KU LEUVEN GENT

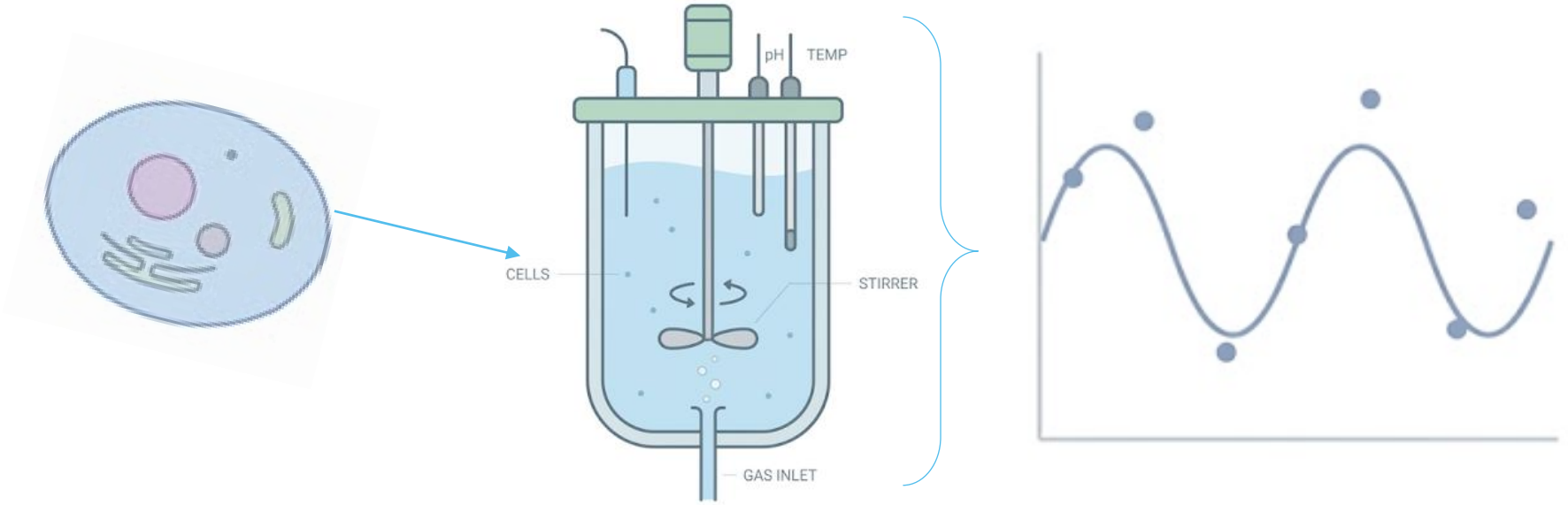
When physics is not enough: *Scientific Machine Learning for Biosystems*

Satyajeet Bhonsale
BioTeC+, KU Leuven

Biology

Process

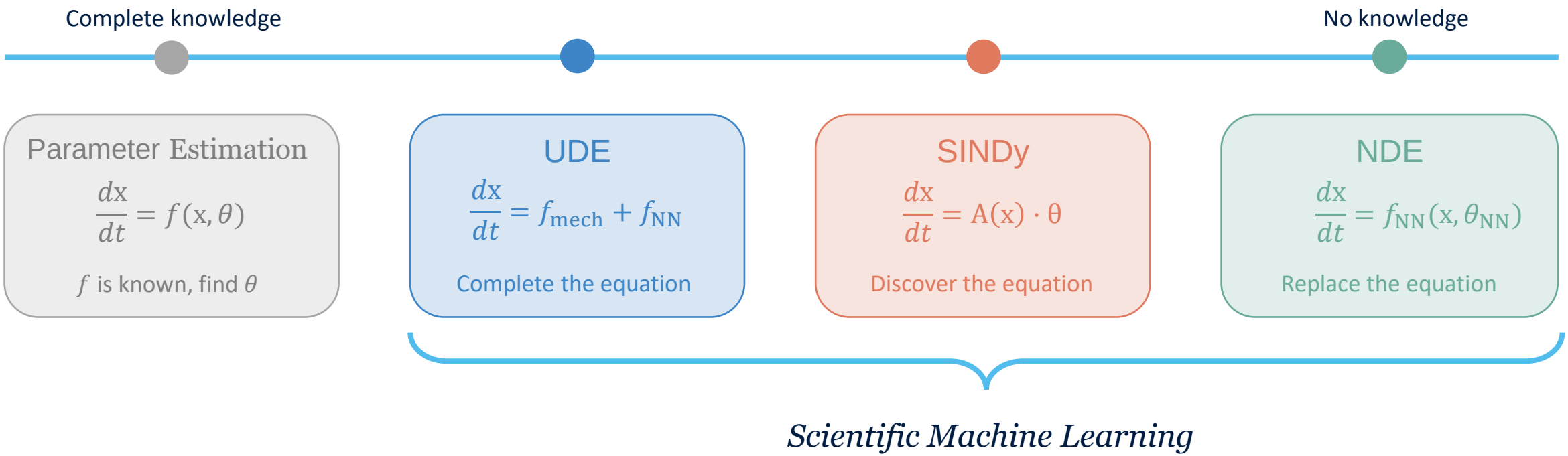
Data



$$\frac{dx}{dt} = f(\text{???})$$

What is the right equation?

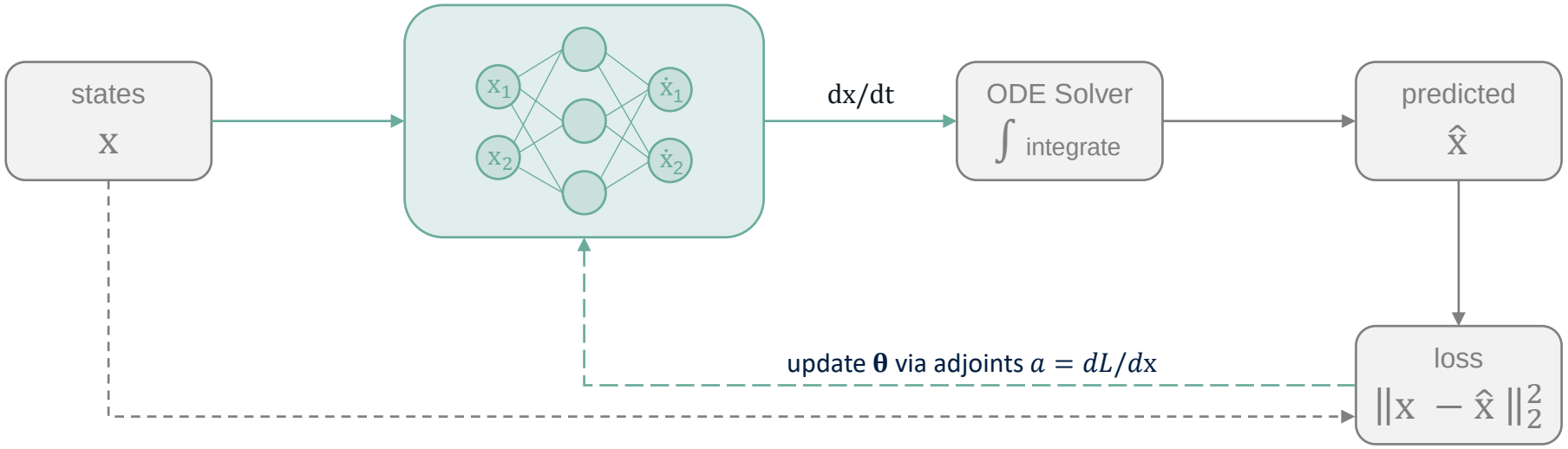
How much f do you already know?



Scientific Machine Learning

NDE: Neural Differential Equations

$$\frac{d\mathrm{x}}{dt} = f_{\mathrm{NN}}(t, \mathrm{x}, \theta_{\mathrm{NN}})$$

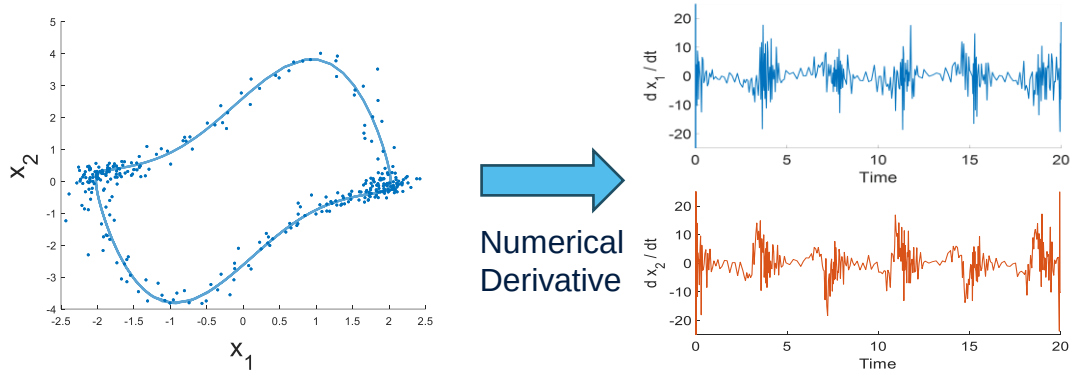


Complete knowledge

No knowledge

SINDy: Sparse Identification of Nonlinear Dynamics

$$\frac{dx}{dt} = \mathrm{A}(\mathrm{x}) \cdot \theta$$



Complete knowledge

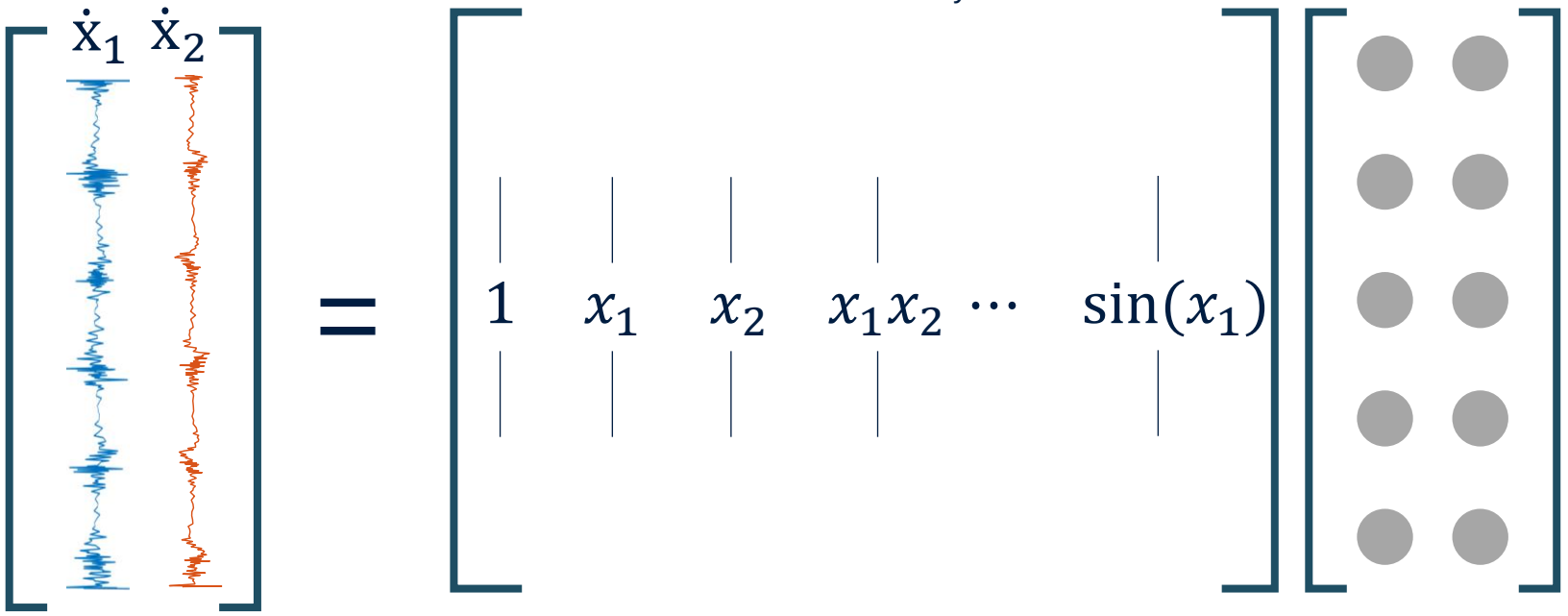
No knowledge

SINDy: Sparse Identification of Nonlinear Dynamics

$$\frac{dx}{dt} = A(x) \cdot \theta$$

Numerical Derivative

Functional Library



Sparse Linear Regression

$$\|\dot{x} - A(x)\theta\|_2^2 + \alpha\|\theta\|_2^2$$

Complete knowledge

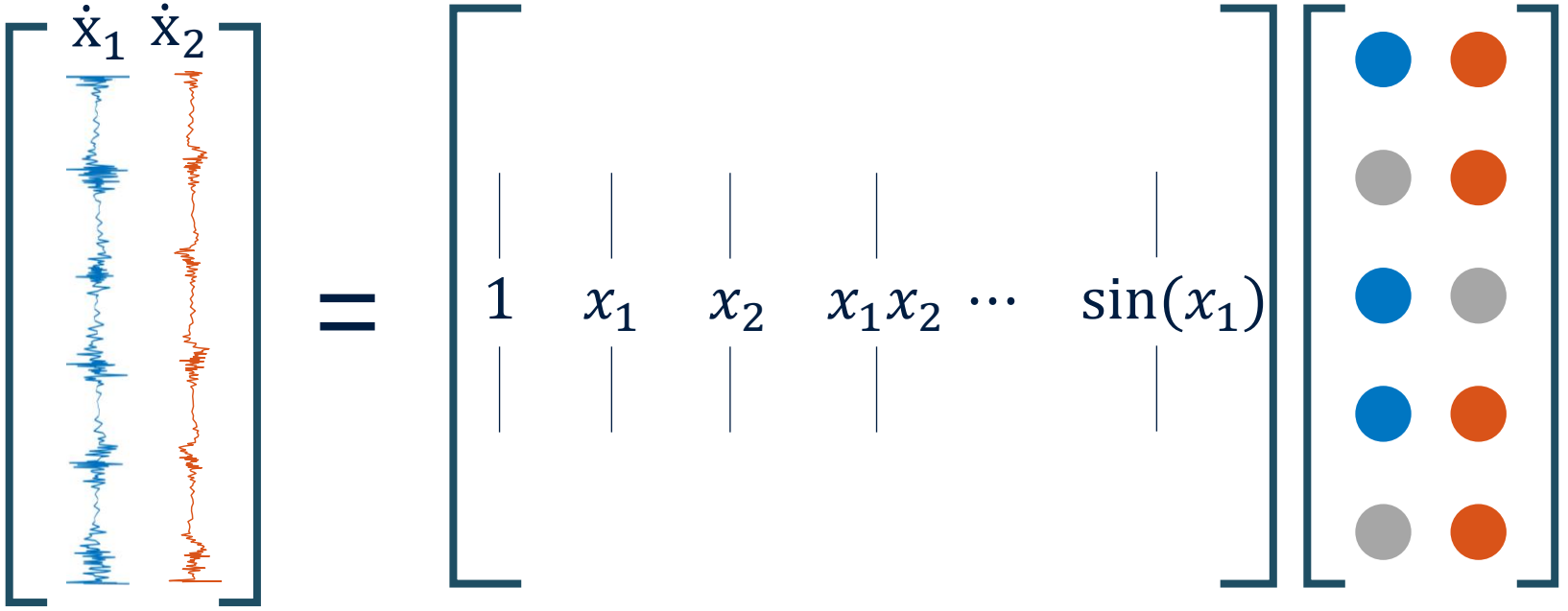
No knowledge

SINDy: Sparse Identification of Nonlinear Dynamics

$$\frac{dx}{dt} = \mathrm{A}(\mathrm{x}) \cdot \theta$$

Numerical Derivative

Functional Library



$$\dot{x}_1 = a_{11} + a_{12}x_2 + a_{13}x_1x_2$$

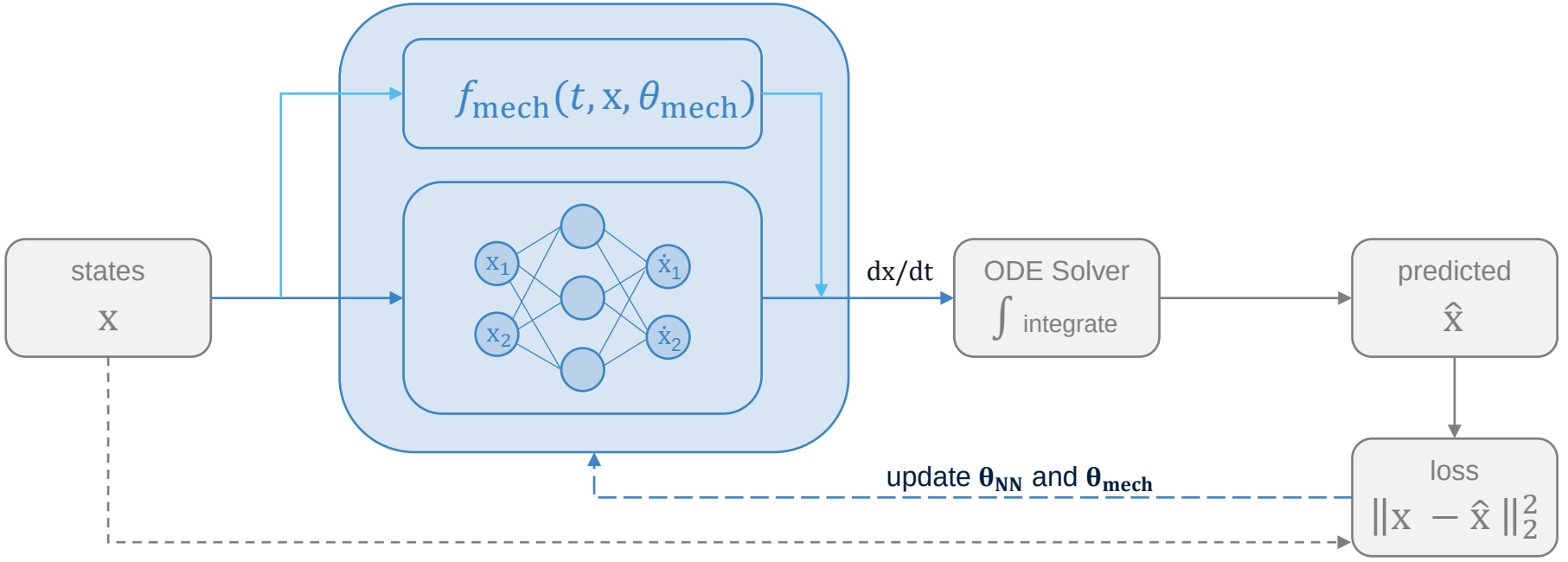
$$\dot{x}_2 = a_{21} + a_{22}x_2 + a_{24}x_1x_2 + a_{25}\sin(x_1)$$

Complete knowledge

No knowledge

UDE: Universal Differential Equations

$$\frac{dx}{dt} = f_{\text{mech}}(t, x, \theta_{\text{mech}}) + f_{\text{NN}}(t, x, \theta_{\text{NN}})$$



Complete knowledge

No knowledge

The Test: Lotka-Volterra Model

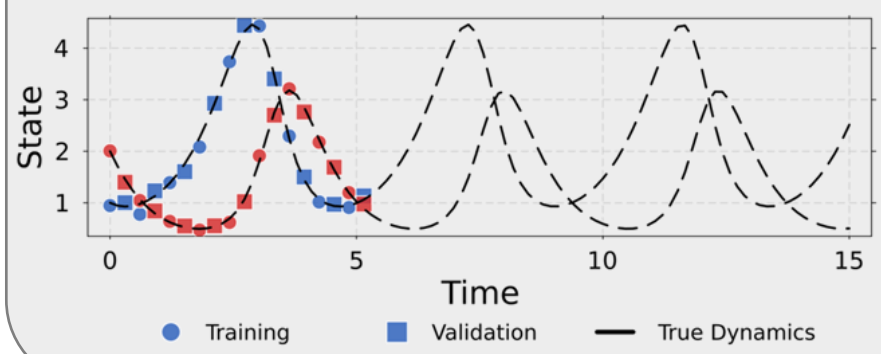
True Governing Equations

$$\frac{dx_1}{dt} = \alpha x_1 - \beta x_1 x_2$$

$$\frac{dx_2}{dt} = -\delta x_2 + \gamma x_1 x_2$$

$$\alpha = 1.3\ \beta = 0.9\ \delta = 1.8\ \gamma = 0.8$$

Experimental Setup



The challenge for each method

NDE: learn everything

SINDy: discover equations

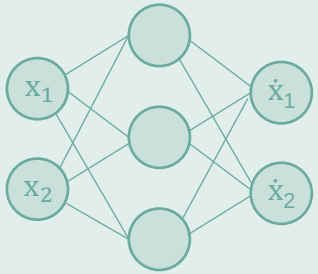
UDE: learn interaction

NDE Results: Accurate, But What Did It Learn?

The Discovered "Model"

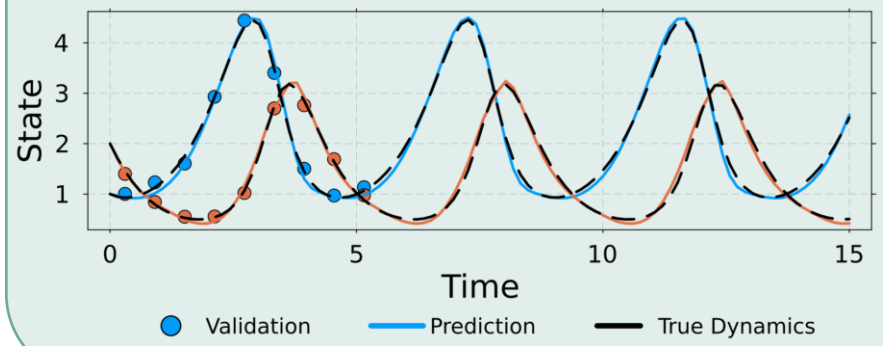
$$\frac{dx_1}{dt} =$$

$$\frac{dx_2}{dt} =$$



1 hidden layer | 3 neurons | RBF

The Modelled Trajectory



The Good

Accurate data fit & stable simulations

The Bad

No interpretable equation or biological insight

The Ugly

Extrapolation is a gamble

SINDy Results: Structure Found, Parameters Lost

The Discovered "Model"

True: $\dot{x}_1 = 1.30x_1 - 0.90x_1x_2$

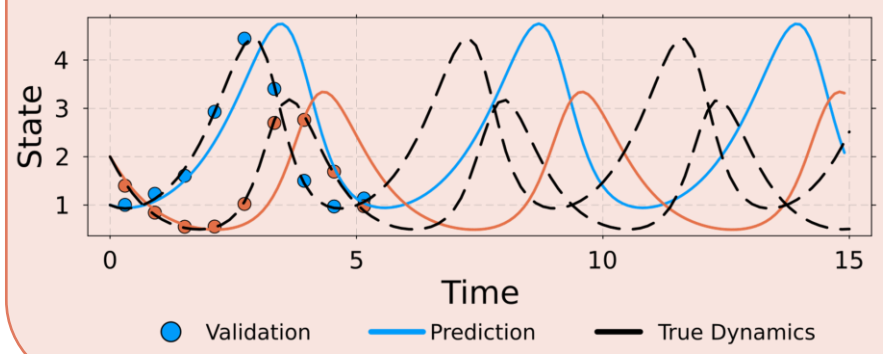
Discovered: $\dot{x}_1 = 1.08x_1 - 0.73x_1x_2$

True: $\dot{x}_2 = -1.80x_1 + 0.80x_1x_2$

Discovered: $\dot{x}_2 = -1.50x_1 - 0.64x_1x_2$

Correct Terms, Wrong Coefficients!

The Modelled Trajectory



The Good

Correct model structure,
Fully interpretable

The Bad

Function library plays a big role

The Ugly

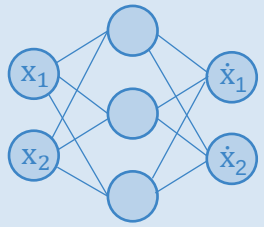
Numerical derivation is a ticking error bomb

UDE Results: Accurate, but Identifiability?

The Discovered "Model"

$$\frac{dx_1}{dt} = 0.31x_1$$
$$\frac{dx_2}{dt} = -1.22x_2$$

$+$

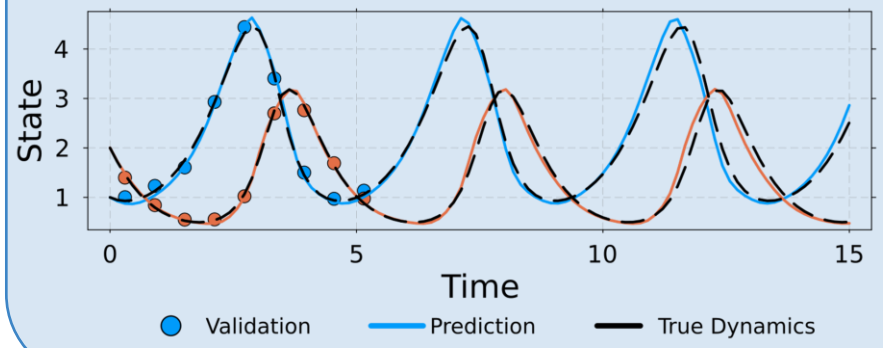


True
$\alpha = 1.3, \delta = 0.9$

$\neq$

Discovered
$\alpha = 0.3, \delta = 1.2$

The Modelled Trajectory



The Good

Accurate data fit & stable simulations

The Bad

Parameters are not unique
NN absorbs part of the physics

The Ugly

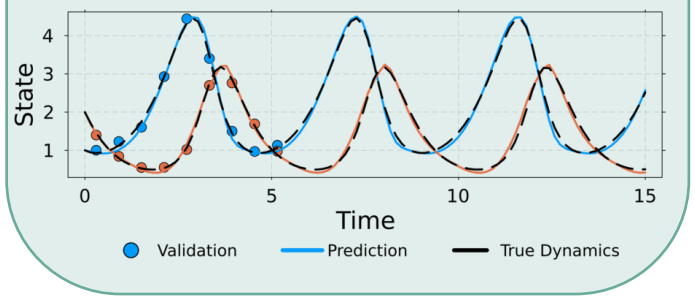
Careful regularization and multistart necessary

What Did Each Method Discover?

NDE

Replace the equation

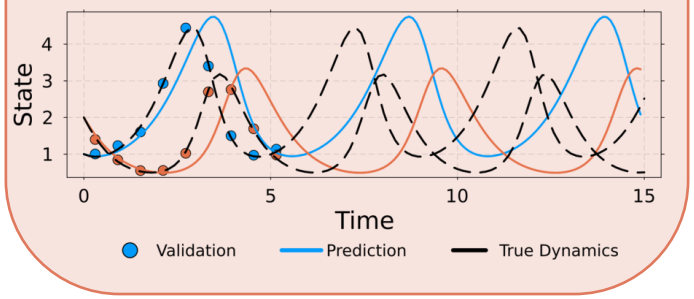
$$\frac{dx_1}{dt} = \dot{x}_1$$
$$\frac{dx_2}{dt} = \dot{x}_2$$



SINDy

Discover the equation

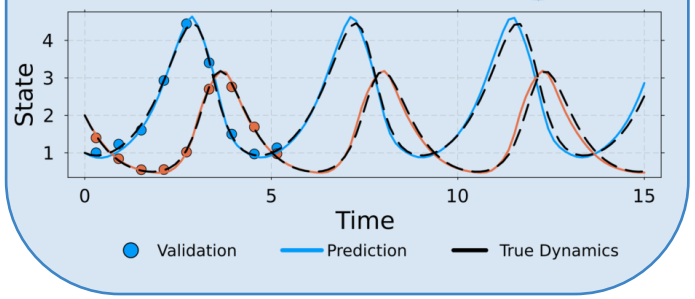
$$dx_1/dt = 1.08x_1 - 0.73x_1x_2$$
$$dx_2/dt = -1.50\,x_1 + 0.64x_1x_2$$



UDE

Complete the equation

$$\frac{dx_1}{dt} = 0.31x_1 + \dot{x}_1$$
$$\frac{dx_2}{dt} = -1.22x_2 + \dot{x}_2$$



Knowledge Axis

No prior knowledge

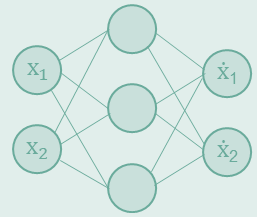
Partially missing knowledge

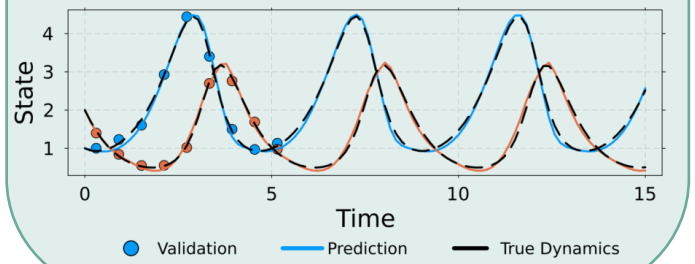
What Did Each Method Discover?

NDE

Replace the equation

$$\frac{dx_1}{dt} = \quad \frac{dx_2}{dt} =$$

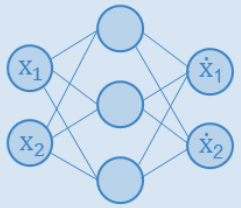


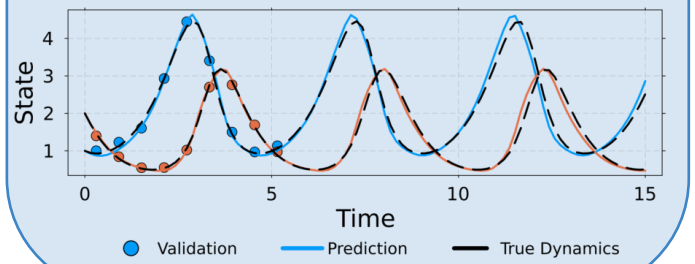


UDE

Complete the equation

$$\frac{dx_1}{dt} = 0.31x_1 \qquad \frac{dx_2}{dt} = -1.22x_2$$

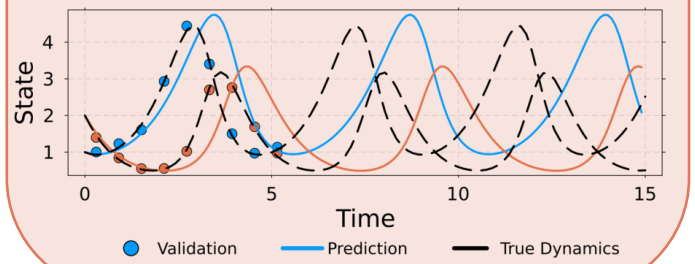




SINDy

Discover the equation

$$dx_1/dt = 1.08x_1 - 0.73x_1x_2$$
$$dx_2/dt = -1.50\,x_1 + 0.64x_1x_2$$

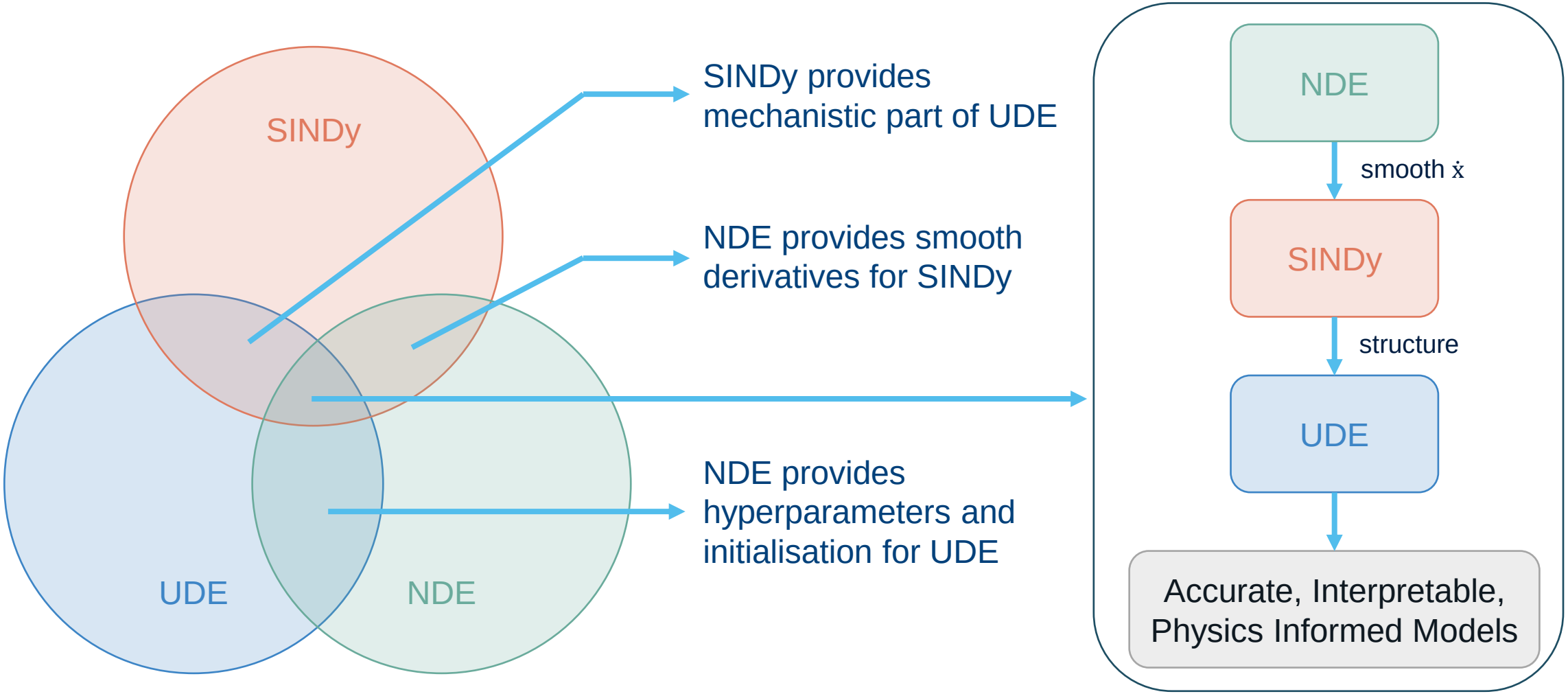


Interpretability Axis

Not interpretable

Interpretable

The SciML Philosophy: Better Together



What lies ahead?

Experiment Design

Which experiments maximise learning?

Equation Discovery

NDE + SINDy + UDE
This Talk

Uncertainty

How much can we trust the model?

Optimisation

How can we use the model?

The Vision

Develop SciML into a comprehensive framework for bioprocess optimisation & control

Key Takeaways

1. **It is possible to discover governing equations from data**
   Not AI/ML replacing models, rather completing them

2. **NDE, SINDy, and UDE are complementary**
   Individually, they trade interpretability for flexibility. Together, they can form a powerful equation discovery pipeline

3. **The field is maturing. but honest challenges remain**
   Identifiability, uncertainty quantification, experimental design are all open problems

Biology follows equations we can't see. SciML helps us find them.