Seminar Presentation

-Vindhya Singh
(M.Sc Informatics)
TUM

Hidden Physics Models: Machine Learning of Non-Linear Partial Differential Equations

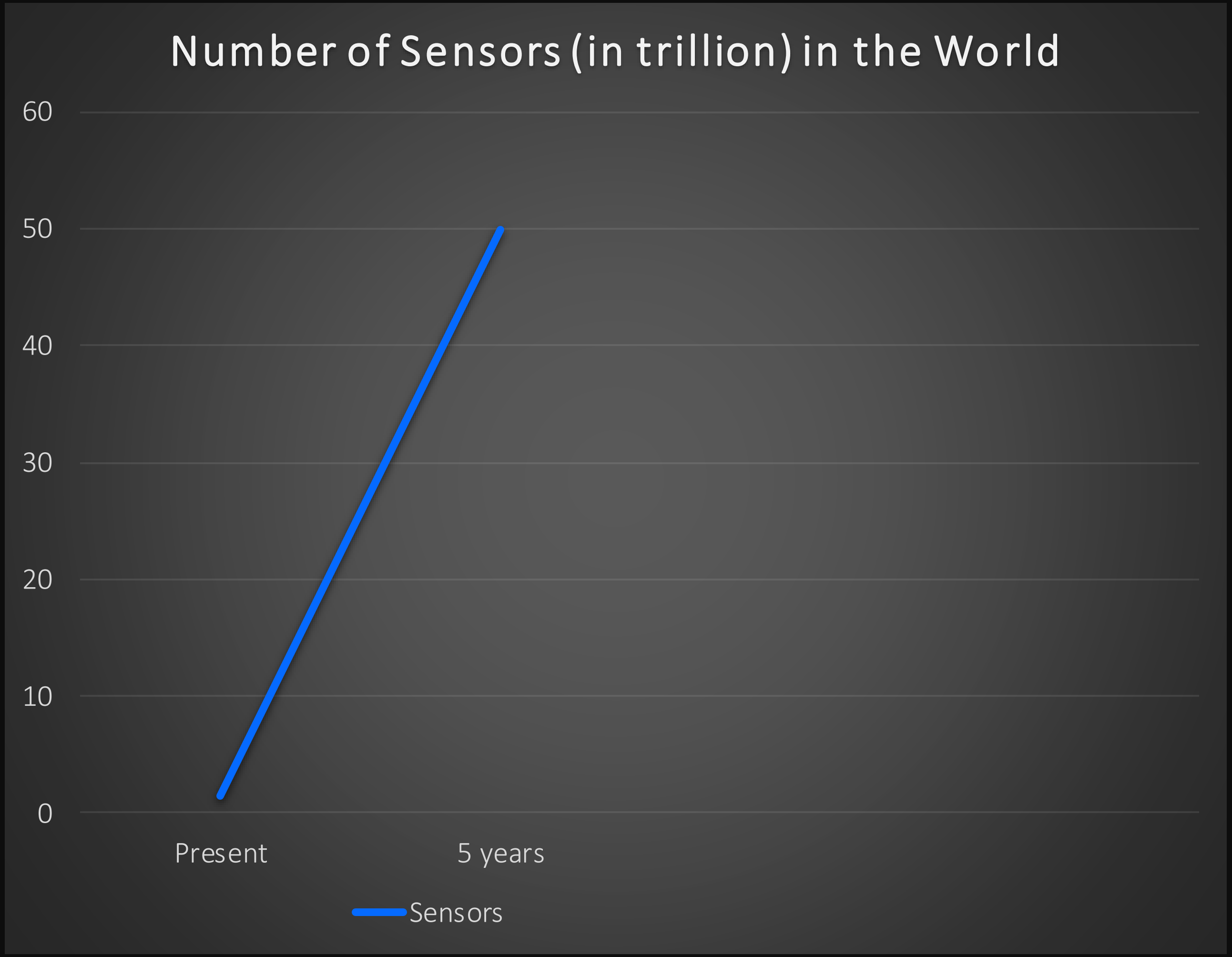
Deep Learning in Physics
Lecturer: Kiwon Um, Philipp Holl, and Nils Thuerey

Original Work by: Maziar Raissi, George Em Karniadakis
Hidden Physics Models: Machine Learning of Nonlinear Partial Differential Equations

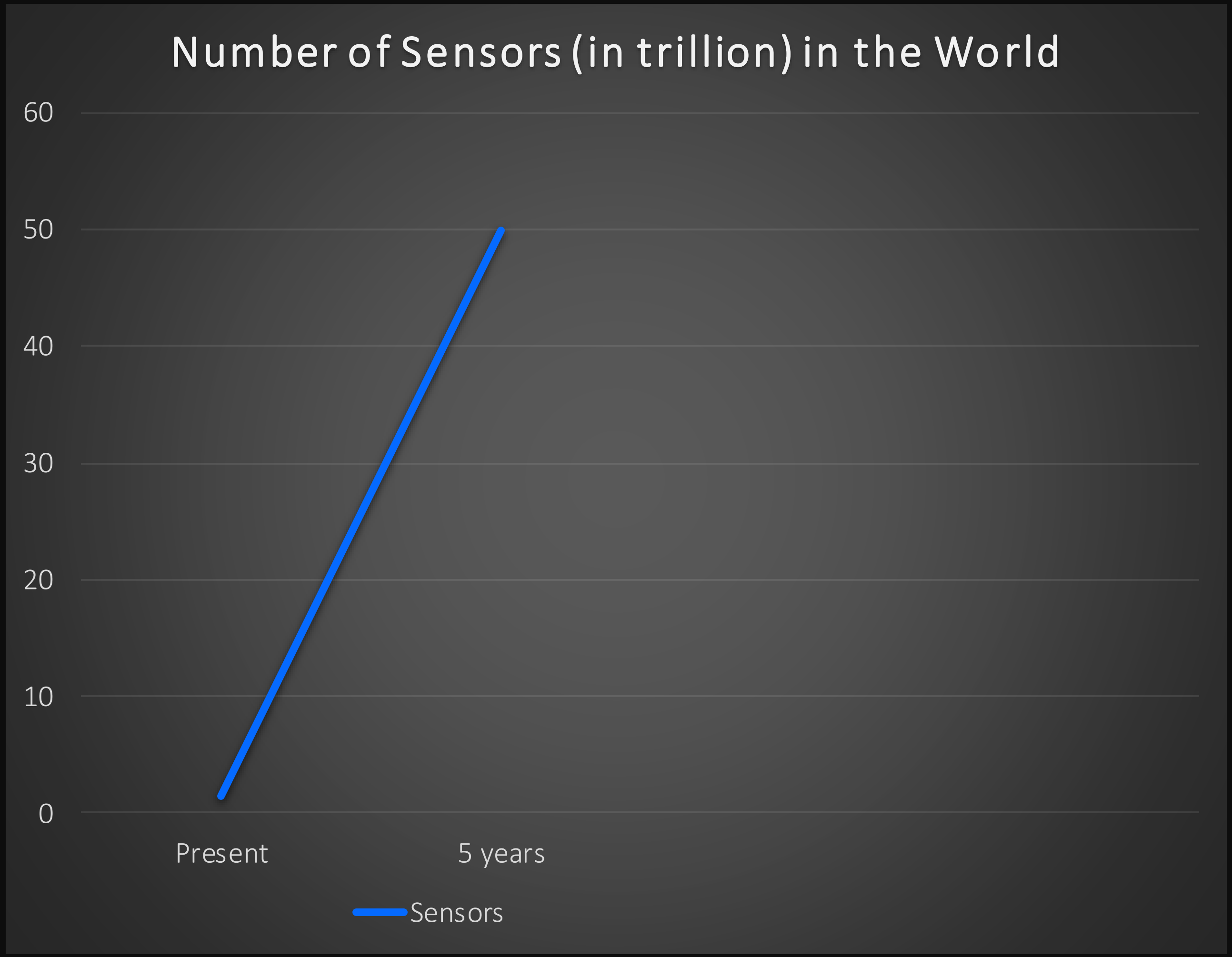
Content

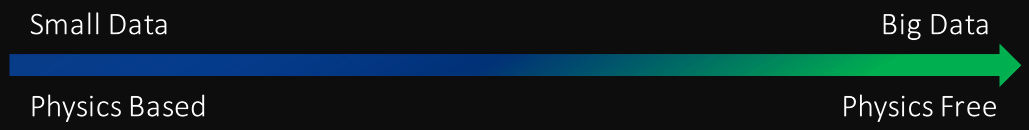
- Hidden Physics Models
- Analysis
- Alternative Approaches

Sample-Efficient Learning



Sample-Efficient Learning



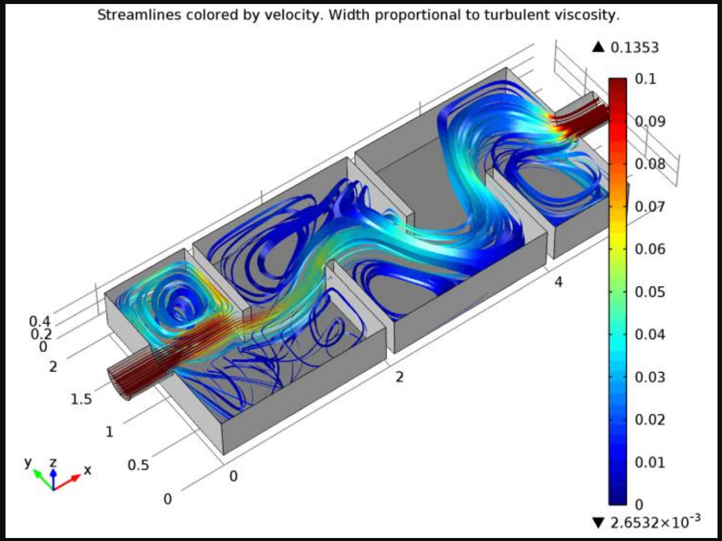


- Scientific data: not easy to obtain
- Big Data: Time Consuming+ Expensive
- Error Prone
- Complexity of the system

Partial Differential Equations

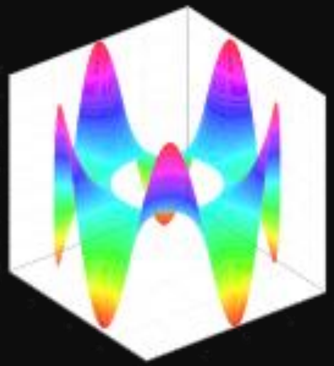
In the general case, a *second-order nonlinear partial differential* equation with two independent variables *has the form*

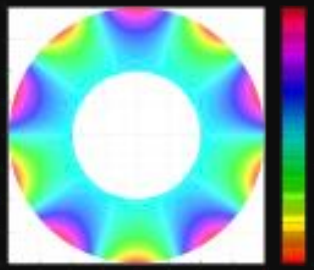
$$F\left(x, y, w, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial^2 w}{\partial x^2}, \frac{\partial^2 w}{\partial x\,\partial y}, \frac{\partial^2 w}{\partial y^2}\right) = 0.$$



Laplace Equation

Navier-Stokes Equation





Hidden Physics Model

- Data-efficient Learning Machines
- Capable of leveraging the laws of Physics

Hidden Physics Model

- Data-efficient Learning Machines
- Capable of leveraging the laws of Physics

Laws of Physics?

Time Dependent and non-linear partial differential Equations

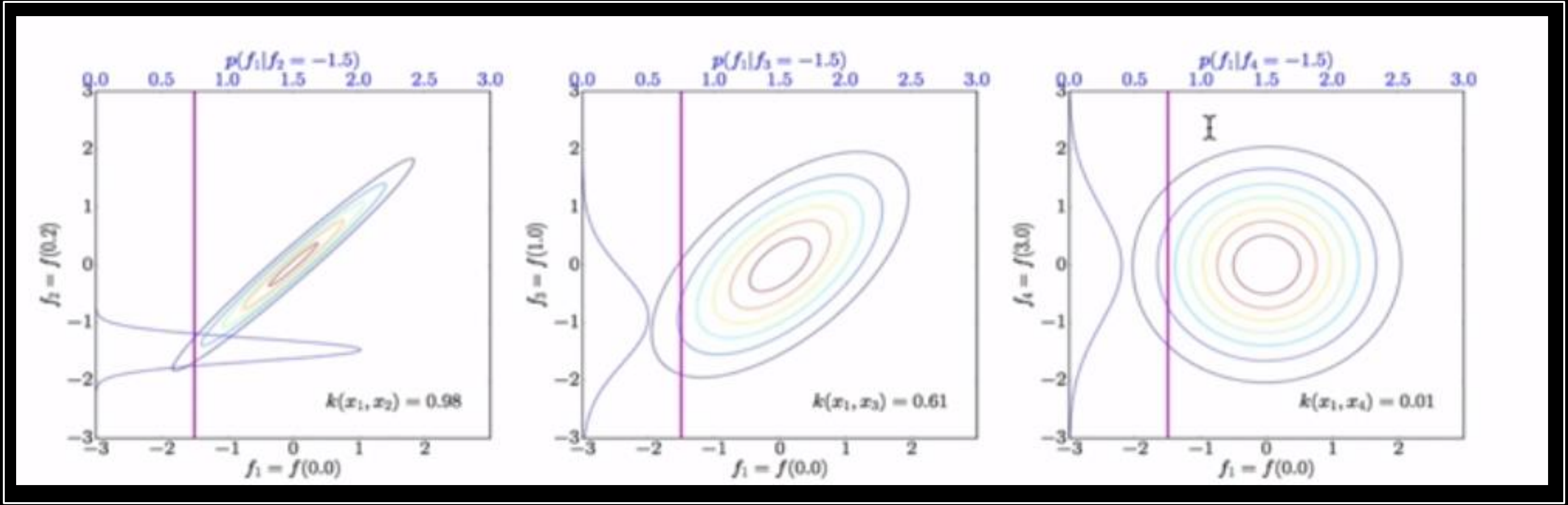
Gaussian Process

A GP can be *completely* defined by a mean function $\mu(t)$ and a variance/kernel function $K(s, t) := Var(X_s, X_t)$, i.e.,

$$X_t \sim GP(\mu(\cdot), K(\cdot, \cdot))$$

Usually $\mu(t)=0$;

$$K(s, t) = \exp(-\Theta || x_s - X_t ||),$$



$$k(x, x'; \theta) = \gamma^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D} w_d^2 (x_d - x'_d)^2\right),$$

Applications in Real World

- Learning
- System Identification
- Data-driven discovery of Partial Differential Equations

Problem Setup

$$h_t + \mathcal{N}_x^{\lambda} h = 0, \;\; x \in \Omega, \;\; t \in [0, T], \qquad (1)$$

$\longrightarrow$ Find the Parameters λ

Problem Setup

$$h_t + \mathcal{N}_x^\lambda h = 0,\ x \in \Omega,\ t \in [0, T], \quad (1)$$

Find the Parameters λ

$$h^n + \Delta t \boxed{\mathcal{N}_x^\lambda} h^n = h^{n-1}. \quad (2)$$

Non-Linear Operator

$$\mathcal{N}_x^\lambda h = \lambda_1 h h_x - \lambda_2 h_{xx}$$

1D Burgers Equation

Problem Setup

$$h_t + \mathcal{N}_x^{\lambda} h = 0,\ \ x \in \Omega,\ \ t \in [0, T], \qquad (1)$$

Find the Parameters λ

$$h^n + \Delta t \boxed{\mathcal{N}_x^{\lambda}} h^n = h^{n-1}. \qquad (2)$$

Non-Linear Operator

$$\boxed{\mathcal{L}_x^{\lambda}} h^n = h^{n-1}. \qquad (3)$$

Linear Operator

Euler Time Stepping Scheme

$$u^n = u^{n-1} + \Delta t\ \mathcal{L}_x u^{n-1}.$$

Forward Euler	$u^n = u^{n-1} + \Delta t \mathcal{L}_x u^{n-1}$
Backward Euler	$u^n = u^{n-1} + \Delta t \mathcal{L}_x u^n$
Trapezoidal Rule	$u^n = u^{n-1} + \frac{1}{2}\Delta t \mathcal{L}_x u^{n-1} + \frac{1}{2}\Delta t \mathcal{L}_x u^n$

The Basic Model

Placing a Gaussian Process Prior,

$$h^n(x) \sim \mathcal{GP}(0, k(x, x', \theta)). \tag{4}$$

Entire structure of the linear operator can be captured in the multi-output Gaussian Process,

$$\boxed{\begin{bmatrix} h^n \\ h^{n-1} \end{bmatrix} \sim \mathcal{GP}\left(0, \begin{bmatrix} k^{n,n} & k^{n,n-1} \\ k^{n-1,n} & k^{n-1,n-1} \end{bmatrix}\right). \qquad (5)}$$

Hidden physics model

Learning

The parameters λ of the operators (both Linear and Non-Linear) can be learned via

$$- \log p(\boldsymbol{h}|\theta, \lambda, \sigma^2) = \frac{1}{2}\boldsymbol{h}^T \boldsymbol{K}^{-1}\boldsymbol{h} + \frac{1}{2}\log|\boldsymbol{K}| + \frac{N}{2}\log(2\pi), \qquad (6)$$

$$\boldsymbol{K} = \begin{bmatrix} k^{n,n}(\boldsymbol{x}^n, \boldsymbol{x}^n) & k^{n,n-1}(\boldsymbol{x}^n, \boldsymbol{x}^{n-1}) \\ k^{n-1,n}(\boldsymbol{x}^{n-1}, \boldsymbol{x}^n) & k^{n-1,n-1}(\boldsymbol{x}^{n-1}, \boldsymbol{x}^{n-1}) \end{bmatrix} + \sigma^2\boldsymbol{I}.$$

Learning

The parameters λ of the operators (both Linear and Non-Linear) can be learned via

targets fitting the training data   penalizes model complexity

$$-\log p(\boldsymbol{h}|\theta, \lambda, \sigma^2) = \frac{1}{2}\boxed{\boldsymbol{h}^T\boldsymbol{K}^{-1}\boldsymbol{h}} + \frac{1}{2}\boxed{\log|\boldsymbol{K}|} + \frac{N}{2}\log(2\pi), \tag{6}$$

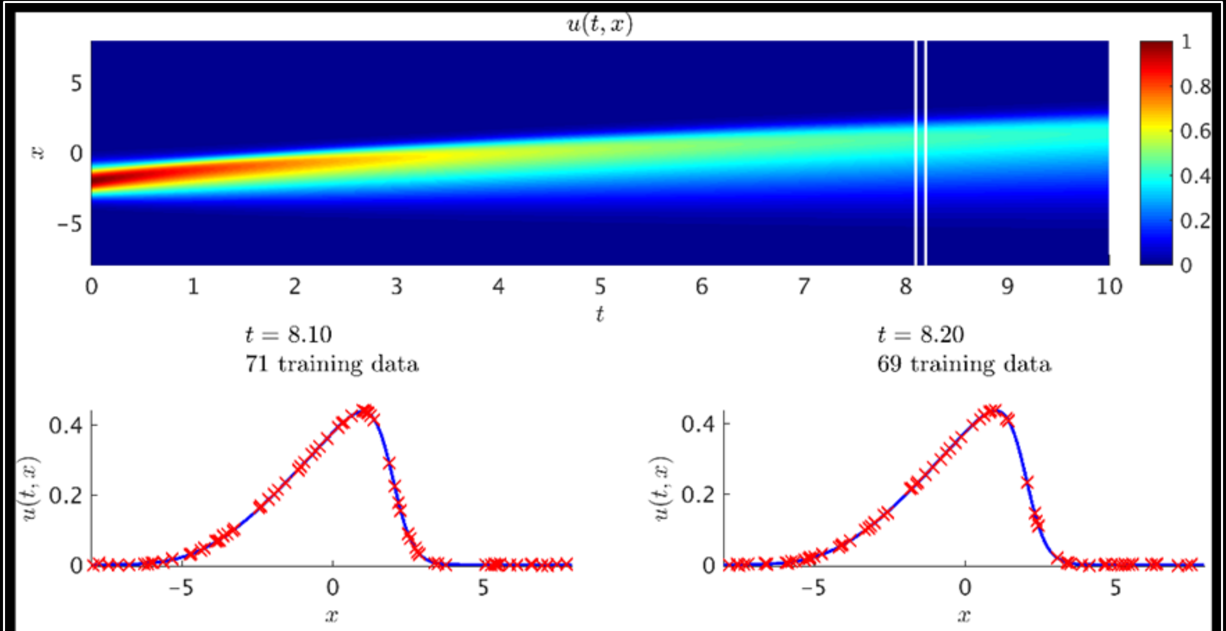
$$\boldsymbol{K} = \begin{bmatrix} k^{n,n}(\boldsymbol{x}^n, \boldsymbol{x}^n) & k^{n,n-1}(\boldsymbol{x}^n, \boldsymbol{x}^{n-1}) \\ k^{n-1,n}(\boldsymbol{x}^{n-1}, \boldsymbol{x}^n) & k^{n-1,n-1}(\boldsymbol{x}^{n-1}, \boldsymbol{x}^{n-1}) \end{bmatrix} + \sigma^2\boldsymbol{I}.$$

Results

Burgers' Equation

(Fluid mechanics, non-linear acoustics, traffic flow)

$$u_t + \lambda_1 u u_x - \lambda_2 u_{xx} = 0, \quad (7)$$



Correct PDE	$u_t + uu_x - 0.1u_{xx} = 0$
Identified PDE (clean data)	$u_t + 1.028uu_x - 0.101u_{xx} = 0$
Identified PDE (1% noise)	$u_t + 1.017uu_x - 0.094u_{xx} = 0$

Burgers' Equation

$$u_t + \lambda_1 u u_x - \lambda_2 u_{xx} = 0, \qquad (7)$$

	Clean Data		1% Noise		5% Noise	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
First Quartile	1.0247	0.0942	0.9168	0.0784	0.3135	0.0027
Median	1.0379	0.0976	1.0274	0.0919	0.8294	0.0981
Third Quartile	1.0555	0.0987	1.1161	0.1166	1.2488	0.1543

Table 1: *Burgers' equation:* Resulting statistics for the learned parameter values.

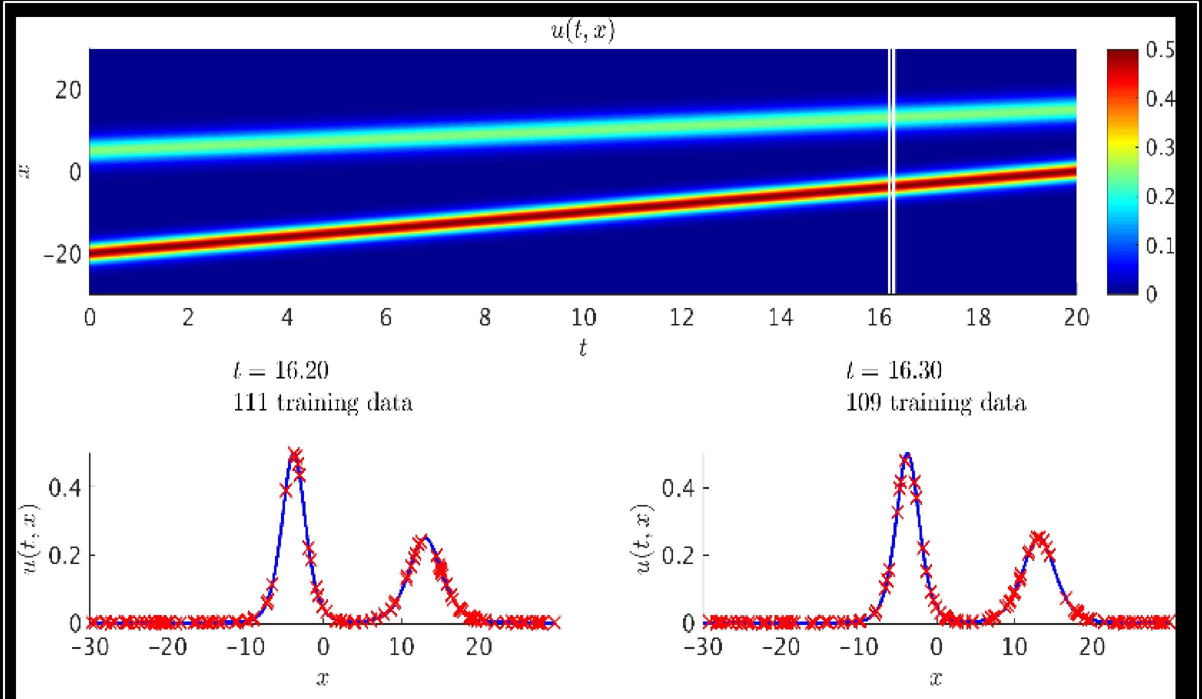
		$\Delta t = 0.1$	$\Delta t = 0.5$	$\Delta t = 1.0$	$\Delta t = 1.5$
Clean Data	λ_1	1.0283	1.1438	1.2500	1.2960
	λ_2	0.1009	0.0934	0.0694	0.0431
1% Noise	λ_1	1.0170	1.1470	1.2584	1.3063
	λ_2	0.0935	0.0939	0.0711	0.0428

Table 2: *Burgers' equation:* Effect of increasing the gap Δt between the pair of snapshots.

KdV Equation

(Waves on shallow water surfaces)

$$u_t + \lambda_1 u u_x + \lambda_2 u_{xxx} = 0, \qquad (8)$$



Correct PDE	$u_t + 6uu_x + u_{xxx} = 0$
Identified PDE (clean data)	$u_t + 6.115uu_x + 1.047u_{xxx} = 0$
Identified PDE (1% noise)	$u_t + 5.722uu_x + 0.958u_{xxx} = 0$

KdV Equation

$$u_t + \lambda_1 u u_x + \lambda_2 u_{xxx} = 0, \quad (8)$$

	Clean Data		1% Noise		5% Noise	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
First Quartile	5.7783	0.9299	5.3358	0.7885	3.7435	0.2280
Median	5.8920	0.9656	5.5757	0.8777	4.5911	0.6060
Third Quartile	6.0358	1.0083	5.7840	0.9491	5.5106	0.8407

Table 3: *The KdV equation:* Resulting statistics for the learned parameter values.

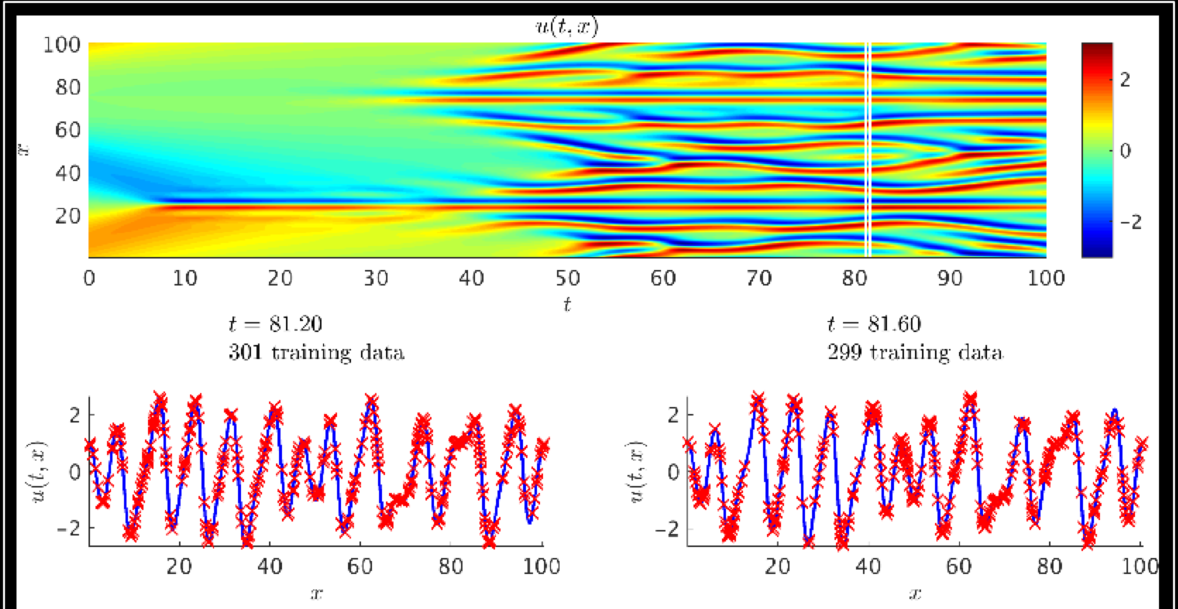
		$\Delta t = 0.1$	$\Delta t = 0.2$	$\Delta t = 0.3$	$\Delta t = 0.4$	$\Delta t = 0.5$
Clean Data	λ_1	6.1145	5.8948	5.4014	4.1779	3.5058
	λ_2	1.0470	0.9943	0.8535	0.4475	0.1816
1% Noise	λ_1	5.7224	5.8288	5.4054	4.1479	3.4747
	λ_2	0.9578	0.9801	0.8563	0.4351	0.1622

Table 4: *The KdV equation:* Effect of increasing the gap Δt between the pair of snapshots.

Kuramoto-Sivashinsky Equation

(Pattern forming system with spatio-temporal behaviour)

$$u_t + \lambda_1 u u_x + \lambda_2 u_{xx} + \lambda_3 u_{xxxx} = 0, \tag{9}$$



Correct PDE	$u_t + uu_x + u_{xx} + u_{xxxx} = 0$
Identified PDE (clean data)	$u_t + 0.952uu_x + 1.005u_{xx} + 0.980u_{xxxx} = 0$
Identified PDE (1% noise)	$u_t + 0.908uu_x + 0.951u_{xx} + 0.927u_{xxxx} = 0$

Kuramoto-Sivashinsky Equation

$$u_t + \lambda_1 u u_x + \lambda_2 u_{xx} + \lambda_3 u_{xxxx} = 0, \qquad (9)$$

	Clean Data			1% Noise			5% Noise		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
First Quartile	0.9603	0.9829	0.9711	0.7871	0.8095	0.5891	-0.0768	0.0834	-0.0887
Median	0.9885	1.0157	0.9970	0.8746	0.9124	0.8798	0.4758	0.5539	0.4086
Third Quartile	1.0187	1.0550	1.0314	0.9565	0.9948	0.9553	0.6991	0.7644	0.7009

Table 5: *Kuramoto-Sivashinsky equation:* Resulting statistics for the learned parameter values.

		$\Delta t = 0.4$	$\Delta t = 0.8$	$\Delta t = 1.2$
Clean Data	λ_1	0.9515	0.5299	0.1757
	λ_2	1.0052	0.5614	0.1609
	λ_3	0.9803	0.5438	0.1647
1% Noise	λ_1	0.9081	0.5124	0.1616
	λ_2	0.9511	0.5387	0.1436
	λ_3	0.9266	0.5213	0.1483

Table 6: *Kuramoto-Sivashinsky equation:* Effect of increasing the gap Δt between the pair of snapshots.

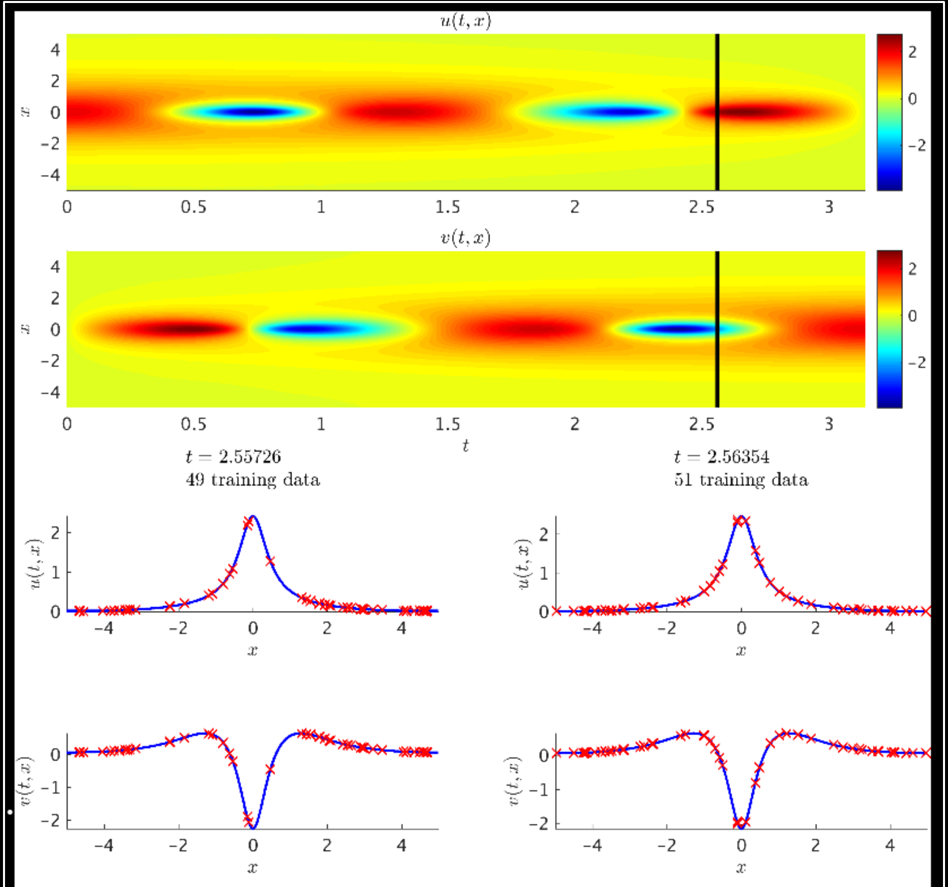
Nonlinear Schrödinger Equation

(Wave propagation in optical fibres)

$$ih_t + \lambda_1 h_{xx} + \lambda_2 |h|^2 h = 0, \qquad (10)$$

$$\begin{aligned} u^n + \Delta t \lambda_1 v^n_{xx} + \Delta t \lambda_2 [(u^{n-1})^2 + (v^{n-1})^2] v^n &= u^{n-1}, \\ v^n - \Delta t \lambda_1 u^n_{xx} - \Delta t \lambda_2 [(u^{n-1})^2 + (v^{n-1})^2] u^n &= v^{n-1}, \end{aligned}$$

$$\begin{bmatrix} u^n \\ v^n \\ u^{n-1} \\ v^{n-1} \end{bmatrix} \sim \mathcal{GP}\left(0, \begin{bmatrix} k^{n,n}_{u,u} & k^{n,n}_{u,v} & k^{n,n-1}_{u,u} & k^{n,n-1}_{u,v} \\ k^{n,n}_{v,u} & k^{n,n}_{v,v} & k^{n,n-1}_{v,u} & k^{n,n-1}_{v,v} \\ k^{n-1,n}_{u,u} & k^{n-1,n}_{u,v} & k^{n-1,n-1}_{u,u} & k^{n-1,n-1}_{u,v} \\ k^{n-1,n}_{v,u} & k^{n-1,n}_{v,v} & k^{n-1,n-1}_{v,u} & k^{n-1,n-1}_{v,v} \end{bmatrix}\right)$$



Correct PDE	$ih_t + 0.5h_{xx} + \vert h\vert^2 h = 0$
Identified PDE (clean data)	$ih_t + 0.506h_{xx} + 0.995\,\vert h\vert^2 h = 0$
Identified PDE (1% noise)	$ih_t + 0.476h_{xx} + 0.999\,\vert h\vert^2 h = 0$

Nonlinear Schrödinger Equation

$$ih_t + \lambda_1 h_{xx} + \lambda_2 |h|^2 h = 0, \qquad (10)$$

	Clean Data		1% Noise		5% Noise	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
First Quartile	0.4950	0.9960	0.3714	0.9250	-0.1186	0.6993
Median	0.5009	1.0001	0.4713	0.9946	0.4259	0.9651
Third Quartile	0.5072	1.0039	0.5918	1.0670	0.9730	1.2730

Table 7: *Nonlinear Schrödinger equation:* Resulting statistics for the learned parameter values.

		$\Delta t = 0.0063$	$\Delta t = 0.0628$	$\Delta t = 0.1257$	$\Delta t = 0.1885$
Clean Data	λ_1	0.5062	0.4981	0.3887	0.3097
	λ_2	0.9949	0.8987	0.7936	0.7221
1% Noise	λ_1	0.4758	0.4976	0.3928	0.3128
	λ_2	0.9992	0.9011	0.7975	0.7255

Table 8: *Nonlinear Schrödinger equation:* Effect of increasing the gap Δt between the pair of snapshots.

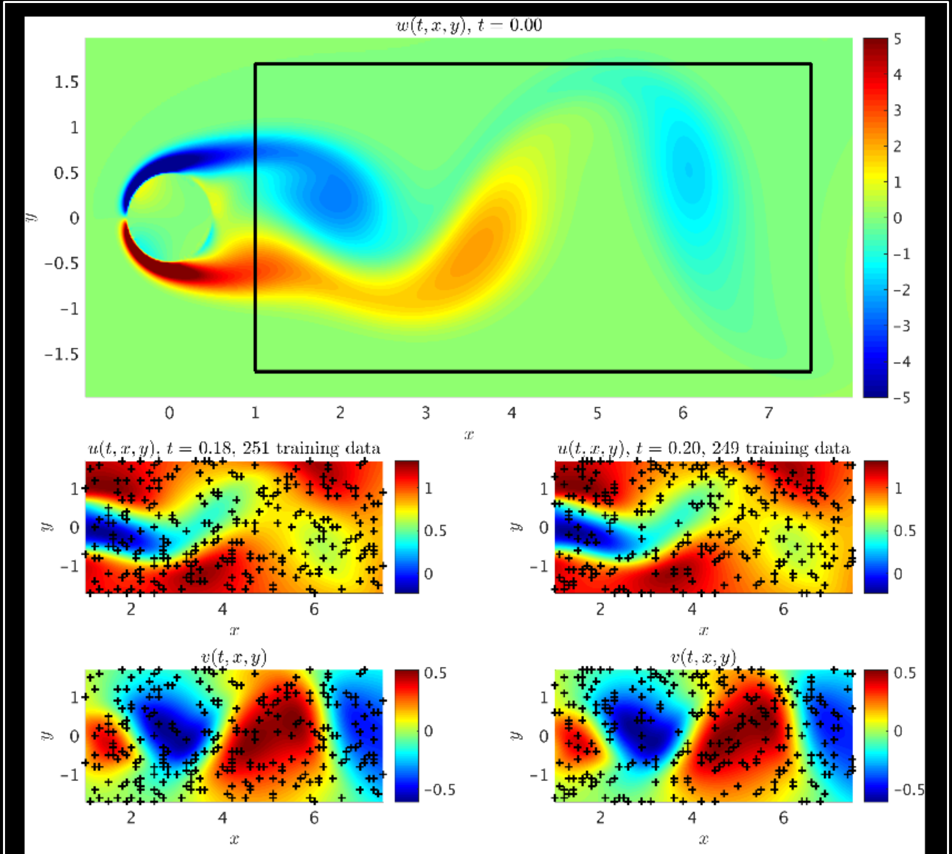
Navier-Stokes Equation

(Weather, Ocean Currents, Blood flow)

$$
\begin{aligned}
u_t + \lambda_1(u u_x + v u_y) &= -p_x + \lambda_2(u_{xx} + u_{yy}),\\
v_t + \lambda_1(u v_x + v v_y) &= -p_y + \lambda_2(v_{xx} + v_{yy}),
\end{aligned}
$$

$$u^n = \psi_y^n, \qquad v^n = -\psi_x^n,$$

$$\begin{bmatrix} u^n \\ v^n \end{bmatrix} \sim \mathcal{GP}\left(0, \begin{bmatrix} k_{u,u}^{n,n} & k_{u,v}^{n,n} \\ k_{v,u}^{n,n} & k_{v,v}^{n,n} \end{bmatrix}\right),$$



Correct PDE	$u_t + (uu_x + vu_y) = -p_x + 0.01(u_{xx} + u_{yy})$ $v_t + (uv_x + vv_y) = -p_y + 0.01(v_{xx} + v_{yy})$
Identified PDE (clean data)	$u_t + 0.983(uu_x + vu_y) = -p_x + 0.00826(u_{xx} + u_{yy})$ $v_t + 0.983(uv_x + vv_y) = -p_y + 0.00826(v_{xx} + v_{yy})$
Identified PDE (1% noise)	$u_t + 0.849(uu_x + vu_y) = -p_x + 0.01399(u_{xx} + u_{yy})$ $v_t + 0.849(uv_x + vv_y) = -p_y + 0.01399(v_{xx} + v_{yy})$

Navier-Stokes Equation

$$u_t + \lambda_1(uu_x + vu_y) = -p_x + \lambda_2(u_{xx} + u_{yy}),$$
$$v_t + \lambda_1(uv_x + vv_y) = -p_y + \lambda_2(v_{xx} + v_{yy}),$$

	Clean Data		1% Noise		5% Noise	
	λ_1	λ_2	λ_1	λ_2	λ_1	λ_2
First Quartile	0.9854	0.0069	0.8323	0.0057	0.5373	0.0026
Median	0.9928	0.0077	0.8717	0.0063	0.6498	0.0030
Third Quartile	1.0001	0.0086	0.9102	0.0070	0.7619	0.0046

Table 9: *Navier-Stokes equations:* Resulting statistics for the learned parameter values.

		$\Delta t = 0.02$	$\Delta t = 0.04$	$\Delta t = 0.06$	$\Delta t = 0.08$	$\Delta t = 1.0$
Clean Data	λ_1	0.9834	0.9925	0.9955	0.9976	1.0021
	λ_2	0.0083	0.0072	0.0058	0.0040	0.0027
1% Noise	λ_1	0.8488	0.9298	0.9597	0.9726	0.9791
	λ_2	0.0140	0.0110	0.0088	0.0069	0.0053

Table 10: *Navier-Stokes equations:* Effect of increasing the gap Δt between the pair of snapshots.

		$\Delta t = 0.02$	$\Delta t = 0.01$	$\Delta t = 0.005$
Clean Data	λ_1	0.9834	0.9688	0.9406
	λ_2	0.0083	0.0091	0.0104
1% Noise	λ_1	0.8488	0.7384	0.6107
	λ_2	0.0140	0.0159	0.0217

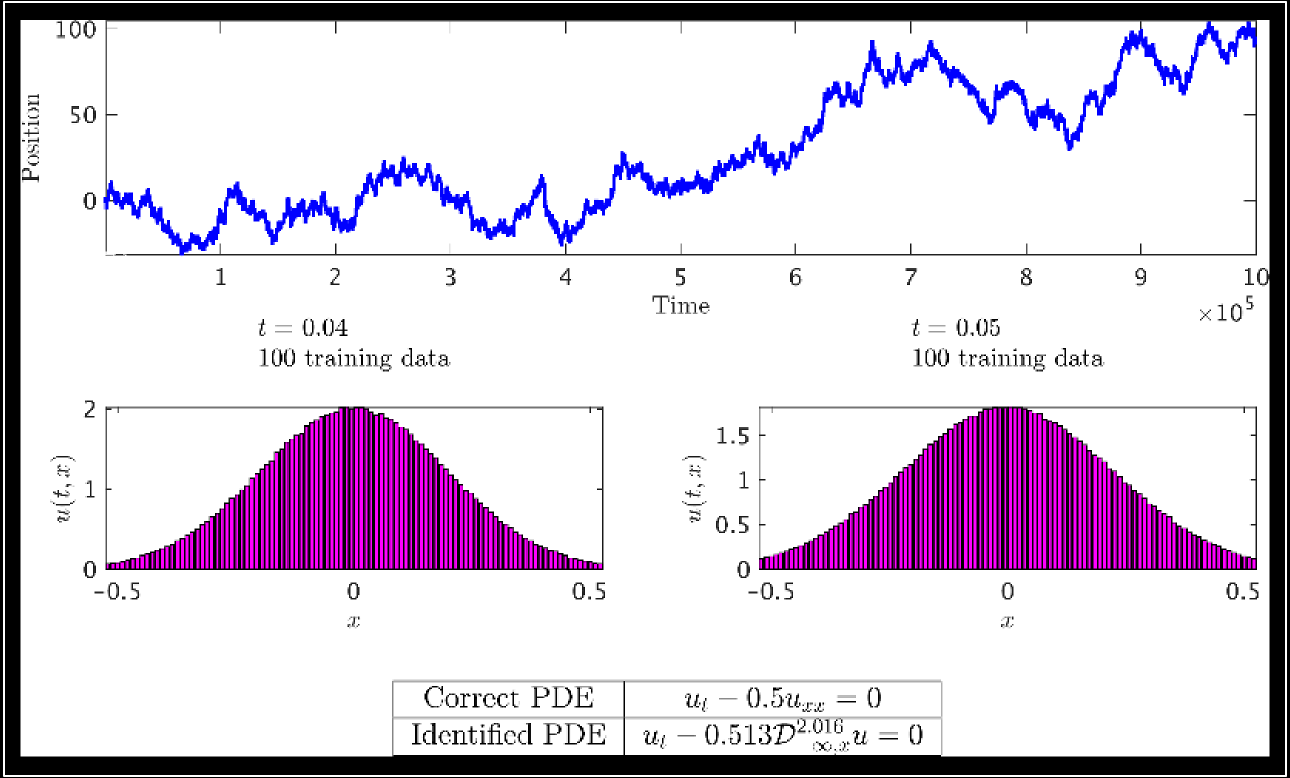
Table 11: *Navier-Stokes equations:* Effect of decreasing the gap Δt between the pair of snapshots.

Fractional Equation

$$u_t - \lambda_1 \mathcal{D}^{\lambda_2}_{-\infty,x} u = 0,$$

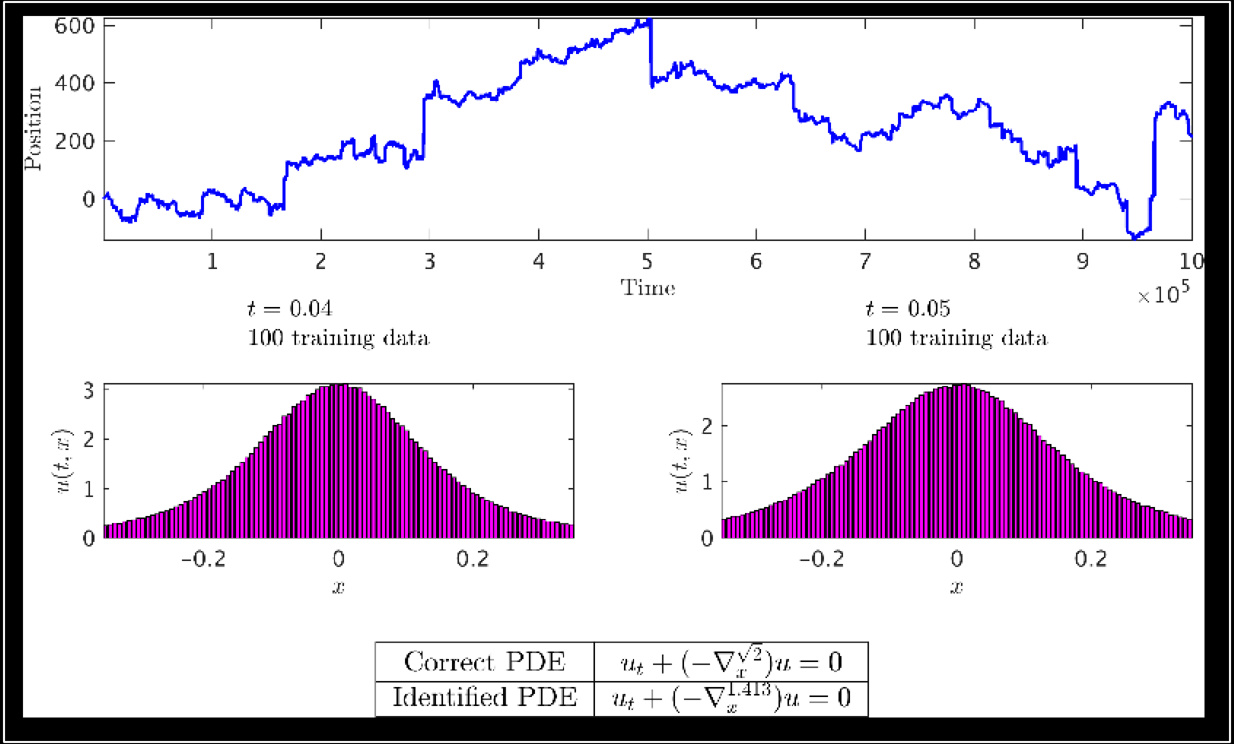
$$[1 - \Delta t \lambda_1 (-iw')^{\lambda_2}] \boxed{\widehat{k}(w, w'; \theta)},$$

Fourier Transform of the kernel



Fractional Equation

$$u_t - \lambda_1 \mathcal{D}^{\lambda_2}_{-\infty,x} u = 0,$$



Patterns in the results?

More data, less noise, and a smaller gap between the two snapshots ✓

More noise in the data leads to less confidence in the estimated parameter values. ✓

Analysis

Pros

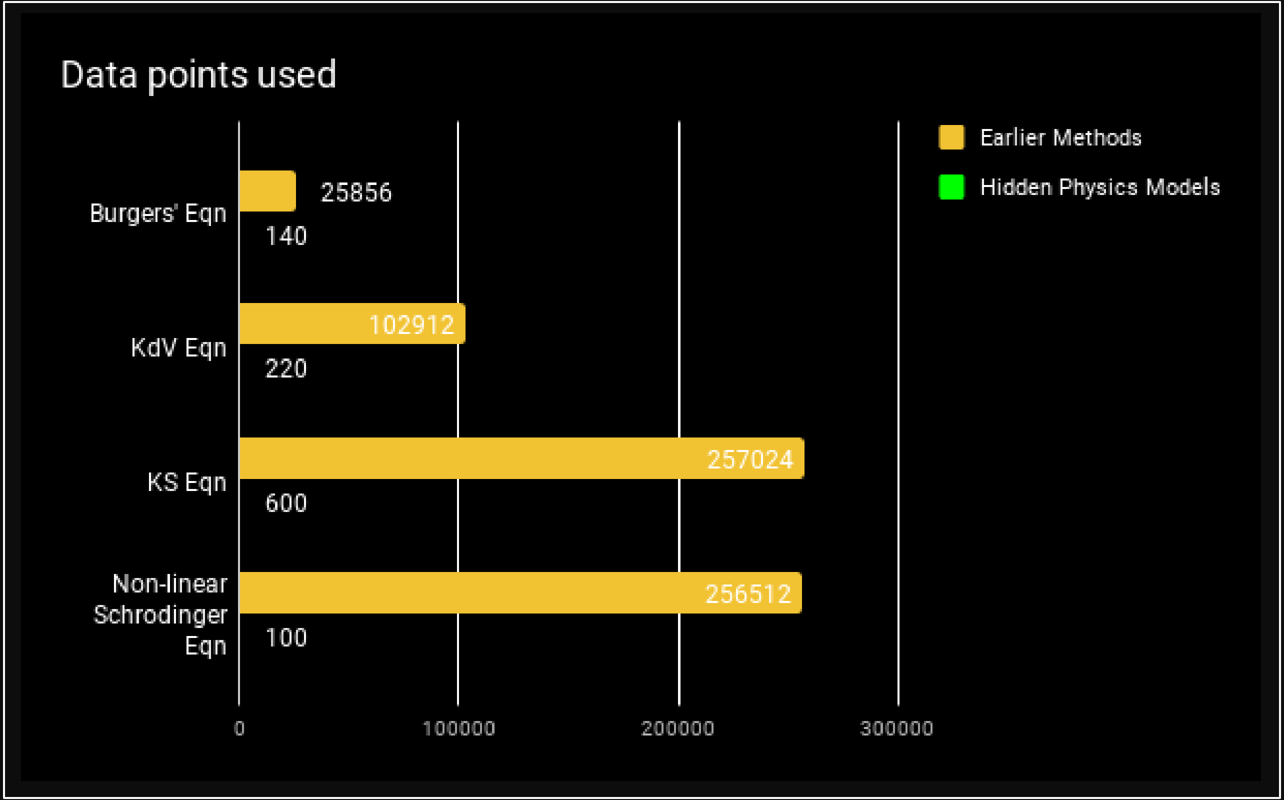
No dictionary

Estimate parameters appearing anywhere

Kernels are obtained analytically

Regular lattice **not essential**

Scattered data



Analysis

Cons

$$-\log p(\boldsymbol{h}|\theta, \lambda, \sigma^2) = \frac{1}{2}\boldsymbol{h}^T\boldsymbol{K}^{-1}\boldsymbol{h} + \frac{1}{2}\log|\boldsymbol{K}| + \frac{N}{2}\log(2\pi), \qquad (6)$$

- Multiple Local Minima

Alternative Approaches

- Assign priors on $(\lambda, \Theta, \{\sigma\}^2)$

- For the limitation around cubic scaling with respect to the total number of training data points,

  - Kalman updates
  - Variational inference
  - Parametric Gaussian processes

Conclusion

1. Wide range application

2. Sample efficient learning

3. Works well in less noisy data and for smaller time steps

References

[1] M. Raissi, P. Perdikaris, G. E. Karniadakis, Numerical gaussian processes for time-dependent and non-linear partial differential equations,. arXiv preprint arXiv:1703.10230 (2017)
[2] S. H. Rudy, S. L. Brunton, J. L. Proctor, J. N. Kutz, Data-driven discovery of partialdifferential equations,. Science Advances 3 (2017).
[3] M. Ebden, Gaussian Processes for Regression: A Quick Introduction. August 2008
[4] J. Hensman, N. Fusi, N. D. Lawrence, Gaussian processes for big data, in: Proceedings of the Twenty-Ninth Conference on Uncertainty in Articial Intelligence,. UAI 2013, Bellevue, WA, USA, August 11-15, 2013.
[5] M. Raissi, Parametric gaussian process regression for big data,. arXiv preprint arXiv:1704.03144 (2017).

Thank You