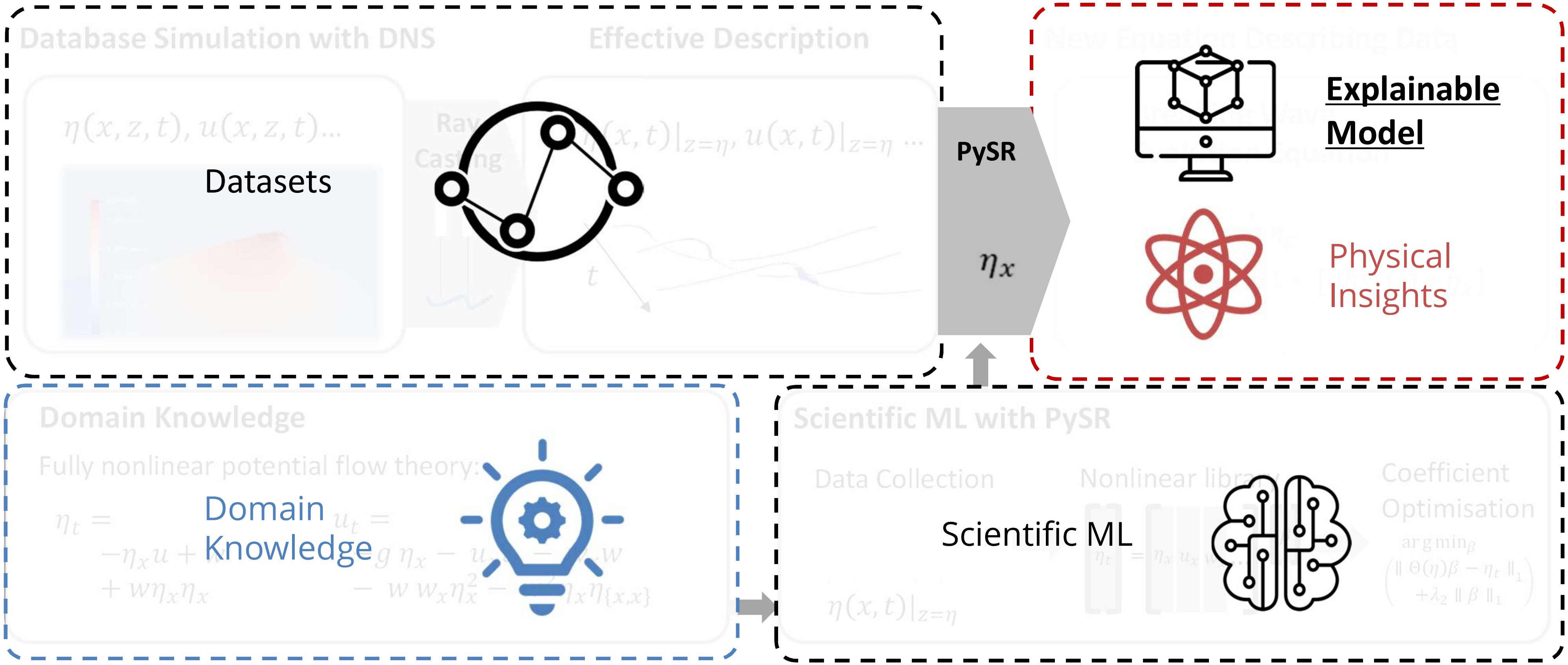


<u>Non-breaking evolution</u>

As expected, for non-breaking evolution, we have less than 1% of difference between the LHS and RHS of the equation.

FNBC equation **works well** for non-breaking evolution.

Machine Learning Approach

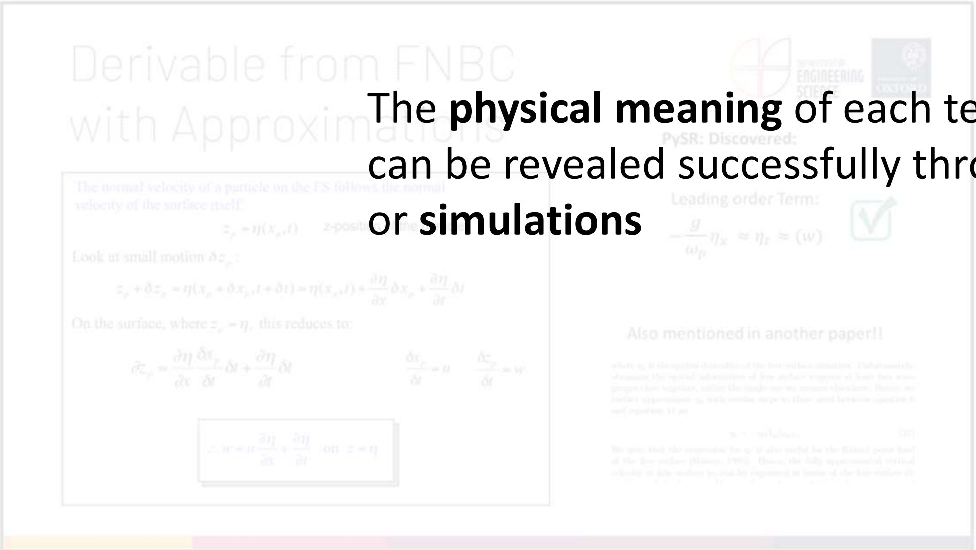


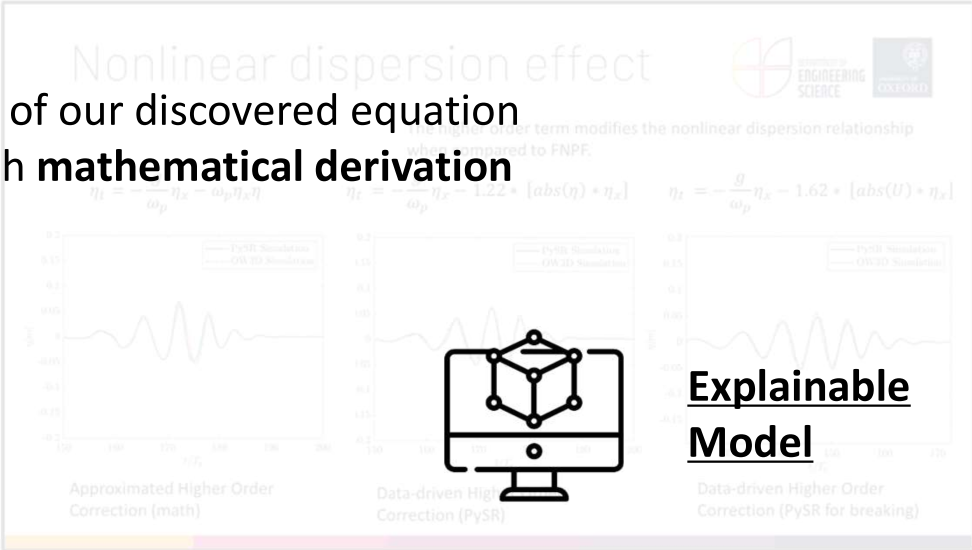
PySR new equation

Discovered boundary condition equation:

Envelope of η

$$\eta_t = -\frac{g}{\omega_p}\eta_x + [-1.6 * abs(B) * \eta_x] + O(3)$$





The **physical meaning** of each term of our discovered equation can be revealed successfully through **mathematical derivation** or **simulations**

Explainable Model

What makes a wave break?

How machine learning can shed light on the underlying physics of breaking waves

Tianning Tang (Tim); Schmidt AI in Science Fellow

28/11/2023

Supervisors and Collaborators

Prof. Thomas Adcock

Prof. Paul Taylor

Prof. Yuntian Chen

Prof. Steve Roberts

Dr. Ben Lambert

Dr. Martin Robinson

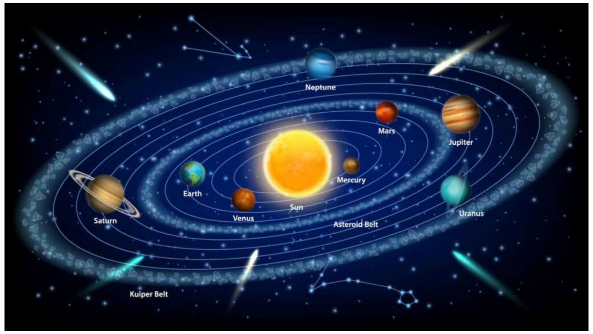
Discovering Physics from Data

Johannes Kepler
(1571 - 1630)

Data used by Kepler (1618)

Planet	Mean distance to sun (AU)	Period (days)	$\frac{R^3}{T^2}$ (10^{-6} AU^3/day^2)
Mercury	0.389	87.77	7.64
Venus	0.724	224.70	7.52
Earth	1	365.25	7.50
Mars	1.524	686.95	7.50
Jupiter	5.20	4332.62	7.49
Saturn	9.510	10759.2	7.43

Constant

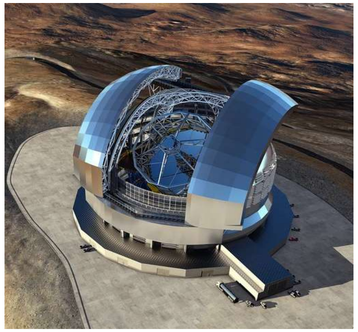


Planetary system

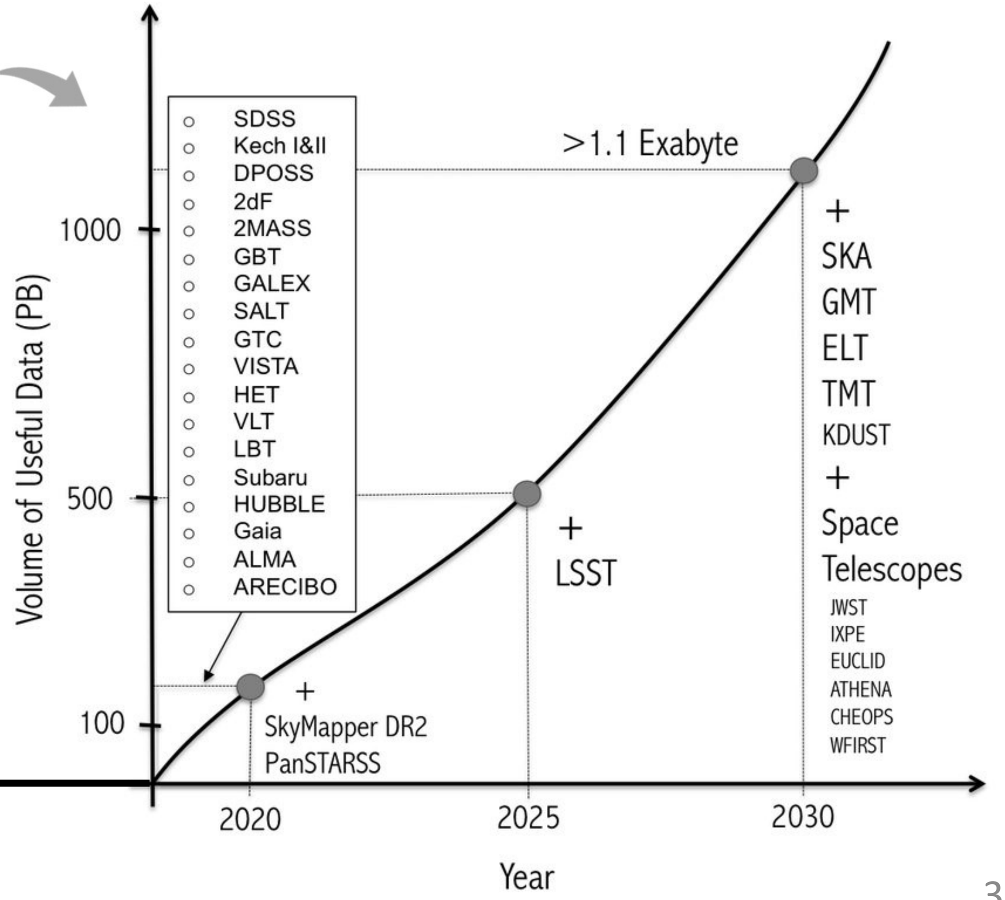
$$\longrightarrow \quad mr\omega^2 = G\frac{mM}{r^2}$$

Newton's law of gravitation (published 1687)

The Astrophysics Data

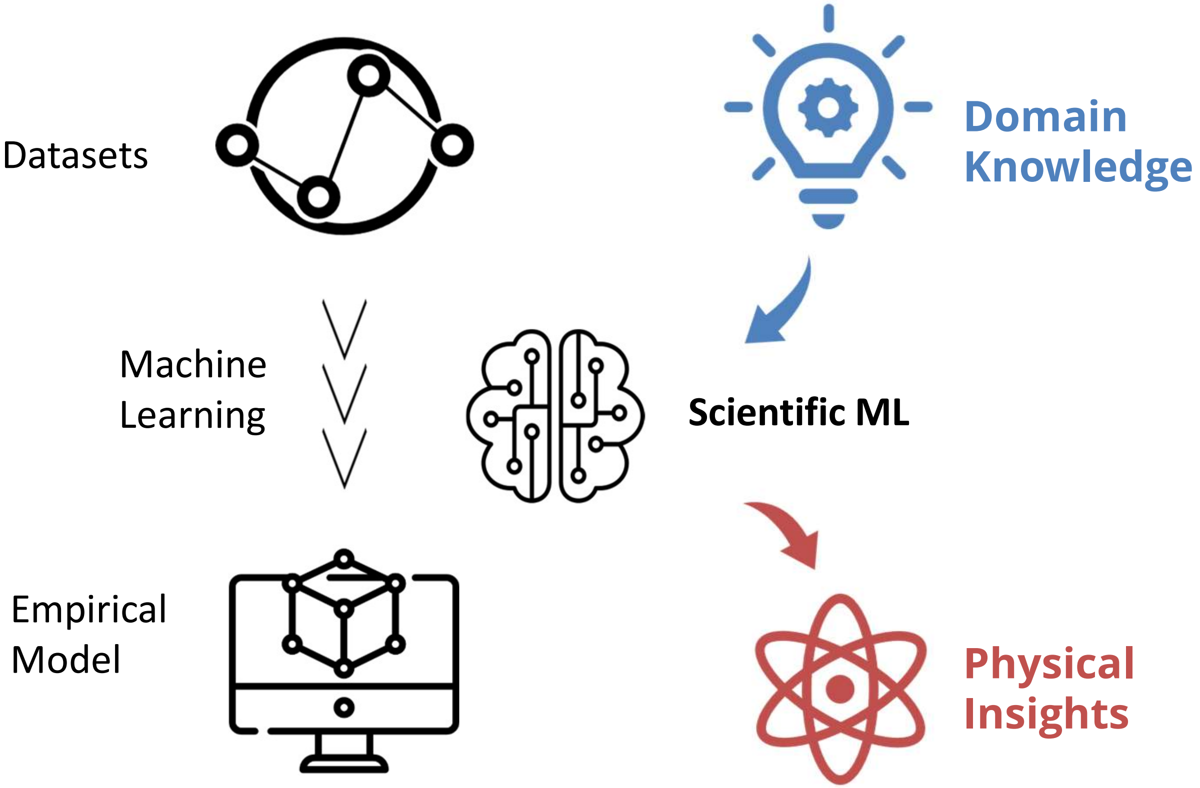


Johannes Kepler
(1571 - 1630)



Can we do better in obtaining physical insights from data after 400 years?

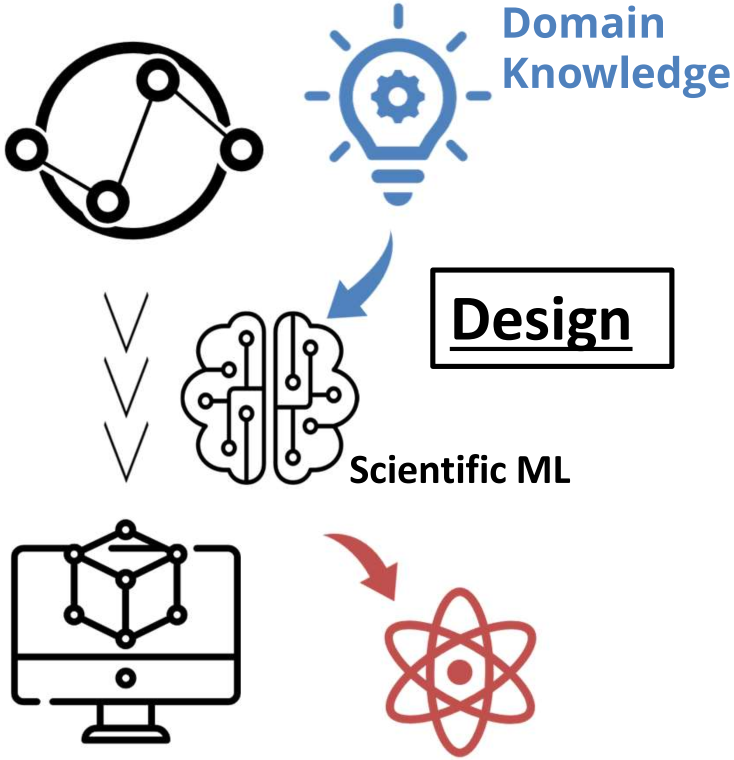
Scientific Machine Learning

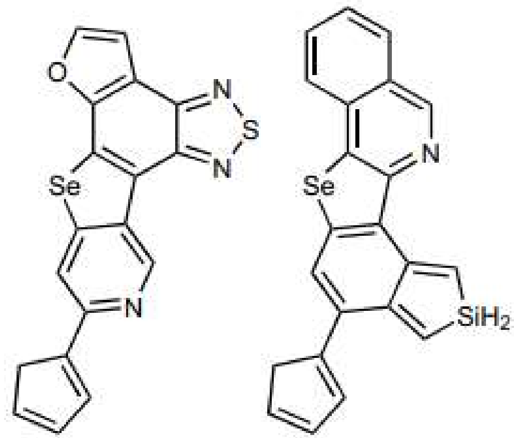


SciML seeks to address **domain-specific** data challenges and extract **insights** from scientific datasets through innovative methodological solutions.

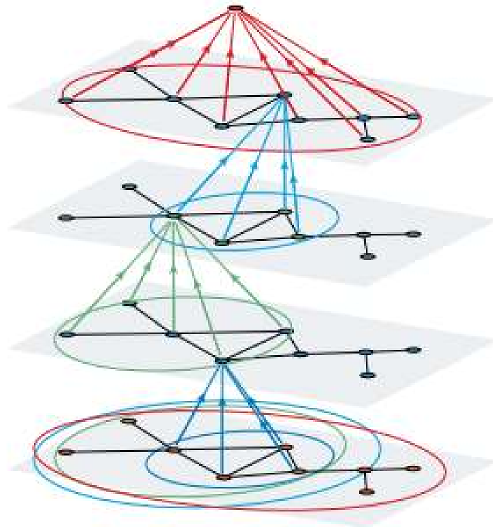
- Brown University

Scientific Machine Learning





Molecular Graphs

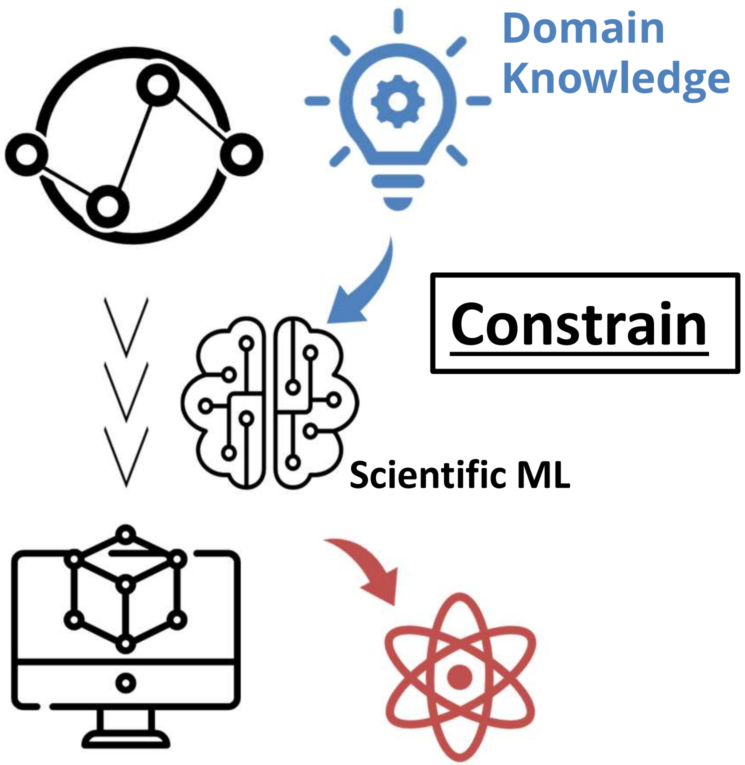


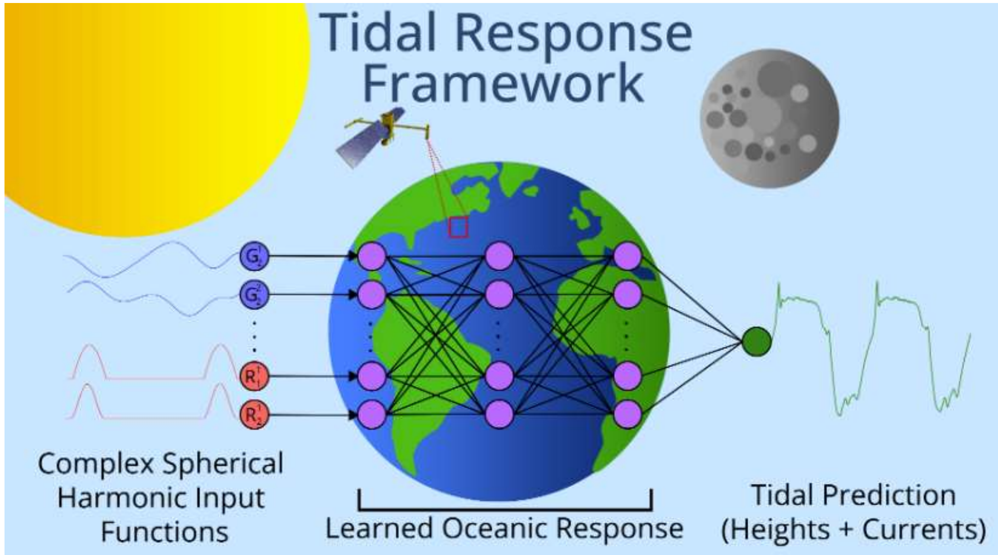
New **architecture** based on graph neural networks

Objective: Predicting chemical properties

Truong Son Hy (2018)

Scientific Machine Learning



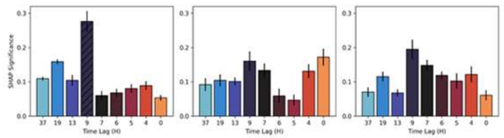


Impulse → Response → Prediction

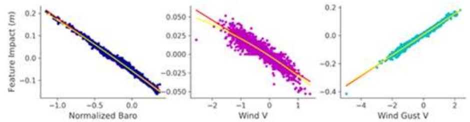
Interpretation

Shap

Oceanic Resonance

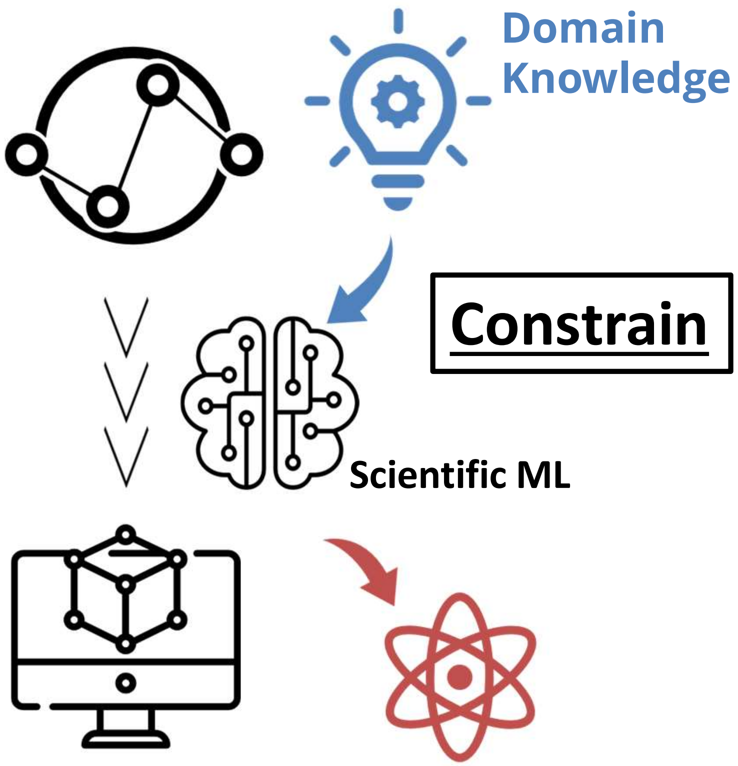


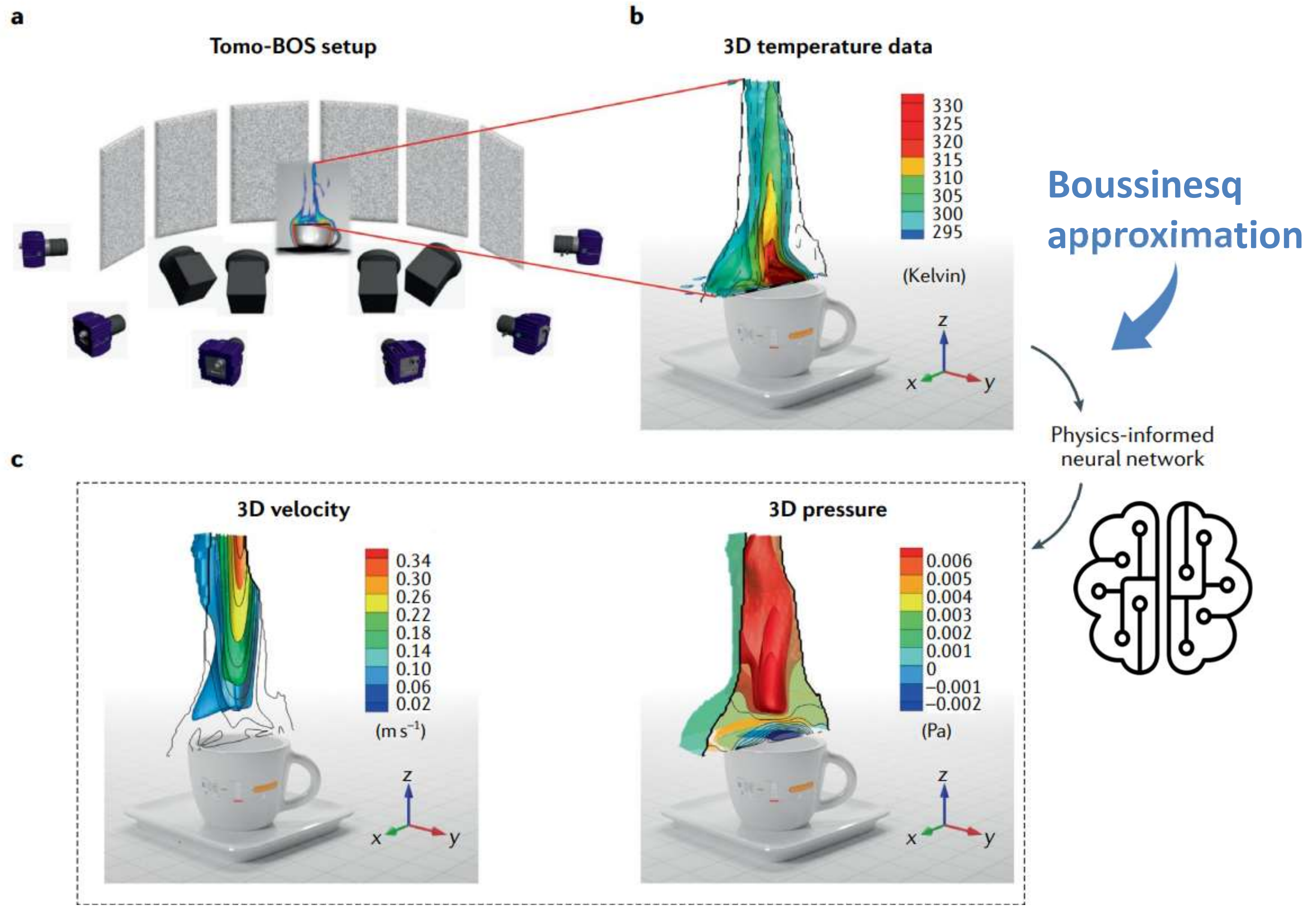
Surge Forecasting



Thomas Monahan (2023, under review)

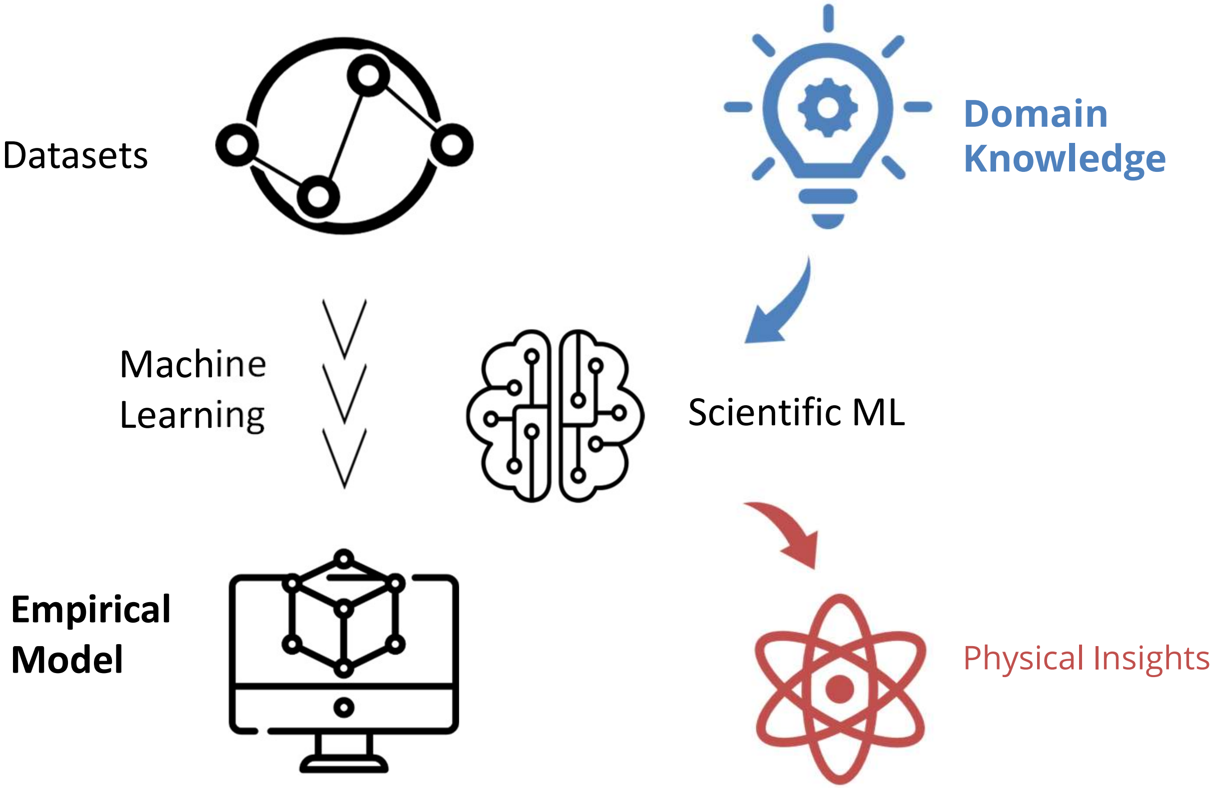
Scientific Machine Learning





Shengze Cai (2021)

Scientific Machine Learning

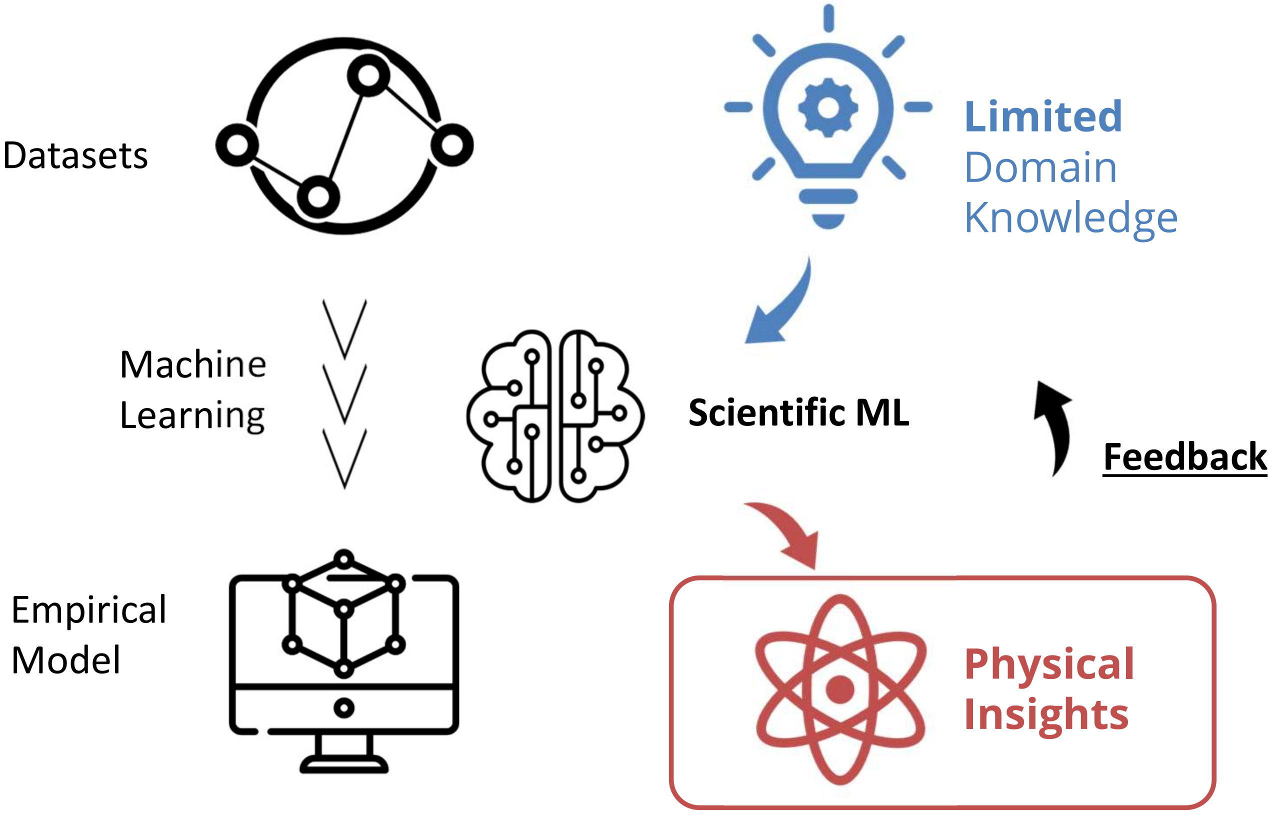


Assumption:

The Domain Knowledge is (close to) sufficient for the underlying system to guide the Empirical Model.

What if the domain knowledge is insufficient?

"Knowledge Discovery"



Knowledge discovery is the process of directly mining important internal principles (i.e., governing equations) from observations and experimental data through machine learning.

"Knowledge Discovery"

Symbolic Regression

Trying to find ***analytical expressions*** of the dataset.
Prior works include Langley et al., 1980s; Koza et al., 1990s; Lipson et al., 2000s etc.

Pool of equations

$1.15y + 0.86$

x^x

$\sin x$

$x^2 y$

…

Select Fittest: $1.15y + 0.86$ x^x

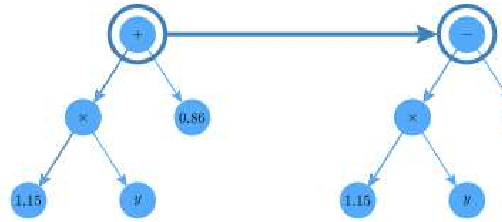
Select an operation:

Mutation or Crossover etc.

New equations to replace old ones:

$x + 0.86$ $(1.15y)^x$

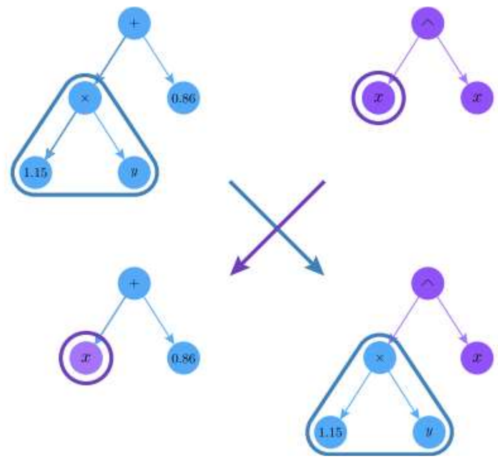
Mutation:



$1.15y + 0.86$ $1.15y - 0.86$

Crossover:

$1.15y + 0.86$ x^x



$x + 0.86$ $(1.15y)^x$

Symbolic Regression

https://github.com/MilesCranmer/PySR

Wave Breaking

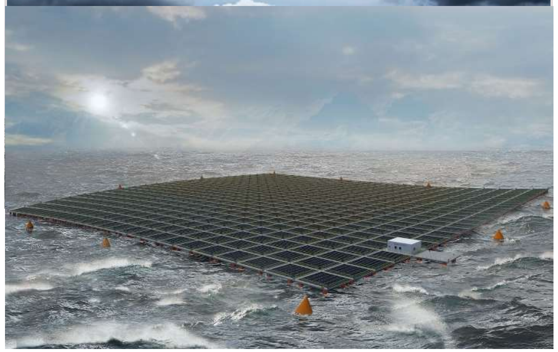
Wave Breaking

Wave breaking occurs where the wave amplitude reaches the critical point that the crest self- disassembled

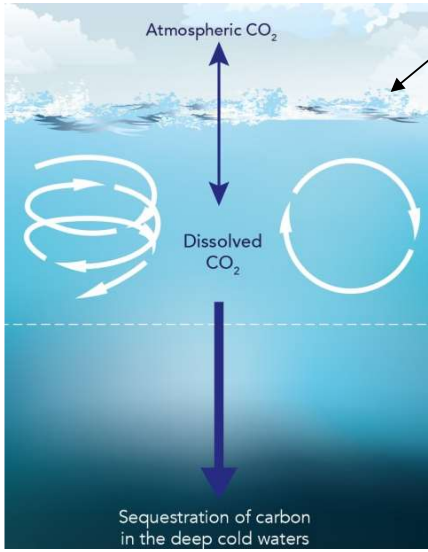
- **Wave Breaking on beach**

- **Wave Breaking in Storm**

- **Renewable Energy**
- Offshore wind
- Wave energy converter
- Offshore floating solar etc.



- **Carbon cycle**



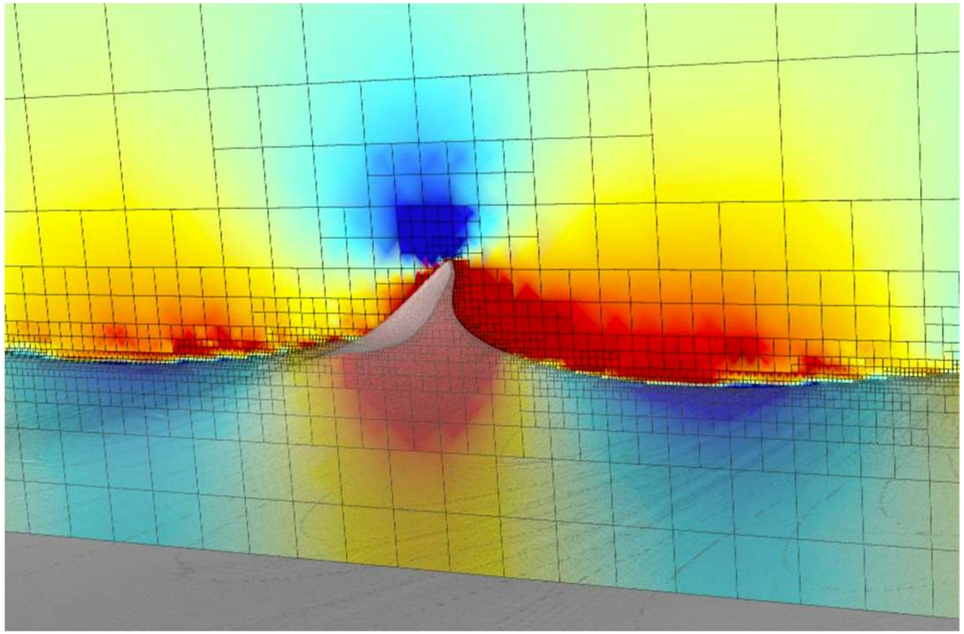
Wave breaking

Water movement enhancing gas exchanges

Background

However, modelling these breaking waves requires DNS solving the **Navier-Stokes equation** and are very **computational heavy**.

2D breaking wave with Navier Stokes Equations – **3 days on Cluster**

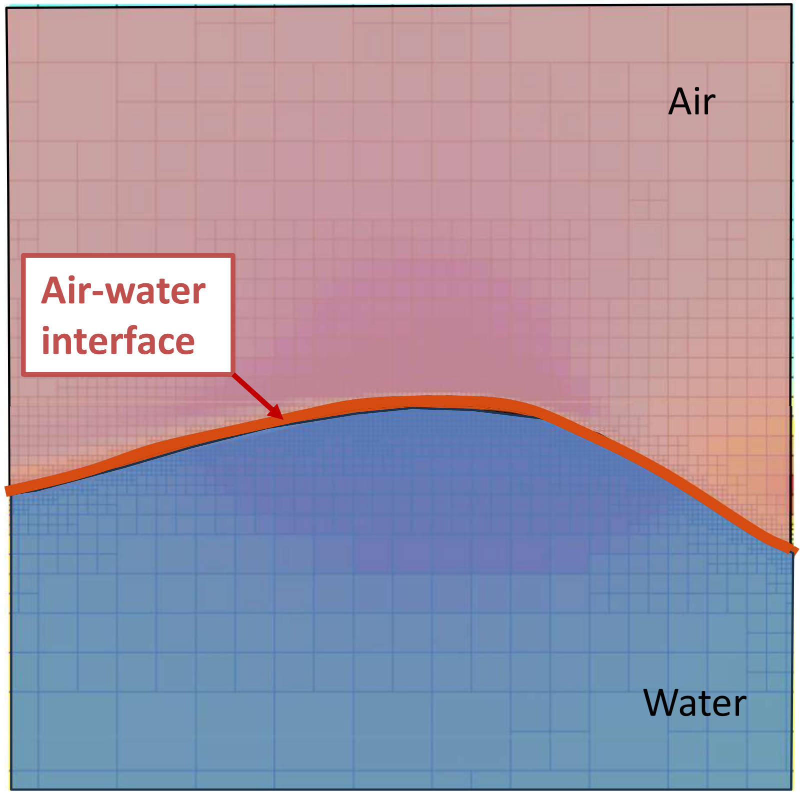


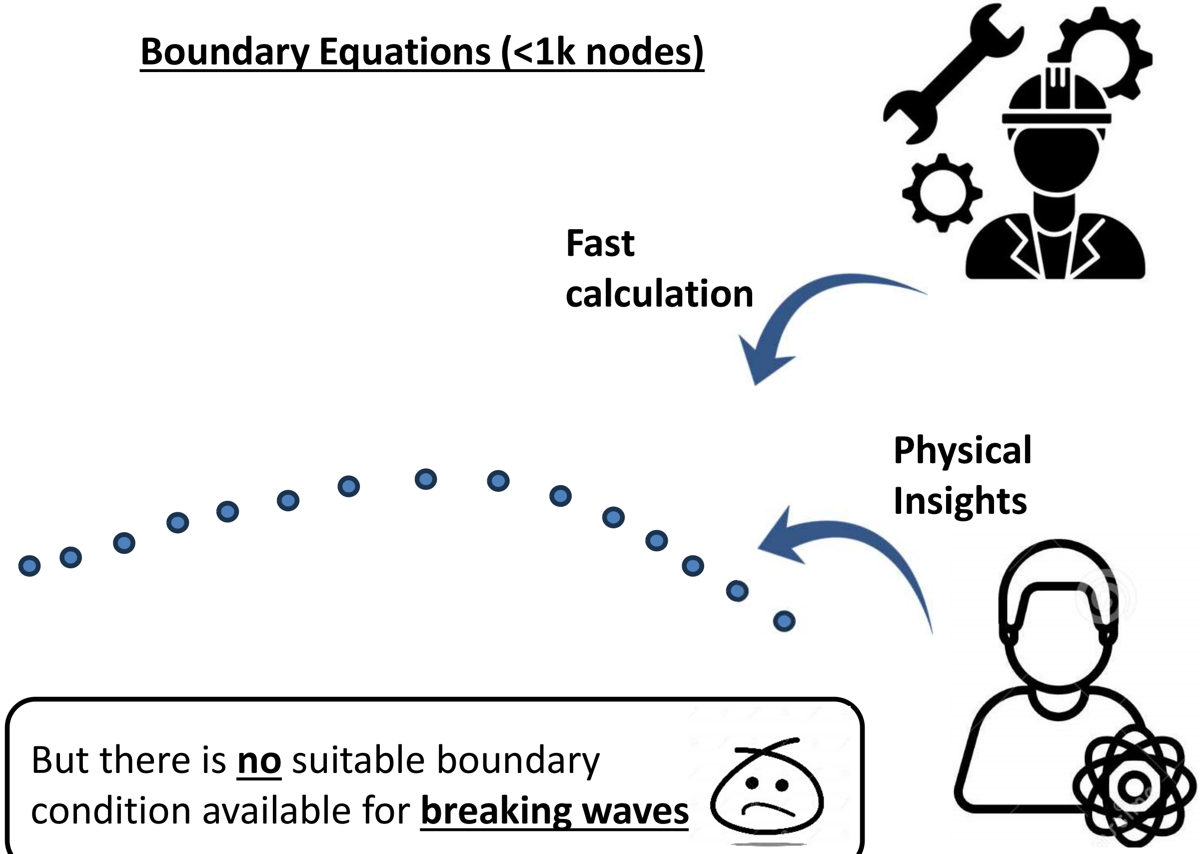
3D breaking wave with Navier Stokes Equations – **3 weeks on Cluster**

Why solving Navier Stokes Equations so <u>slow</u>?

<u>Navier Stokes Equations (500k nodes)</u>

<u>Boundary Equations (<1k nodes)</u>

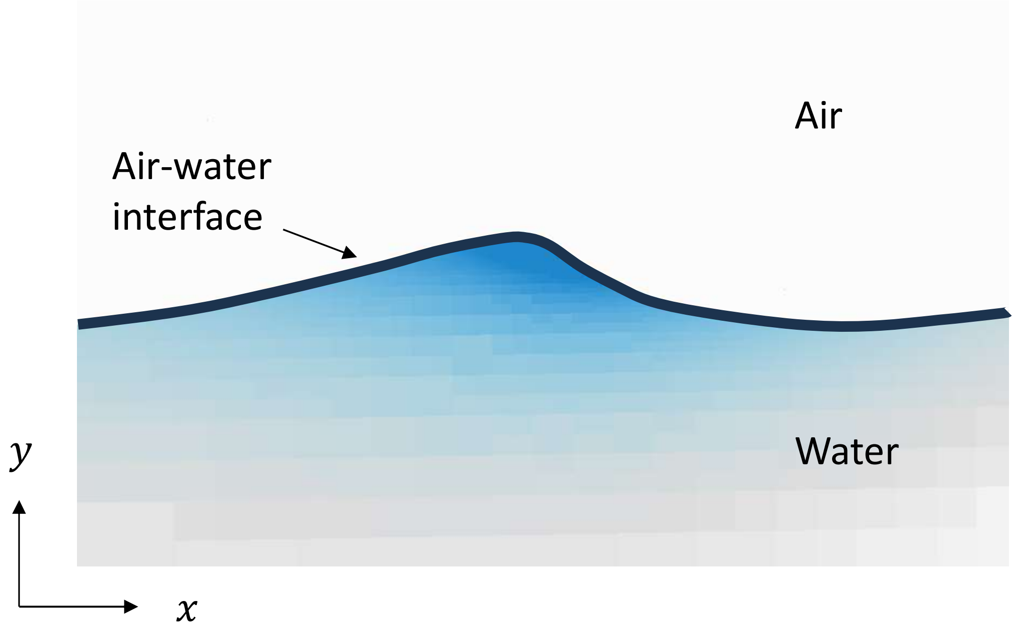




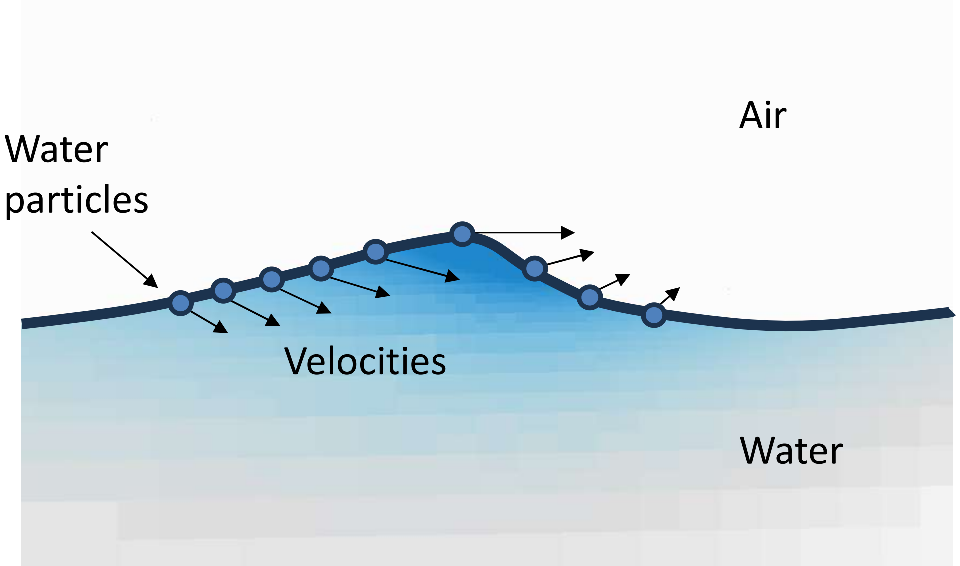
But there is **<u>no</u>** suitable boundary condition available for **<u>breaking waves</u>**

Describe a Wave with Boundary Conditions

Surface Elevation (η)



Velocity in x and y direction (u , w)

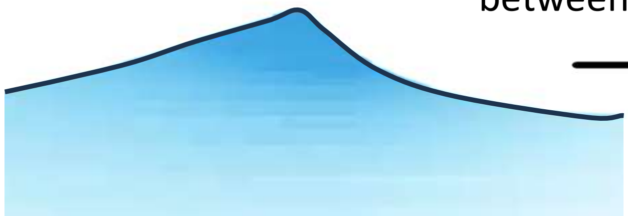


Objective

Simulate wave in time

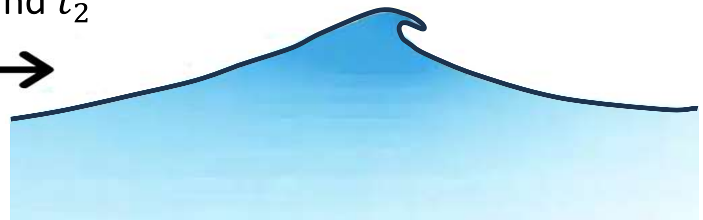
$\eta_{t1 \Rightarrow t2}$: Change of η between t_2 and t_1

$\eta_{t2 \Rightarrow t3}$: Change of η between t_3 and t_2



$\eta \; @ \; t = t_1$

$\eta \; @ \; t = t_2$



$\eta \; @ \; t = t_3$

Governing Boundary Equation for Breaking Waves

$$\eta_t = N(\eta, \eta_x, \eta_{xx}, u, u_x, w \cdots, \mu)$$

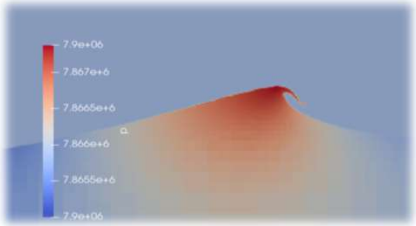
Time derivatives of surface elevation $\eta(x, t)$

Unknown right-hand side (RHS): generally, some spatial derivatives η_x etc, and parameters μ

Machine Learning Approach

Database Simulation with DNS

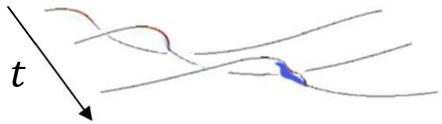
$\eta(x,z,t), u(x,z,t)$...



Ray Casting

Effective Description

$\eta(x,t)|_{z=\eta*}, u(x,t)|_{z=\eta*}$ ...



PySR

New Equation Describing Data

Breaking Wave Evolution Equation

$$\eta_t = -\frac{g}{\omega_p}\eta_x$$
$$-1.64 * \ [abs(U) * \eta_x]$$

Domain Knowledge

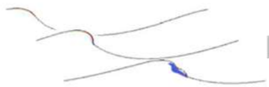
Fully nonlinear boundary conditions :

$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

$$u_t = -g\,\eta_x - u_x u - w_x w - w\,w_x\eta_x^2 - w^2\eta_x\eta_{\{x,x\}}$$

Scientific ML with PySR

Data Collection



$\eta(x,t)|_{z=\eta}$

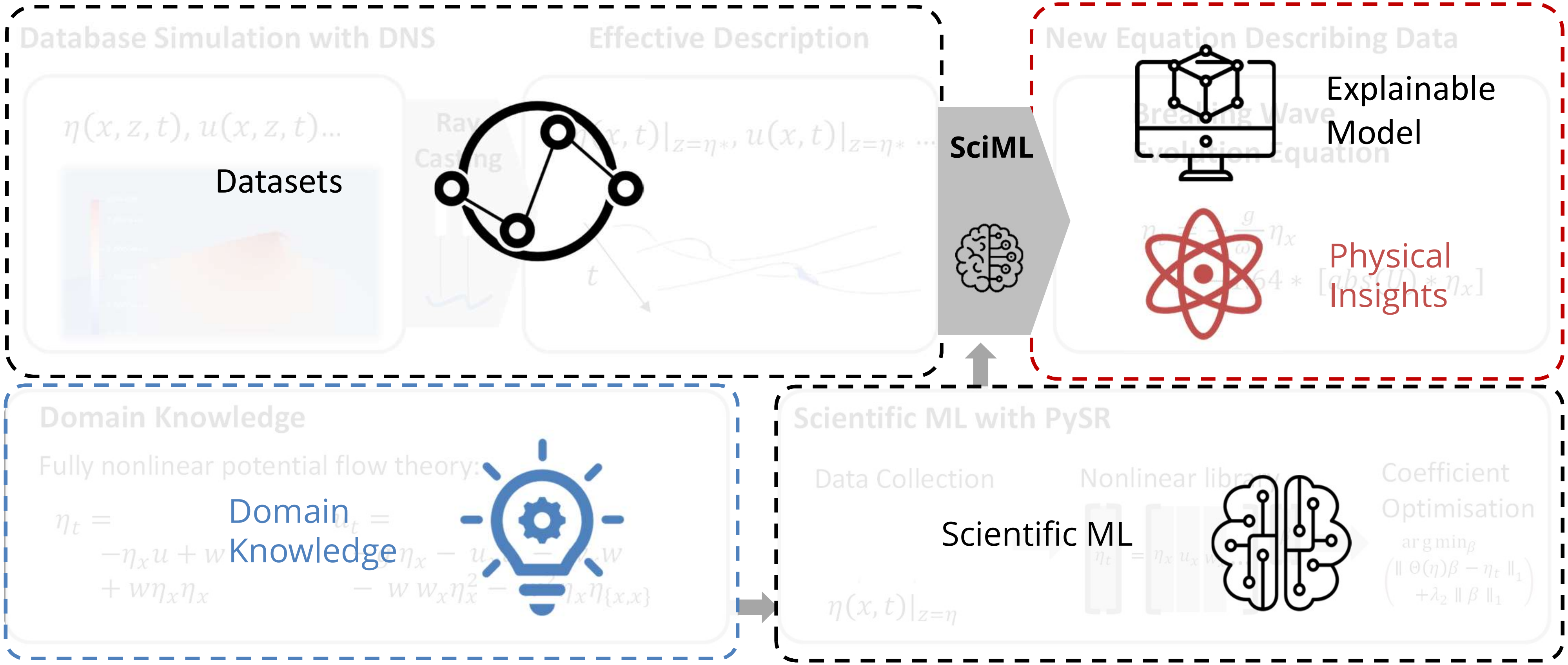
Nonlinear library

$$[\eta_t] = [\eta_x \ u_x \ w_x \ldots][\beta]$$

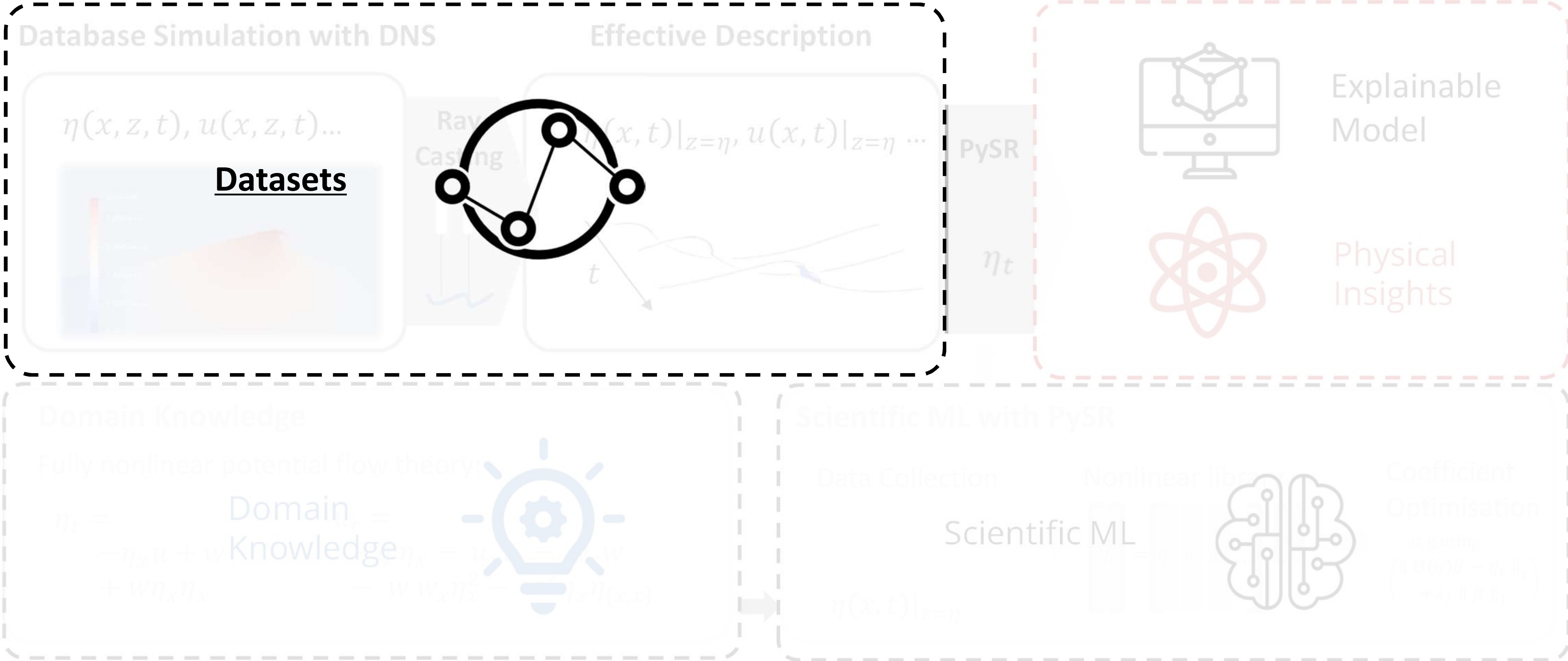
Coefficient Optimisation

$$\arg\min_\beta \left(\| \Theta(\eta)\beta - \eta_t \|_1 + \lambda_2 \| \beta \|_1 \right)$$

Machine Learning Approach



Machine Learning Approach



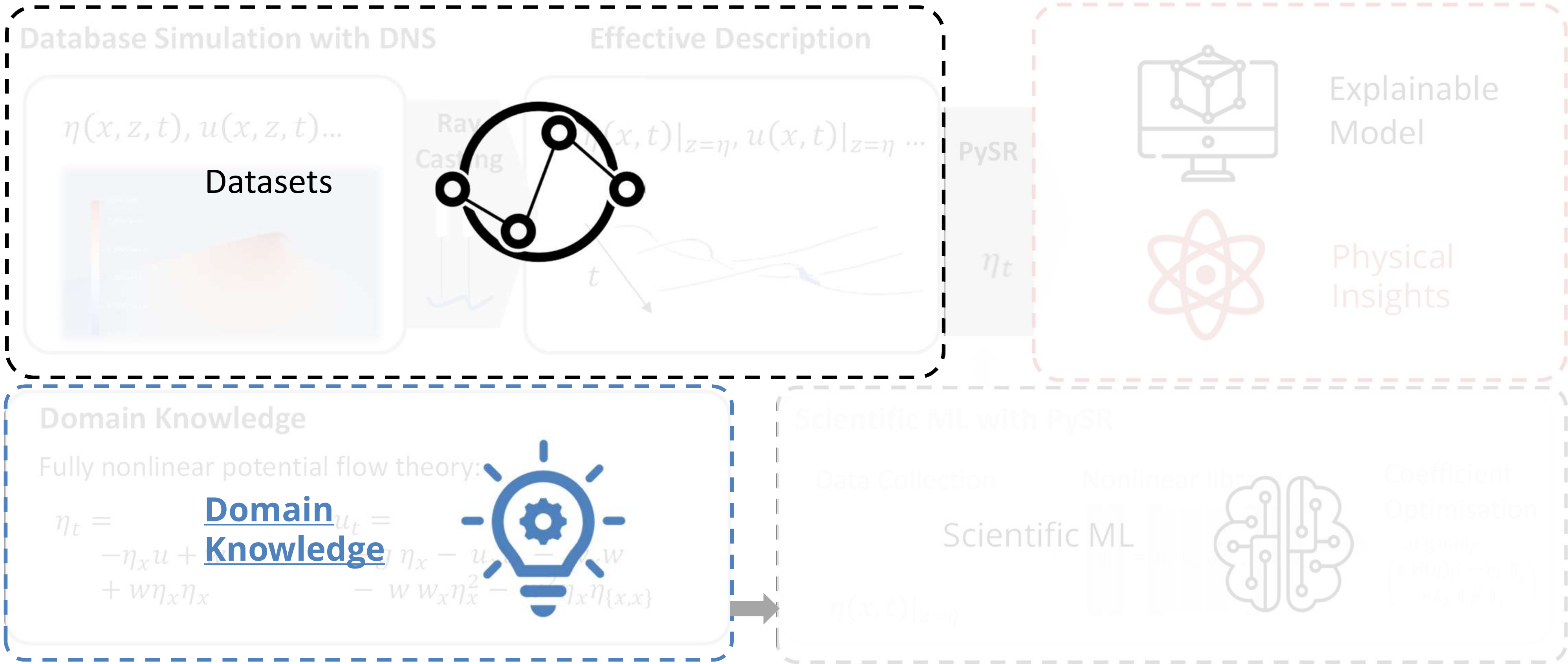
Datasets

Datasets

Over 75 2D breaking wave cases with over 1 million data points

A single case

Machine Learning Approach



Domain Knowledge

Non-breaking evolution

Fully Nonlinear Boundary Conditions (FNBC)

Time derivatives of surface elevation

Spatial derivatives of surface elevation and velocities

$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

η and u are *coupled* in FNBC framework

*Subscripts denote partial differentiation

Domain Knowledge

Non-breaking evolution

Wave breaking in progress

Fully Nonlinear Boundary Conditions (FNBC) → **Navier-Stokes Equation**

Time derivatives of surface elevation

Spatial derivatives of surface elevation and velocities

$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

$$\Downarrow$$

$$\boxed{\eta_t = \mathrm{FNBC}}$$

η and u are *coupled* in FNBC framework

Apply FNBC to breaking waves?

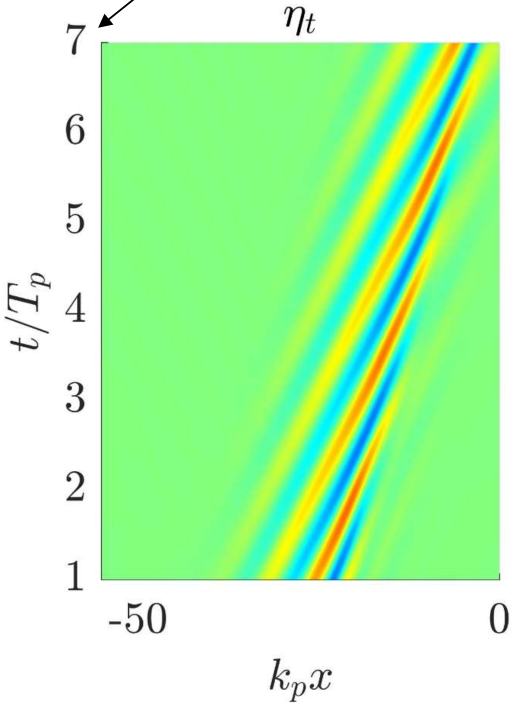
*Subscripts denote partial differentiation

Domain Knowledge

Breaking evolution

LHS:

Time axis

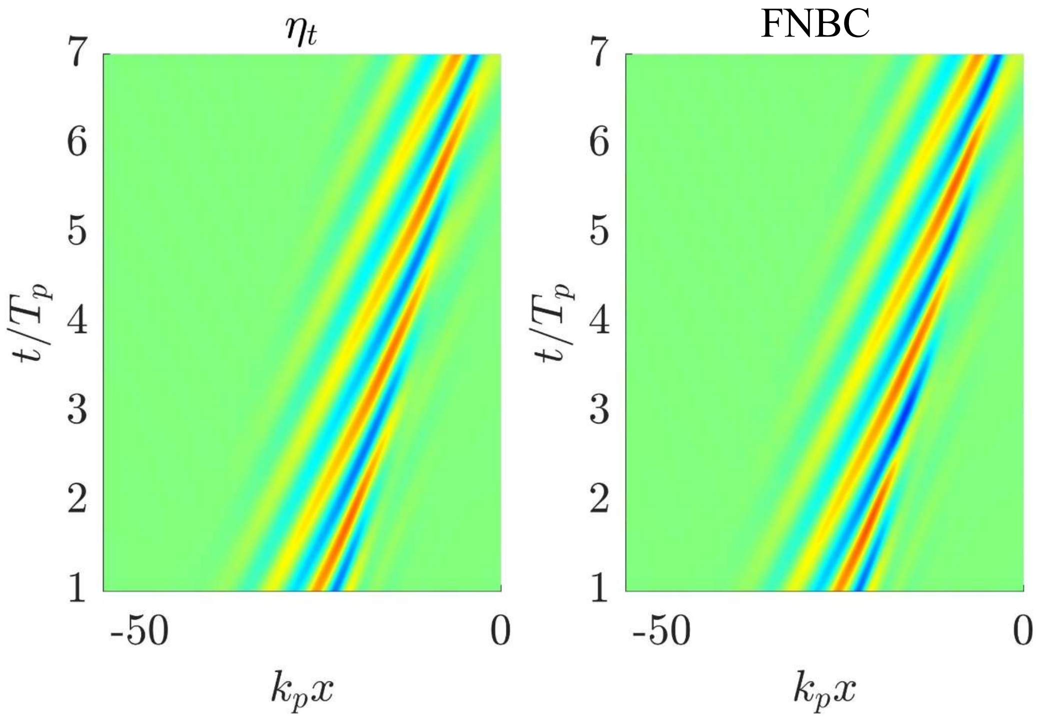


Space axis

Domain Knowledge

Breaking evolution

LHS: **RHS:**



Domain Knowledge

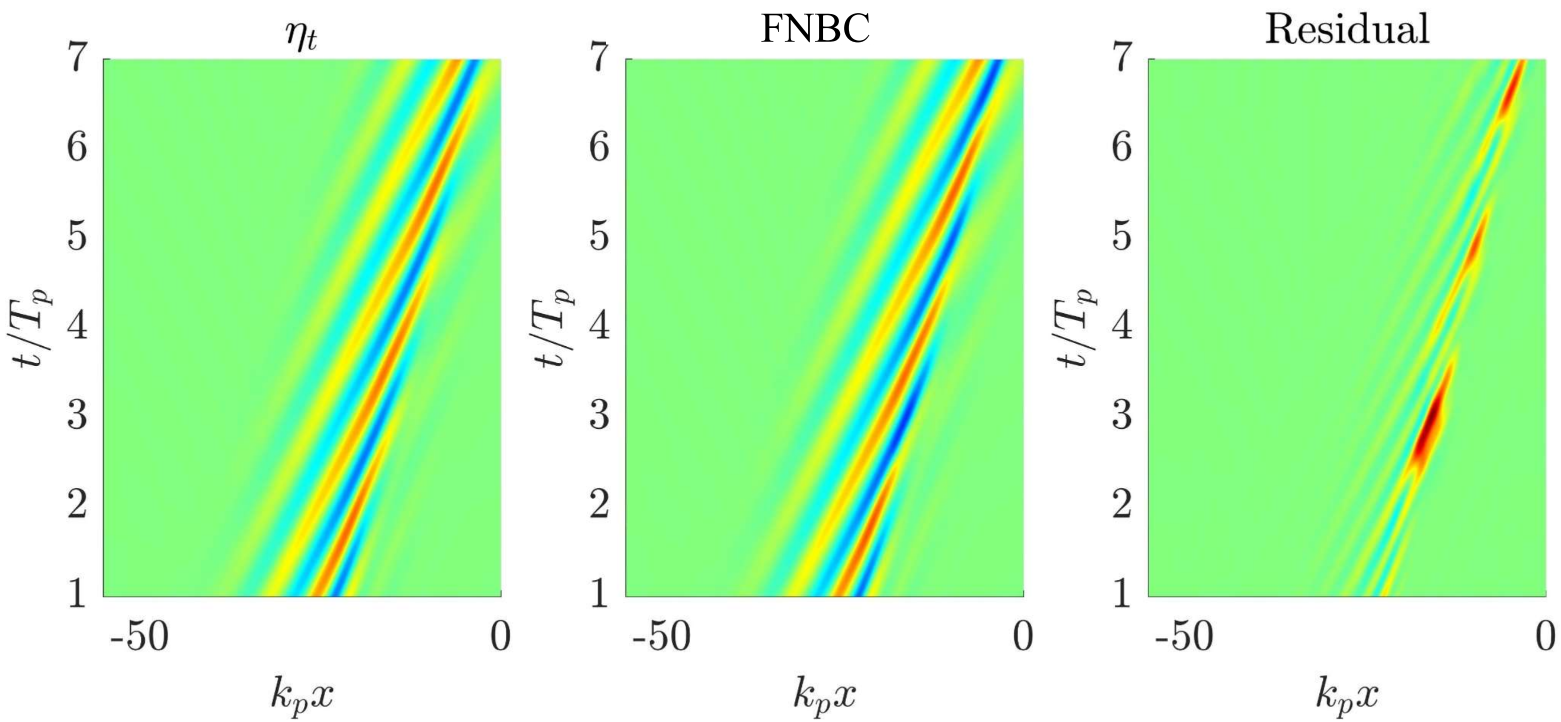
Breaking evolution

LHS:

RHS:

Residual = LHS - RHS

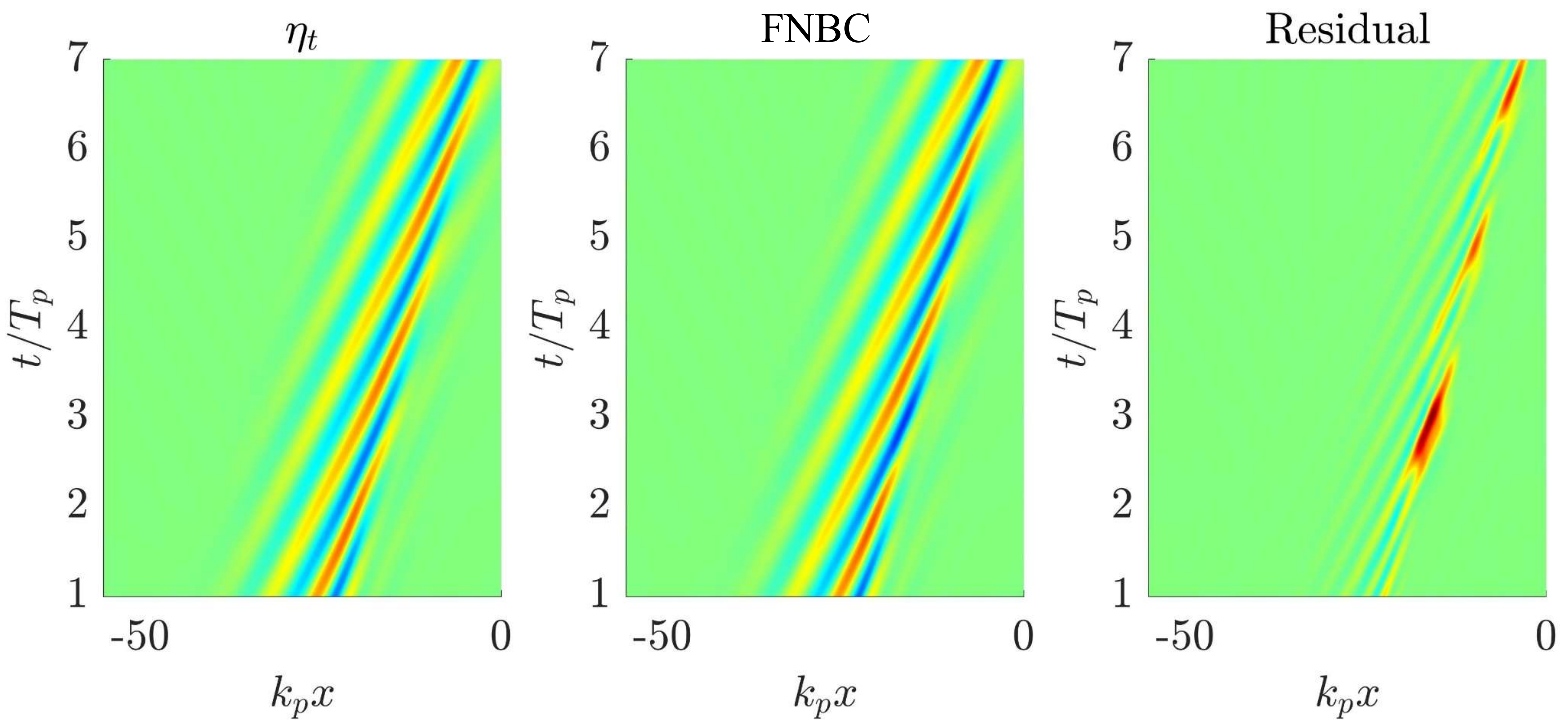
Residual: Difference between two sides of the equation



Domain Knowledge

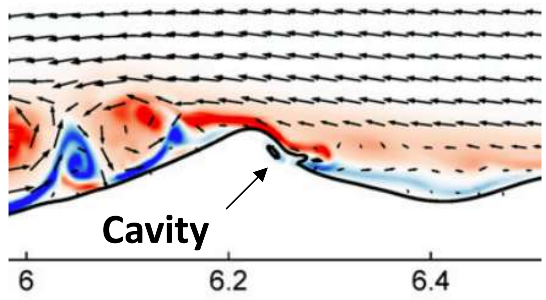
Breaking evolution

LHS: **RHS:** **Residual = LHS - RHS**

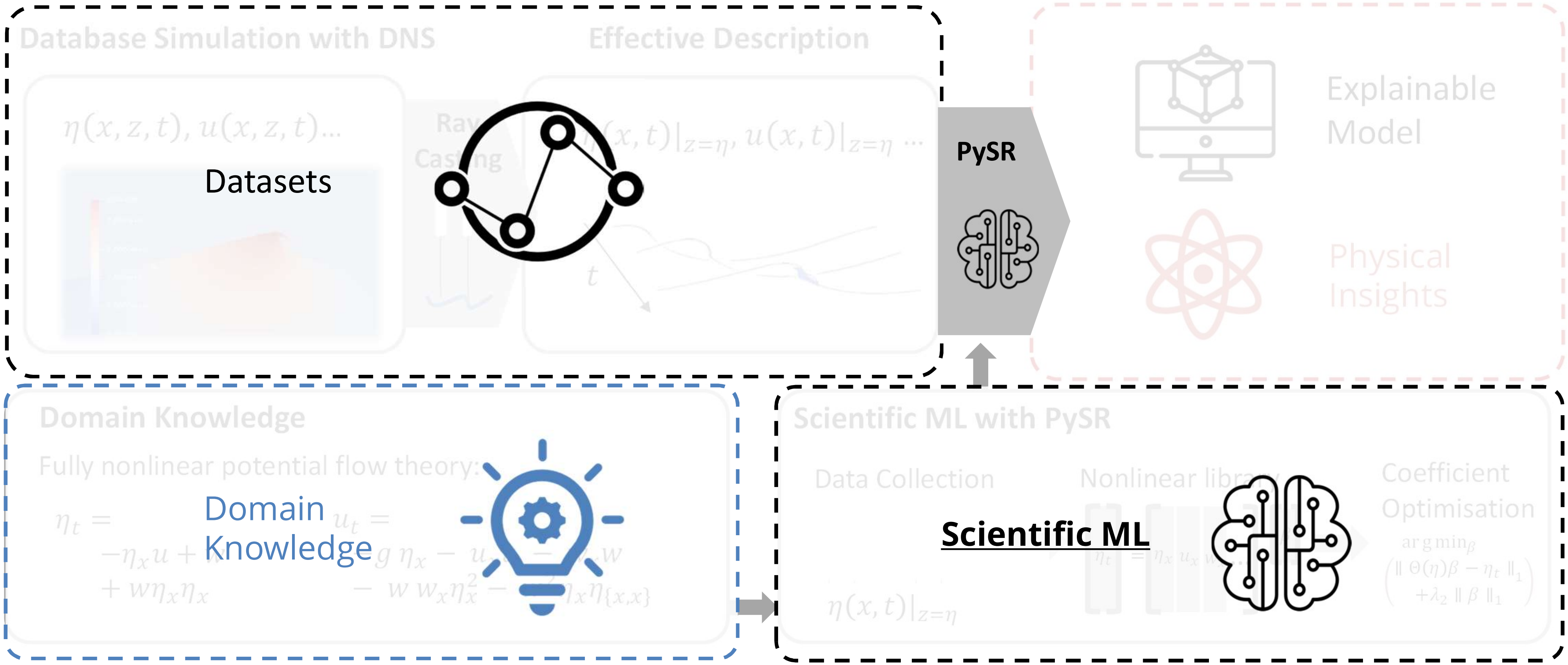


Residual: Difference between two sides of the equation

For breaking evolution, we observe **some deviation** from the FNBC framework at the breaking region.



Machine Learning Approach



Scientific ML

Non-breaking evolution | **Wave breaking in progress** | **Non-breaking evolution**

Fully Nonlinear Boundary Conditions (FNBC) → **Navier-Stokes Equation** → **Fully Nonlinear Boundary Conditions (FNBC)**

$$\boxed{\eta_t = \text{FNBC}}$$

$$\eta_t = N(\eta, \eta_x, u, u_x, w \cdots, \mu)$$

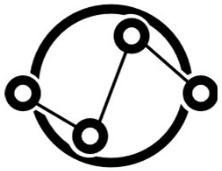
$$\boxed{\eta_t = \text{FNBC}}$$

The deviation from FNBC should be small

$$\boxed{\eta_t = \text{FNBC} + N(\eta, \eta_x, u, u_x, w \cdots, \mu)}$$

Scientific ML with Symbolic Regression

Datasets



Nonlinear library

$$[\eta_t] = [\eta_x \; u_x \; w_x \cdots][\beta]$$

Coefficient Optimisation

$$\arg\min_\beta \begin{pmatrix} \| \Theta(\eta)\beta - \eta_t \|_1 \\ + \lambda_2 \| \beta \|_1 \end{pmatrix}$$

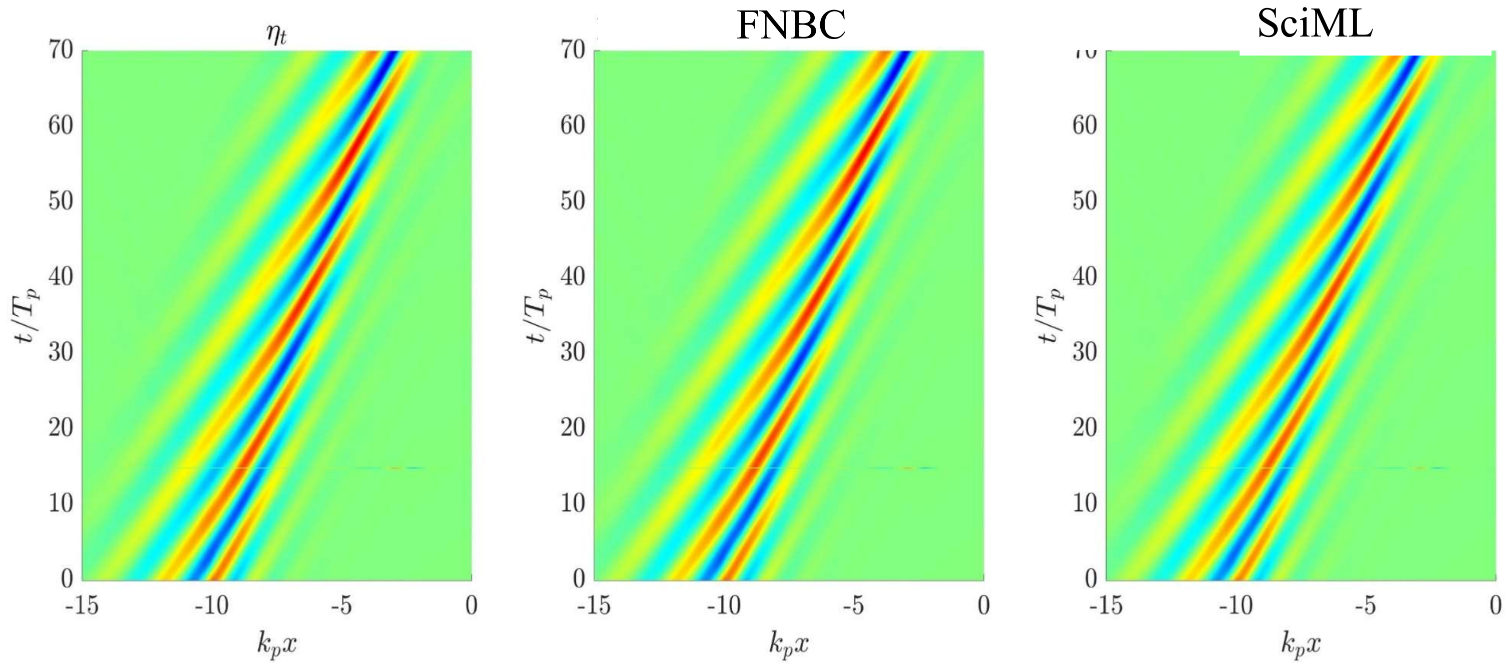
Governing Boundary Equation for Breaking Waves

$$\eta_t = \underbrace{\text{FNBC}}_{\text{Domain Knowledge}} + \underbrace{N(\eta, \eta_x, u, u_x, w \cdots, \mu)}_{\text{Breaking Correction}}$$

Domain Knowledge + Breaking Correction

Results

Non-breaking evolution



Domain Knowledge
from math derivation

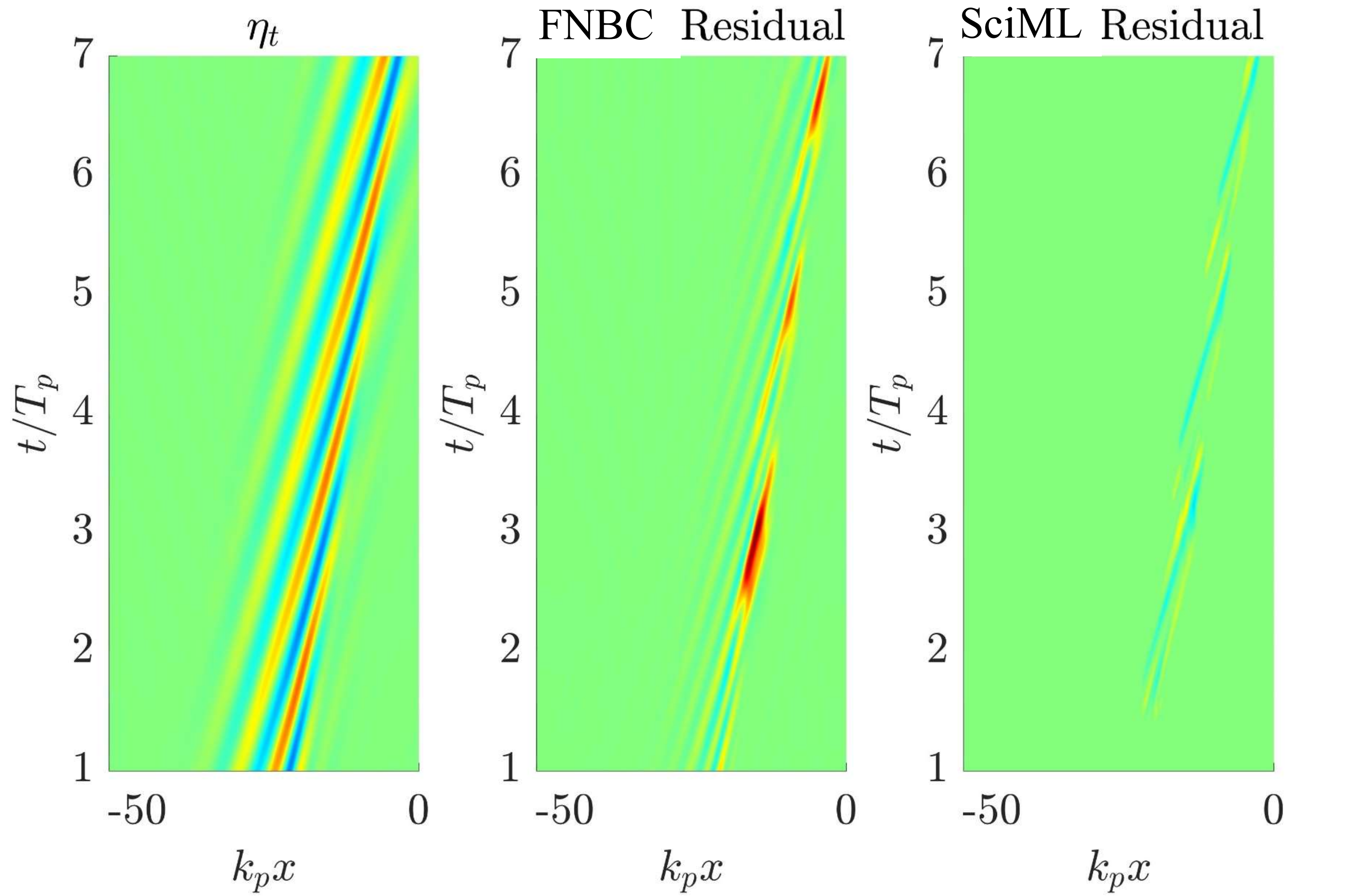
$$-1\eta_x u + 1w + O(3)$$

SciML

$$-1.058\eta_x u + 0.98w + O(3)$$

Scientific ML
discovered equations
from data

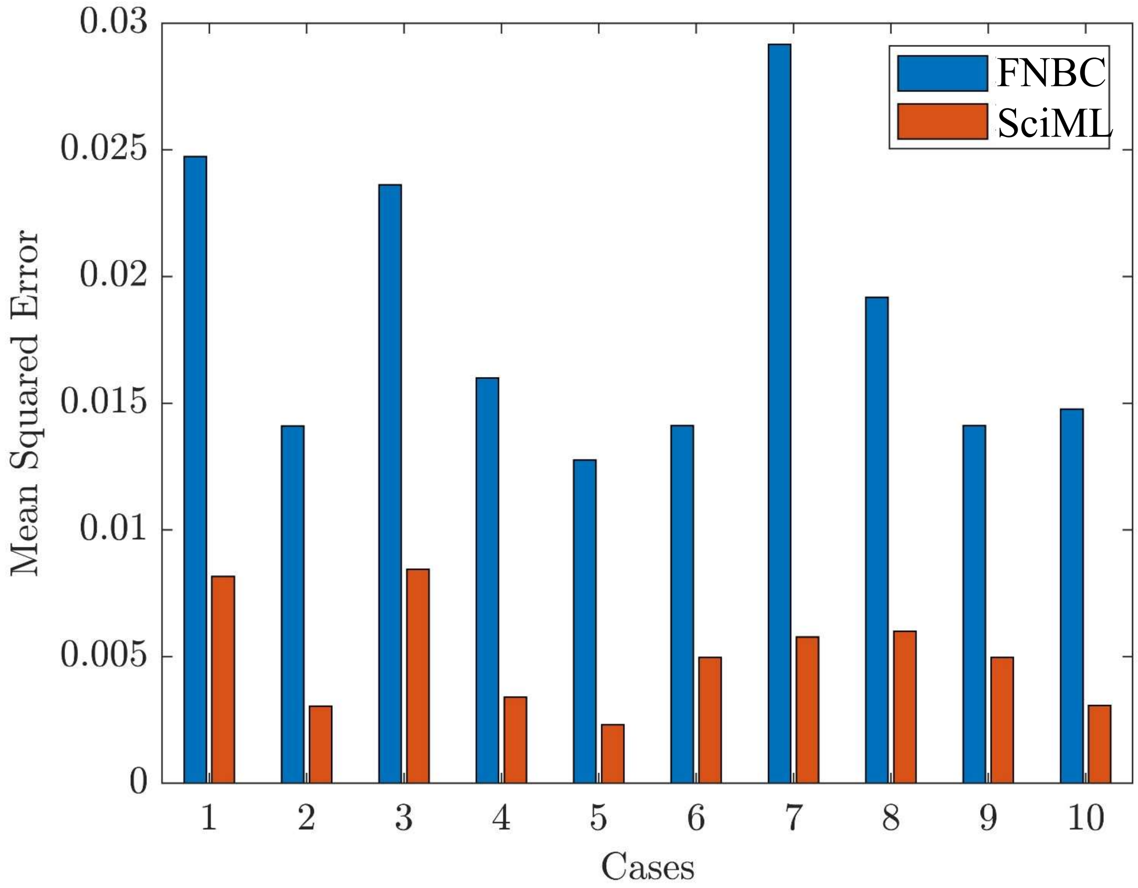
Results



Breaking evolution

Based on the MSE error calculation, the new SciML discovered formulation can reduce over 91% of the residual!

Results – Test dataset



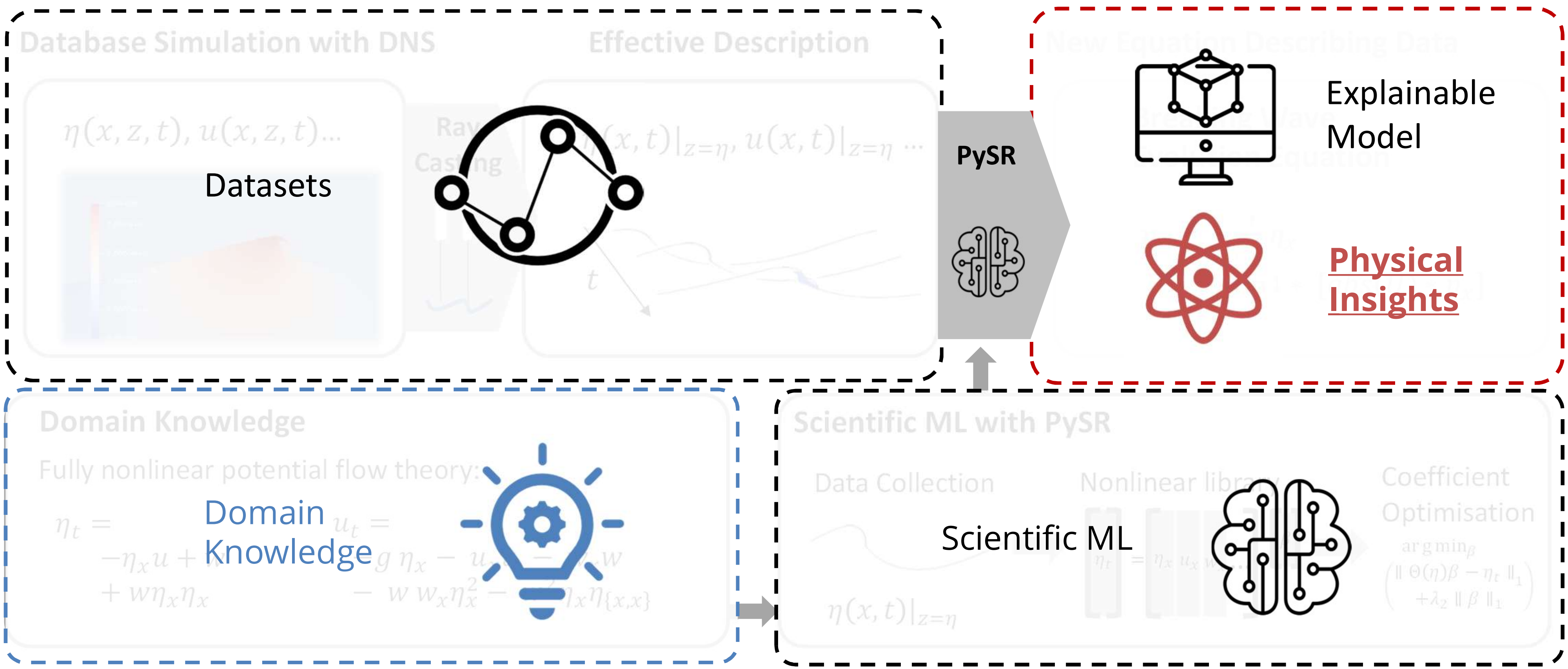
$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2.$$

DNS observed value

SciML/FNBC predicted value

The test dataset also shows significantly reduction in the MSE.

Machine Learning Approach



Why the SciML Discovered Equation Works?

Physical insights

Non-breaking evolution

FNBC framework:

$$\eta_t = -1\eta_x u + 1w + O(3)$$

SciML discovered equation:

$$\eta_t = -1.058\eta_x u + 0.98w + O(3)$$

Breaking evolution

FNBC framework:

$$\eta_t = -1\eta_x u + 1w + O(3)$$

SciML discovered equation:

$$\eta_t = -\frac{g}{\omega_p}\eta_x + 1.6\, abs(U)\eta_x + O(3)$$

Residual is reduced significantly

Physical insights

Non-breaking evolution

FNBC framework:

$$\eta_t = -1\eta_x u + 1w + O(3)$$

SciML discovered equation:

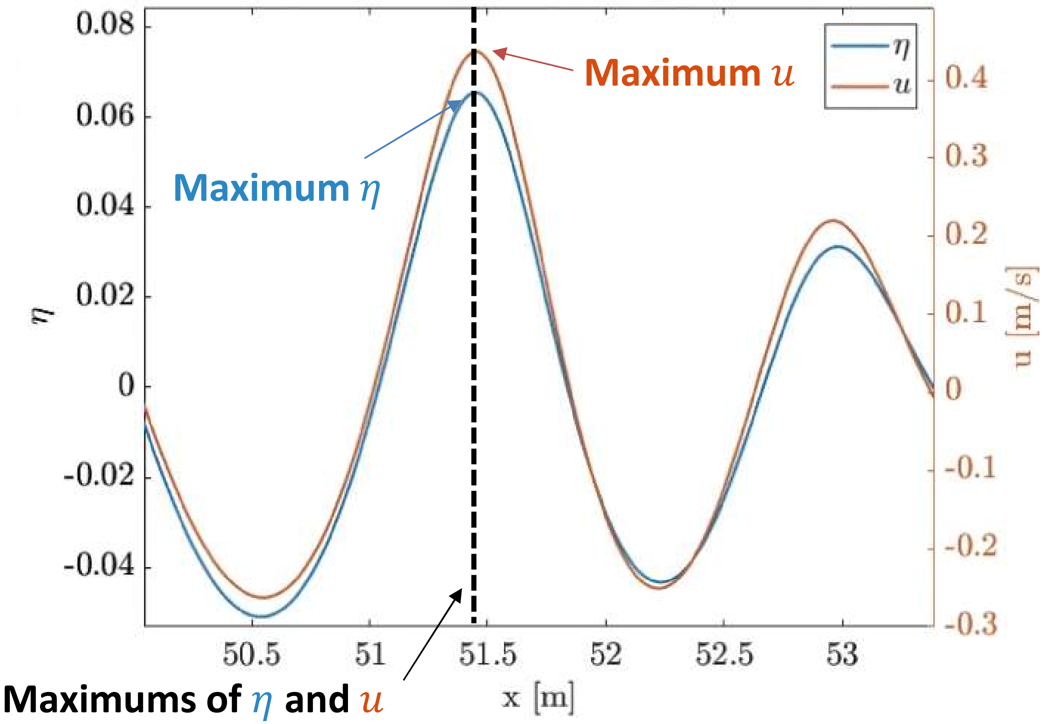
$$\eta_t = -1.058\eta_x u + 0.98w + O(3)$$

Terms with surface elevation

Terms with velocities

FNBC & SciML: ✓

$\eta_t \leftarrow u$ terms & η terms



Maximums of η and u are aligned

Physical insights

DEPARTMENT OF ENGINEERING SCIENCE
UNIVERSITY OF OXFORD

Breaking evolution

Why the SciML uses surface elevation terms (η) **only** ?

FNBC framework:

$$\eta_t = 1w - 1\eta_x u + O(3)$$

SciML discovered equation:

$$\eta_t = -\frac{g}{\omega_p}\eta_x + 1.6\, abs(U)\eta_x + O(3)$$

Terms with surface elevation (η) **only**

FNBC:

$\eta_t \leftarrow$ u terms & η terms

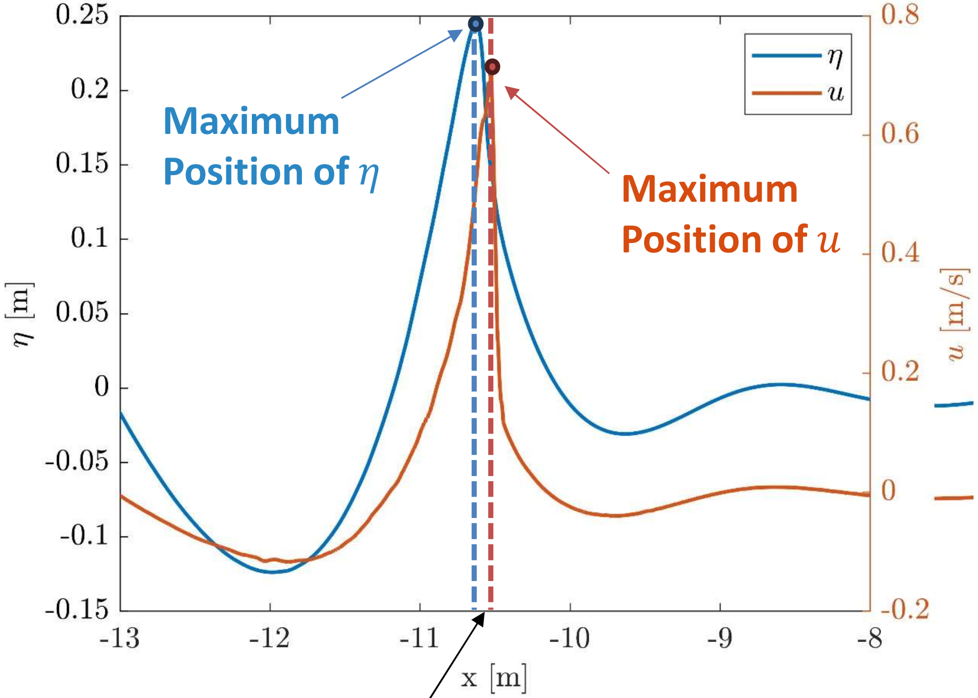
SciML :

$\eta_t \leftarrow$ η terms

$\neq$

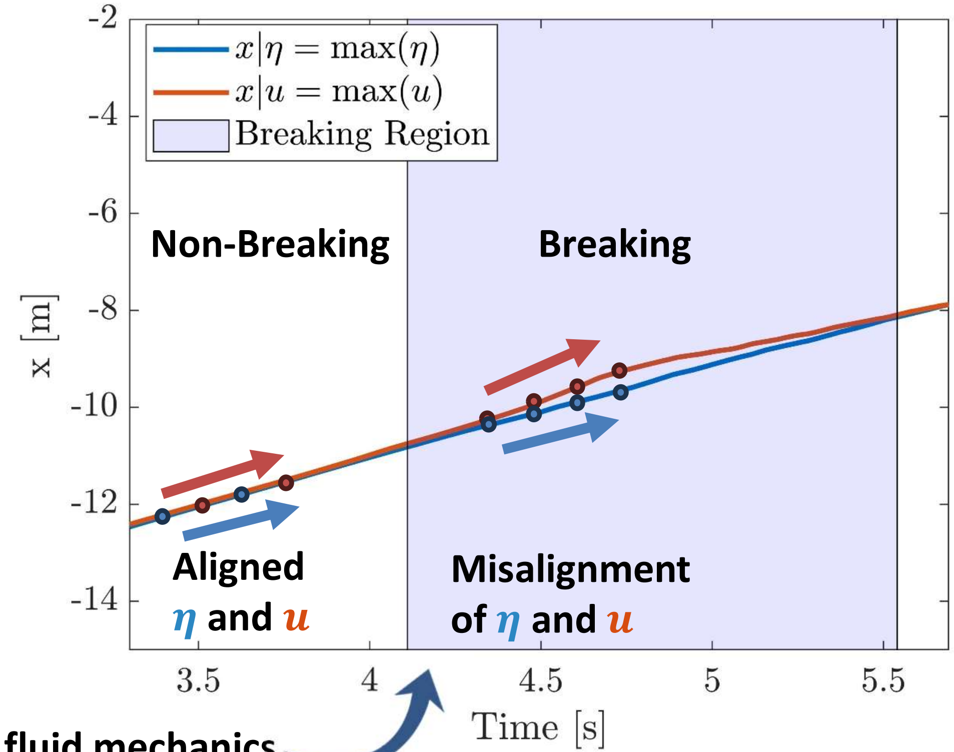
Physical insights

Maximum η and u



Maximums of η and u are NOT aligned

Evolution of the maximum position η and u



New findings in fluid mechanics

Simulation – in progress

New Equation Describing Data

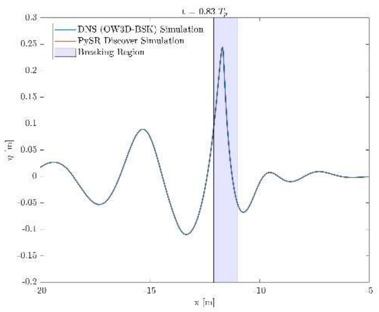
Breaking Wave Evolution Equation

$$\eta_t = -\frac{g}{\omega_p}\eta_x - 3.64 * \; [abs(B) * \eta_x]$$

RK-4 Integration

t

Reduce-order Simulation for breaking region



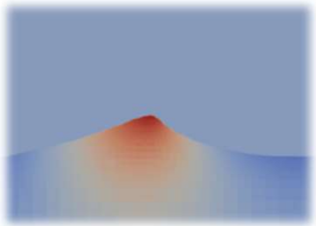
Domain Knowledge

Fully nonlinear boundary conditions:

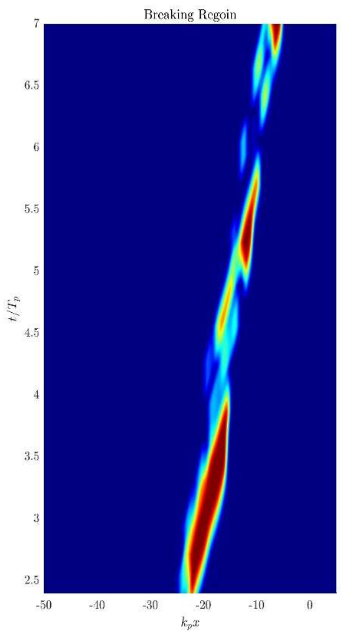
$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

$$u_t = -g\,\eta_x - u_x u - w_x w - w\,w_x\eta_x^2 - w^2\eta_x\eta_{\{x,x\}}$$

Non-Breaking



Space-time Breaking Region Map



Model

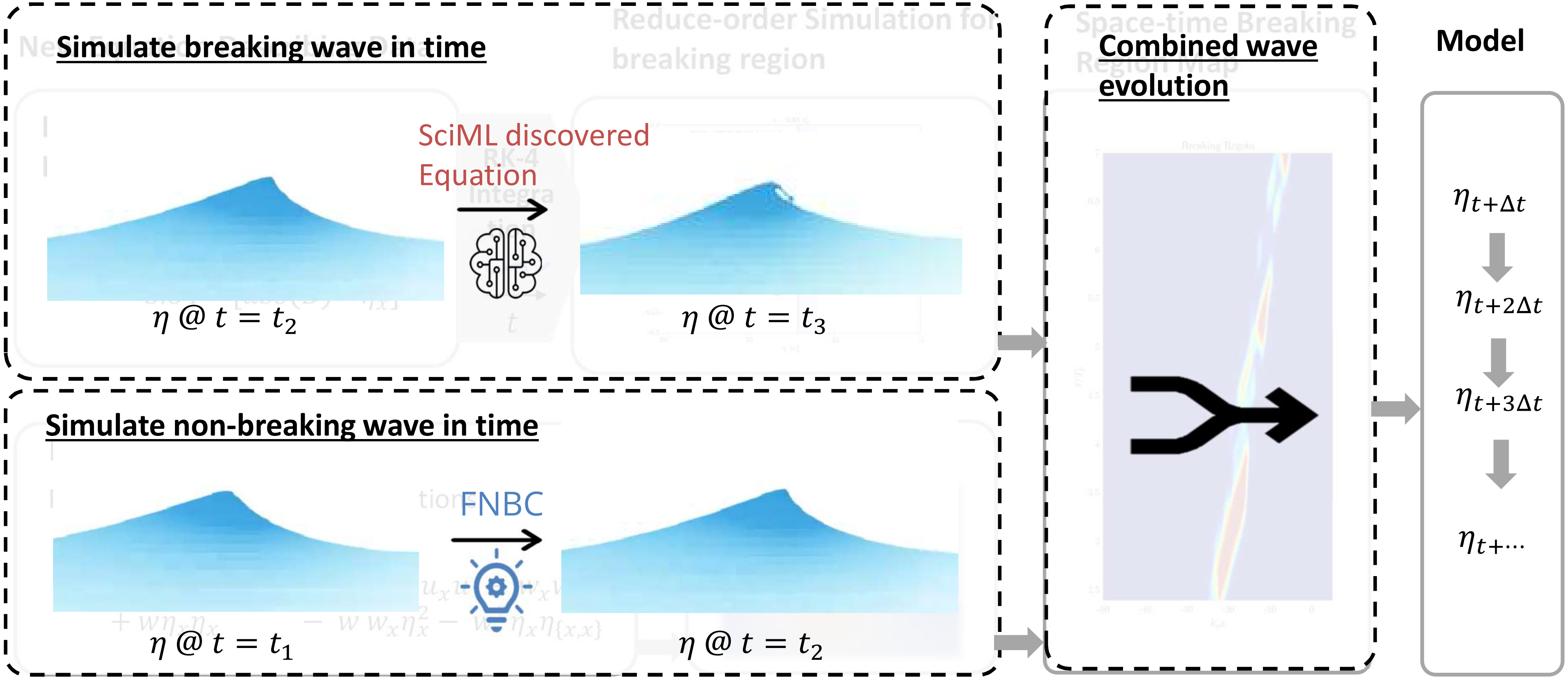
$\eta_{t+\Delta t}$

$\eta_{t+2\Delta t}$

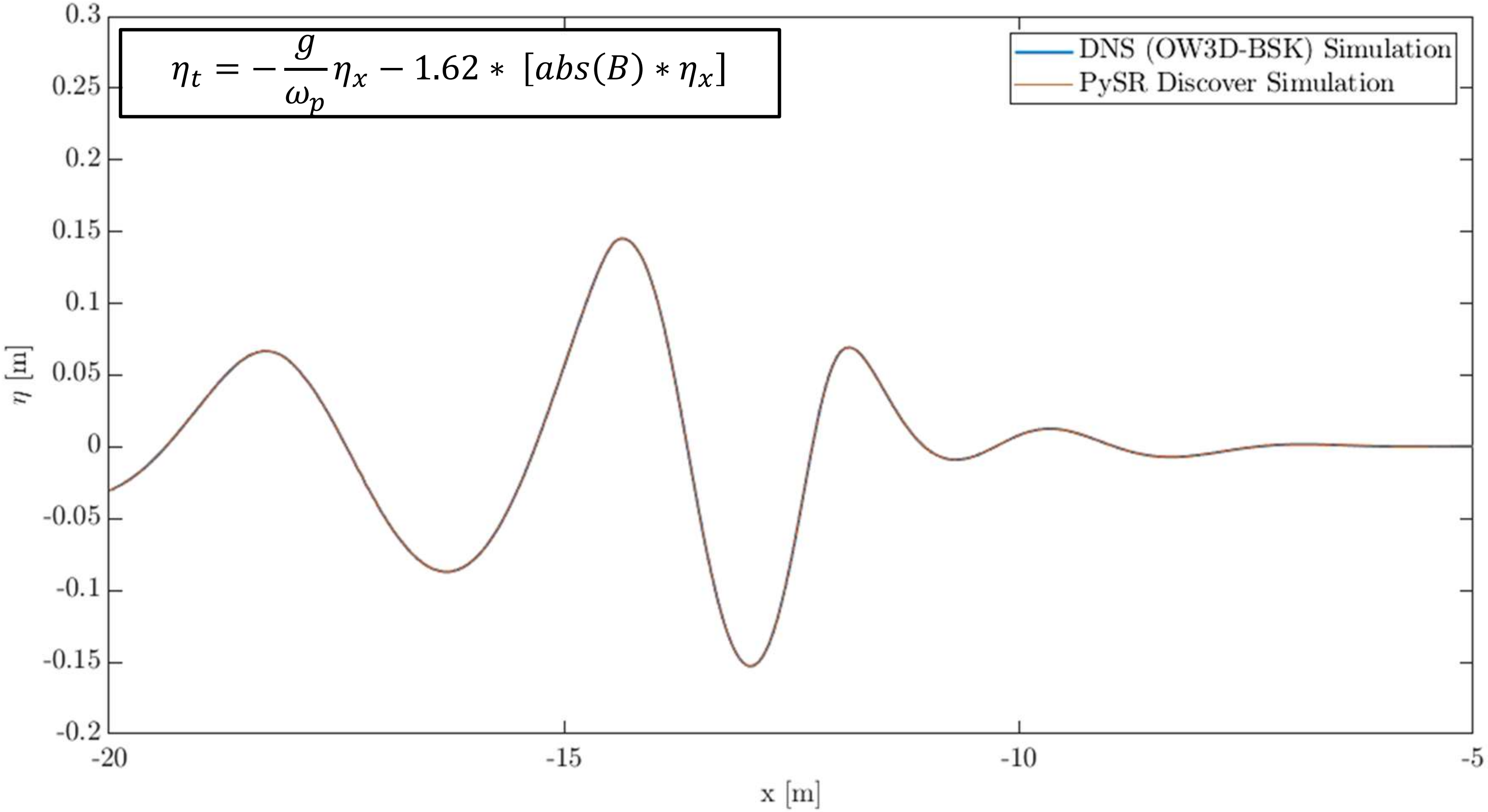
$\eta_{t+3\Delta t}$

$\eta_{t+\cdots}$

Simulation – in progress



Preliminary Results



** Only for 2d deep water spilling breakers so far*

We aim to develop a new model discovered by ML (in-progress) that:

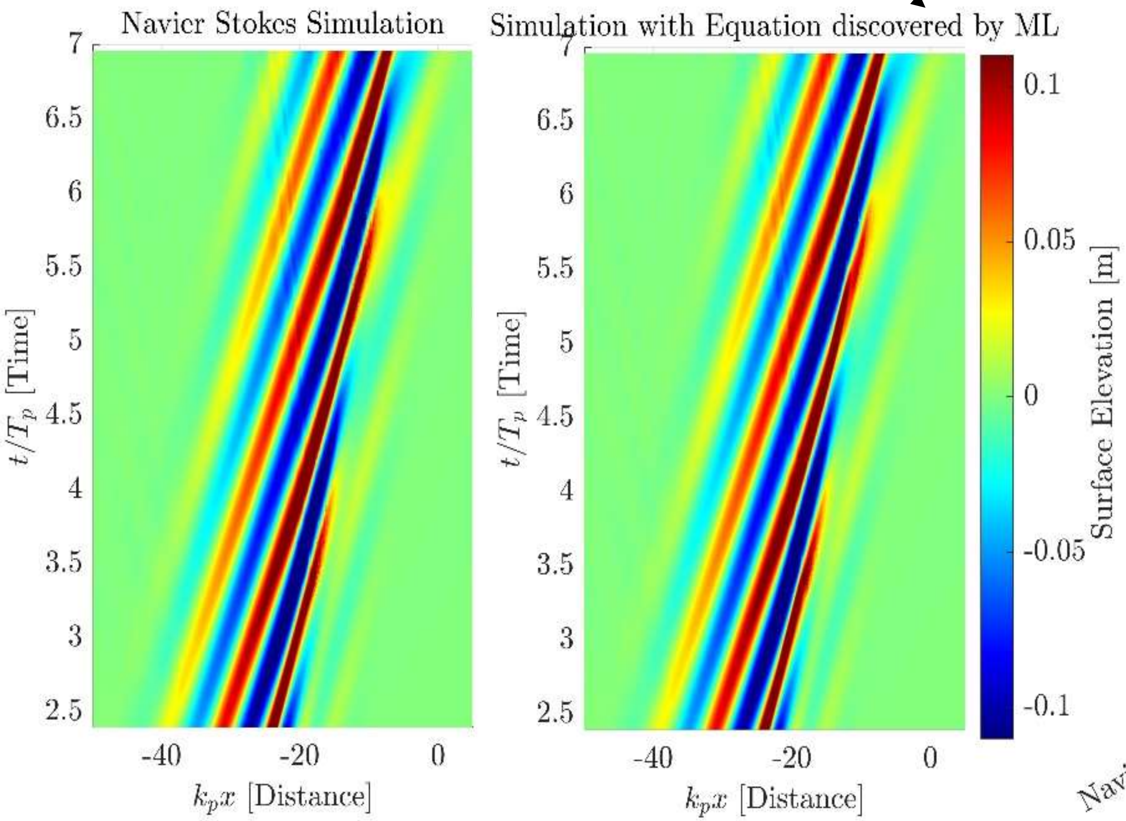
- Overlooks bubbles and white cap details
- Equation based **numerical simulation** (*white box*)
- Very **Fast** (*2 minutes* on desktop vs *3250* of core *hours* on supercomputer)
- Mathematically interpretable
- Directly applicable to various scales of the wave

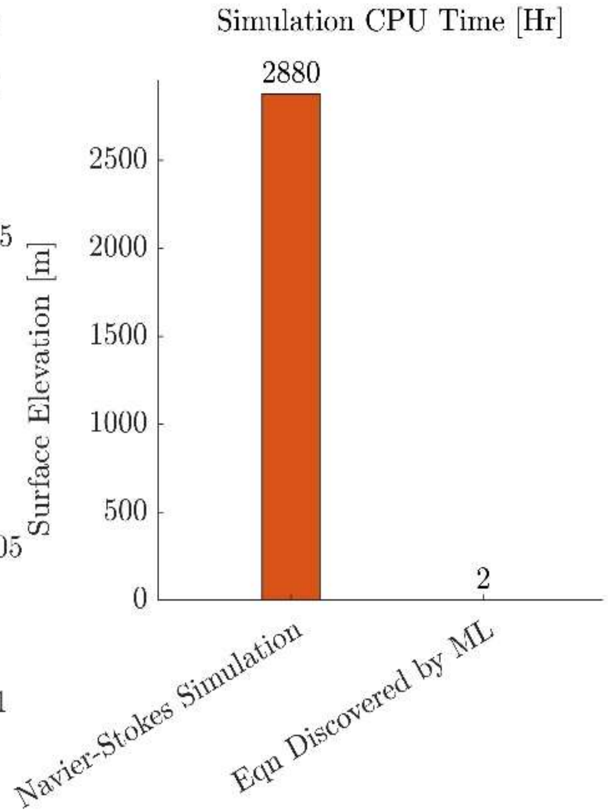
Conclusion

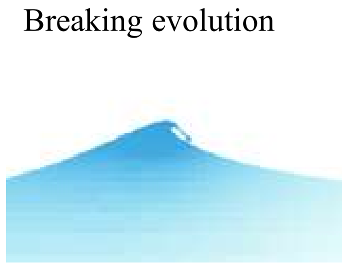
Accurate approximation of breaking wave with Equation discovered by SciML

Significant reduction in modelling time

New physical insights







Misalignment of η and u

Physical Insights

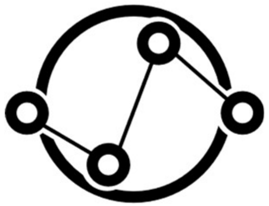
Thank you!

Tianning Tang (Tim);
Email: tianning.tang@eng.ox.ac.uk

Can we now do better with lagre amount of data?

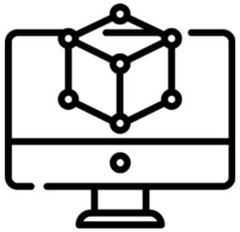
Machine Learning

Datasets

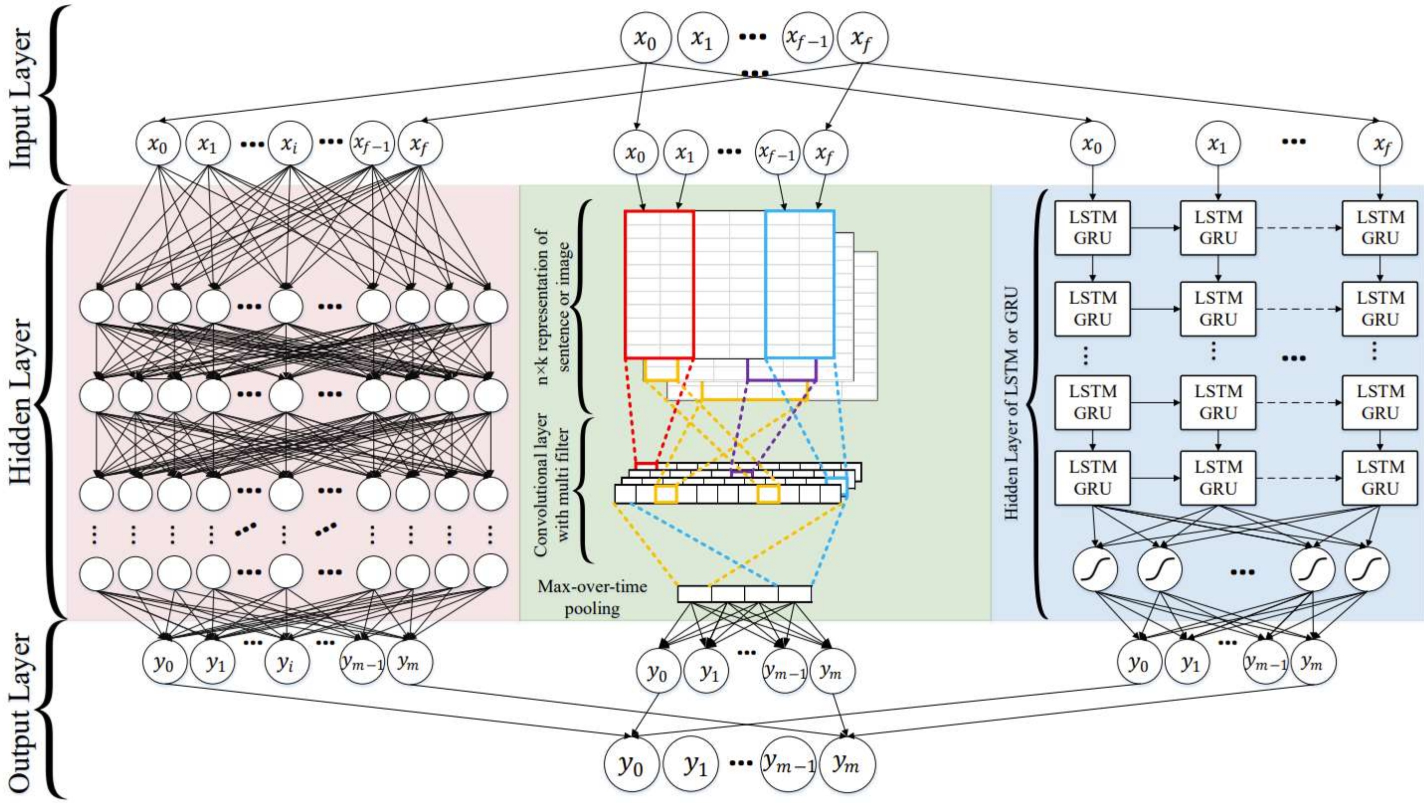


Machine Learning

Empirical Model

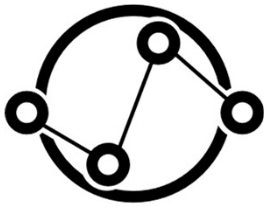


RMDL: Random Multimodel Deep Learning for Classification



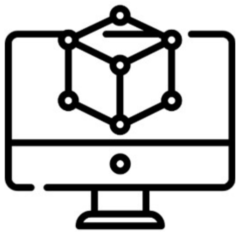
Machine Learning

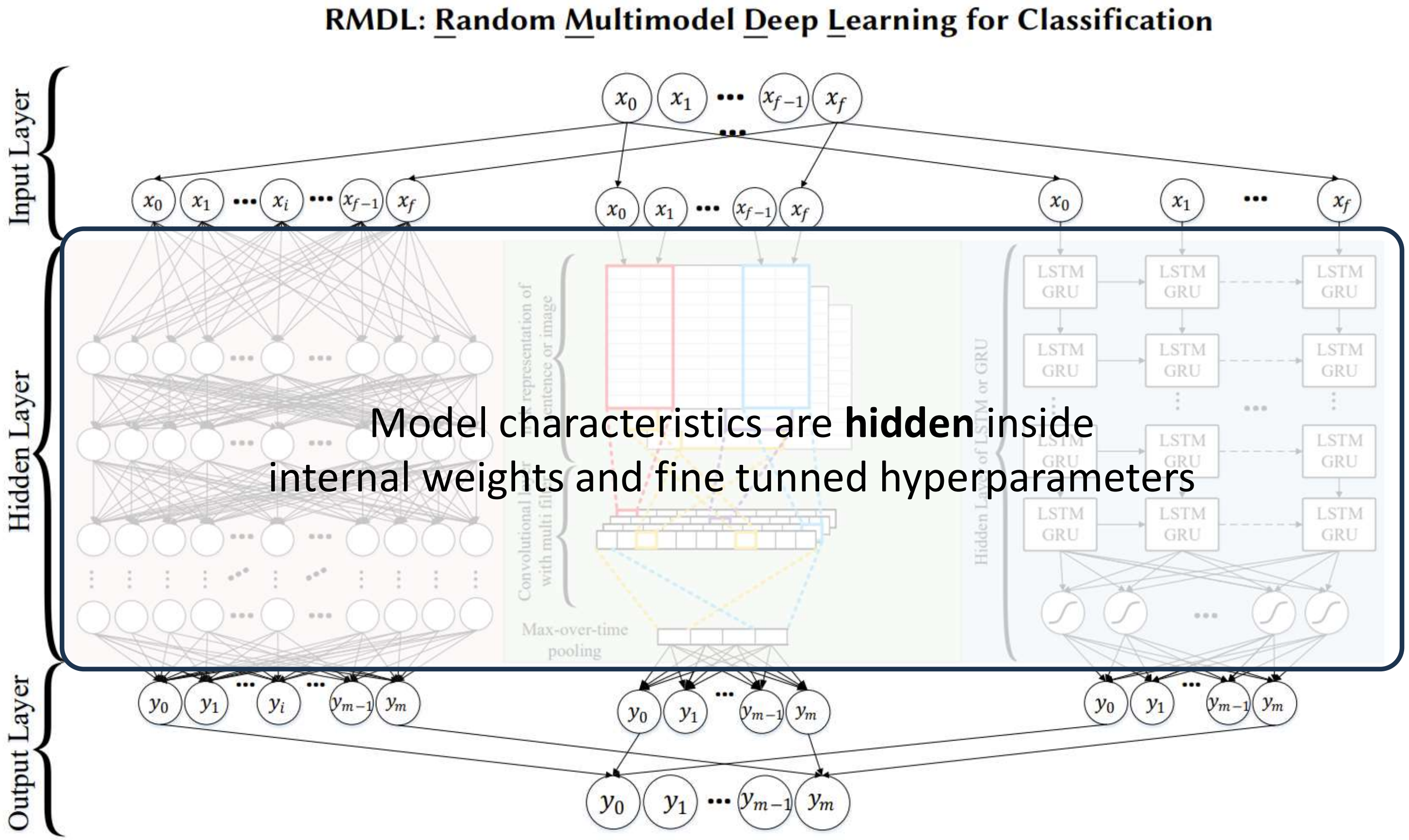
Datasets



Machine
Learning

Empirical
Model





Model characteristics are **hidden** inside
internal weights and fine tunned hyperparameters

Domain Knowledge

Non-breaking evolution

Fully Nonlinear Boundary Conditions (FNBC)

Time derivatives of surface elevation

Spatial derivatives of surface elevation and velocities

$$\eta_t = -\eta_x u + w + w\eta_x\eta_x$$

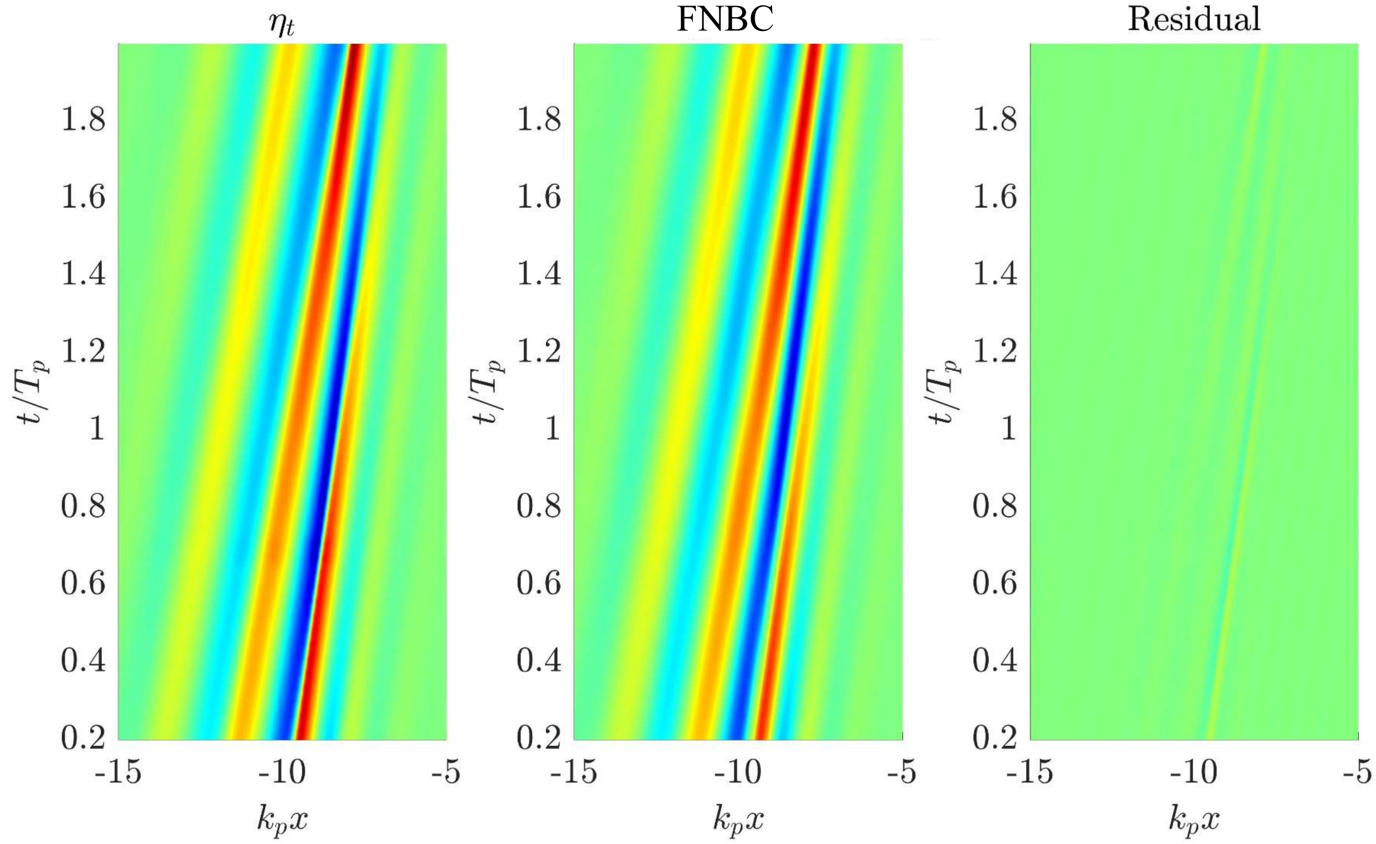
$$\boxed{\eta_t = \mathrm{FNBC}}$$

η and u are *coupled* in FNBC framework

*Subscripts denote partial differentiation

Domain Knowledge

LHS: **RHS:** **Res = LHS - RHS**



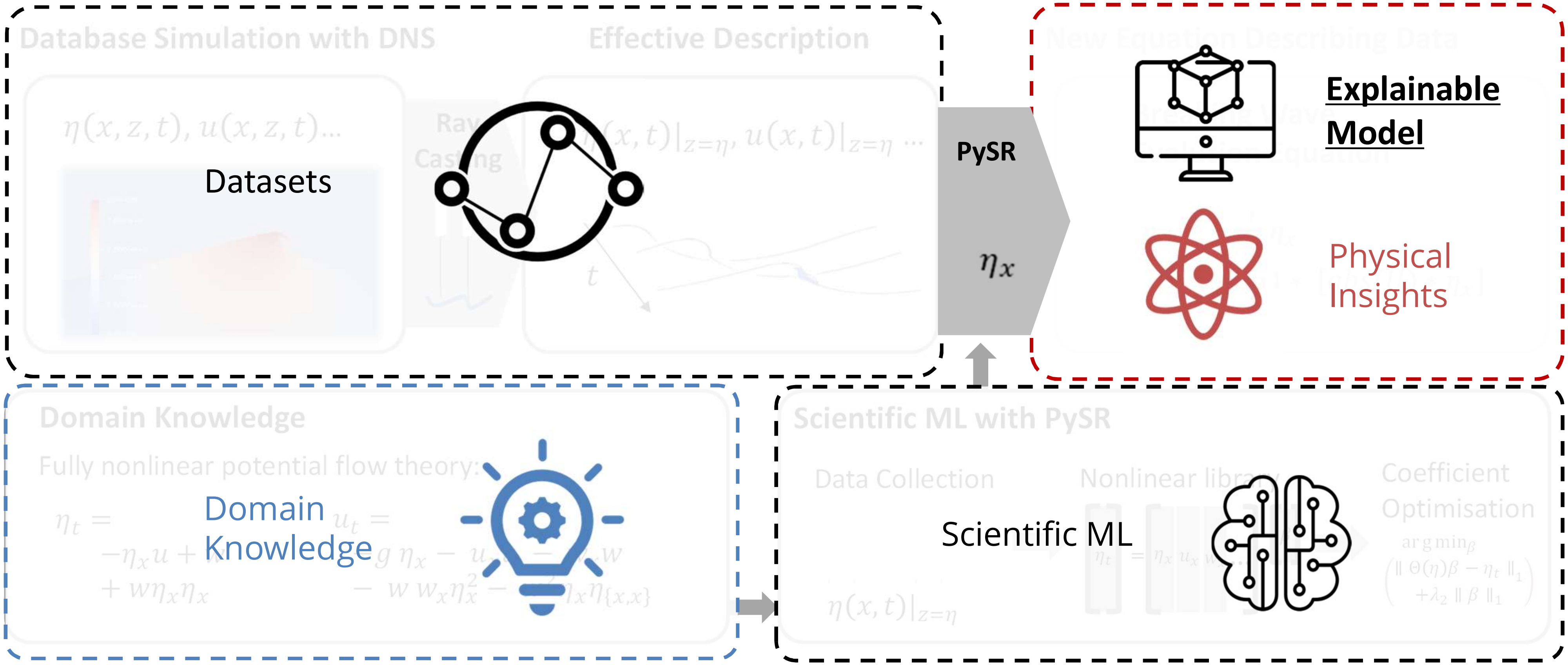
Non-breaking evolution

$$\boxed{\eta_t = \mathrm{FNBC}}$$

As expected, for non-breaking evolution, we have less than 1% of difference between the LHS and RHS of the equation.

FNBC equation **works well** for non-breaking evolution.

Machine Learning Approach

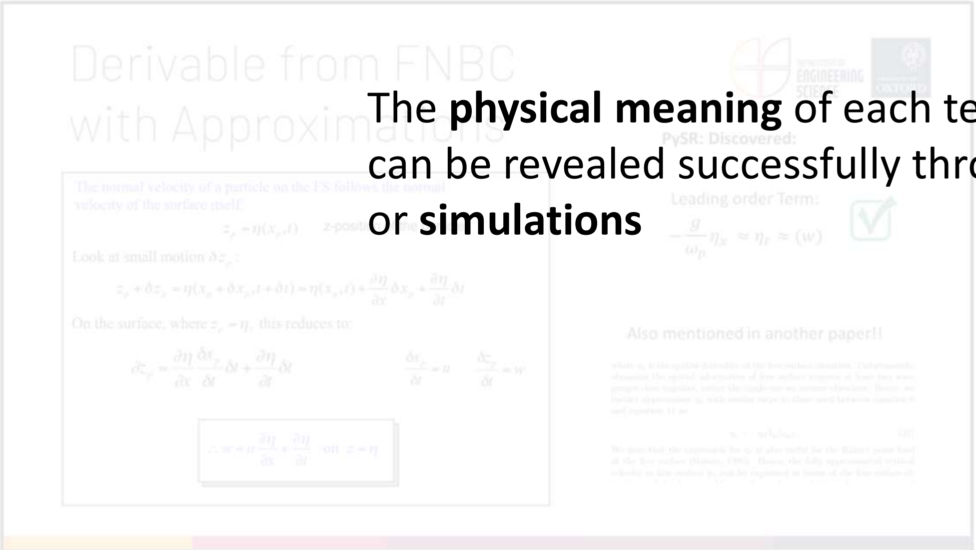


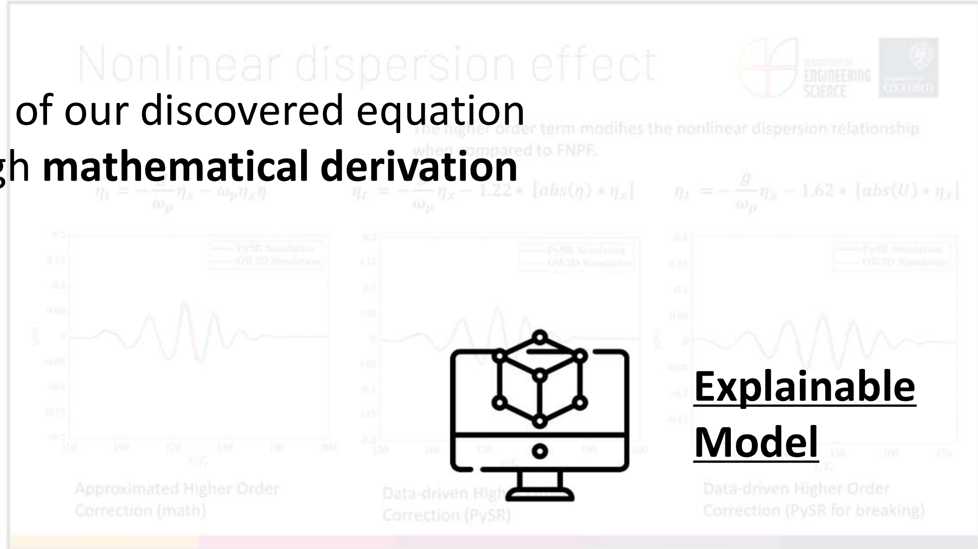
PySR new equation

Discovered boundary condition equation:

Envelope of η

$$\eta_t = -\frac{g}{\omega_p}\eta_x + [-1.6 * abs(B) * \eta_x] + O(3)$$





The **physical meaning** of each term of our discovered equation can be revealed successfully through **mathematical derivation** or **simulations**

Explainable Model