Bridging the Gap Between Machine Learning and Mathematical Models/Analytic Expressions.

CHIMERA: Machine Learning Work Stream

Simon Arridge, Gevik Grigorian, Sandip George
University College London

1. Introduction

Some preliminary facts...

Mathematical models play a really important role in understanding the dynamics of physiological systems, and they can be used to make predictions about the future state of a patient. However, when attempting to model complicated, real-life dynamics mathematically, it becomes increasingly likely that the model will be incomplete. Furthermore, with the vast amount of patient data being recorded, only a small proportion of the data is consulted by clinicians for decision making.

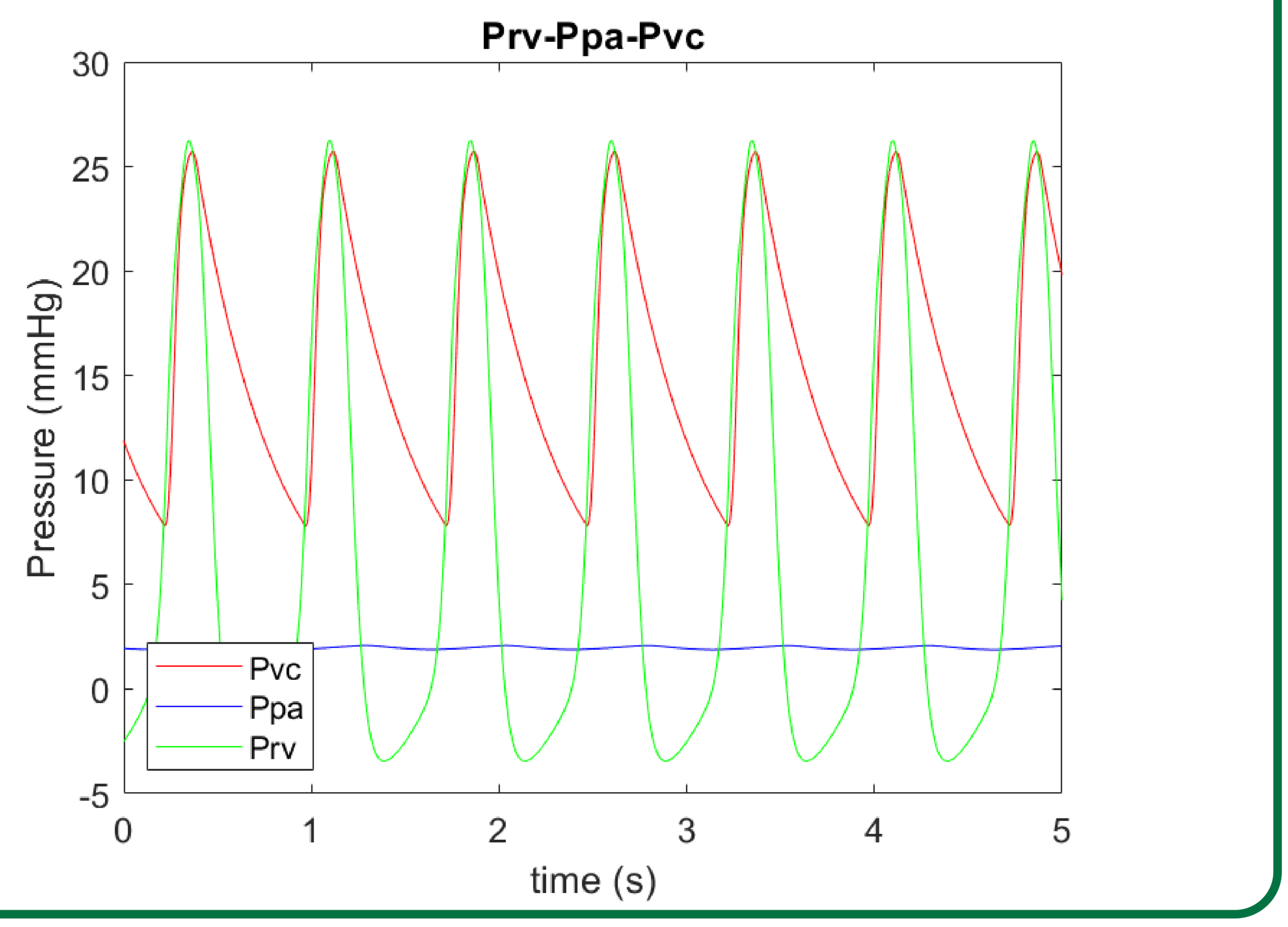
What are our objectives?

- Build hybrid "grey-box" models by 'plugging up' the incomplete parts of mathematical models with neural networks.
- Combine deep learning with symbolic regression to distil analytic expressions from data.

2. Pressure-Volume Model of Cardiovascular System

To capture the main dynamics of the cardiovascular system (CVS), a 10-dimensional system of ODEs representing a pressure-volume model of the CVS is simulated. The CVS is divided into 6 **elastic chambers** - the 10 state variables are the volumes of the 6 chambers and 4 of the 6 flows.

From these 10 state variables, other variables, such as the pressures in each of the chambers, can then be calculated. The figure below shows some of the key results.



3. Universal Differential Equations

Universal differential equations (UDEs) [1] are a hybrid **grey-box** modelling approach, where incomplete mathematical models are complemented using neural networks.

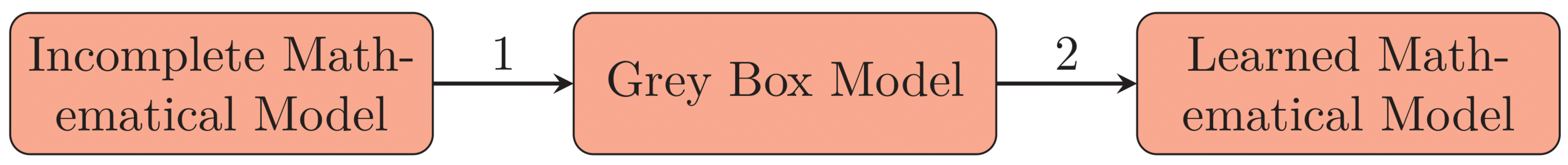
$$\frac{dx_1}{dt} = \alpha x_1 - \beta x_1 x_2 \qquad \frac{dx_2}{dt} = -\gamma x_2 + \delta x_1 x_2$$

$$\longrightarrow$$

$$\frac{dx_1}{dt} = \alpha x_1 - NN_1 \qquad \frac{dx_2}{dt} = -\gamma x_2 + NN_2$$

In the figure above, if the quadratic terms of the Lotka-Volterra model (in red) are unknown, they can be replaced with neural networks (in orange).

The neural networks are trained to capture the missing dynamics (arrow 1 in the figure below), and then **regressed** back down to symbolic form in order to learn the missing terms from the model (arrow 2).



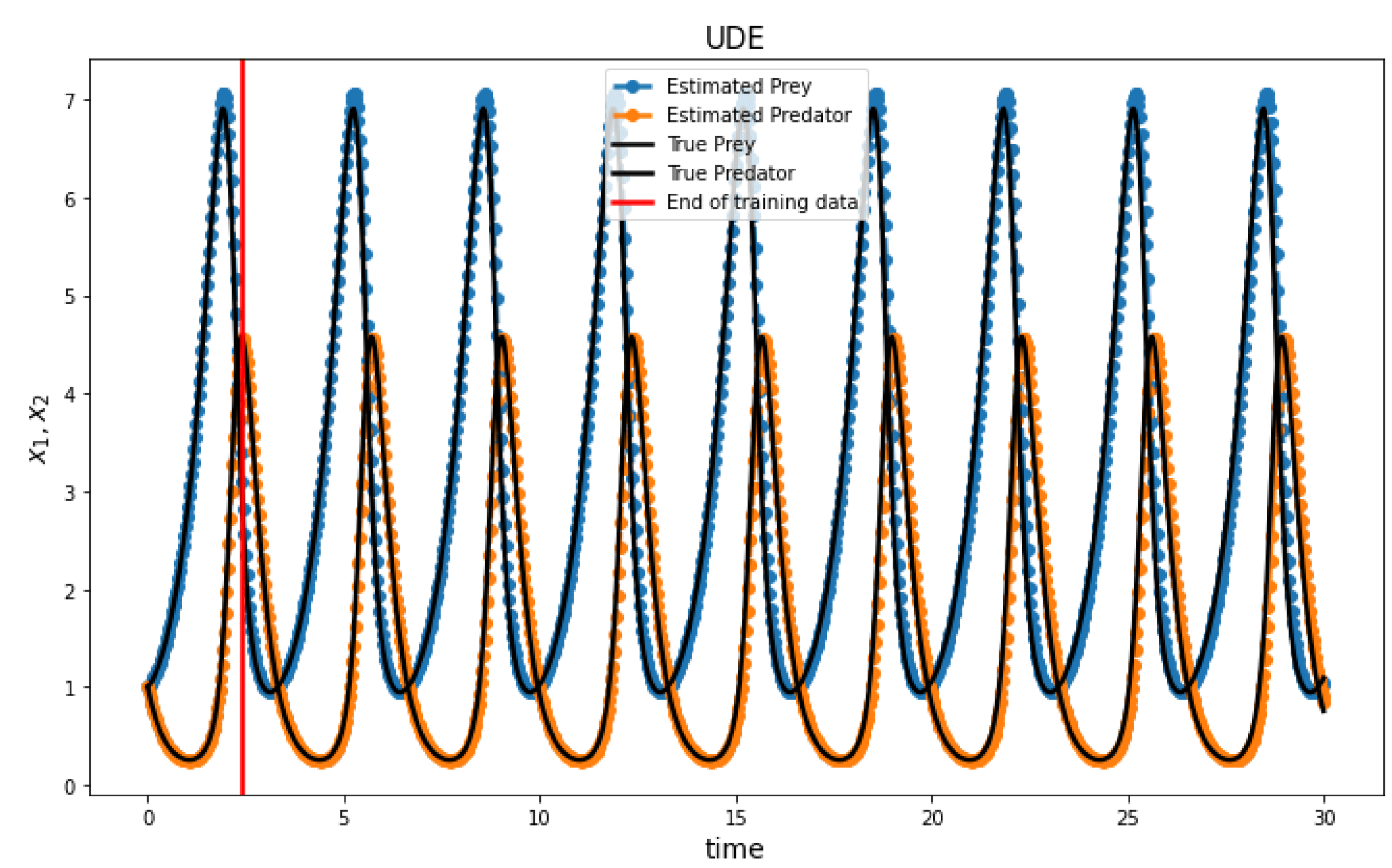
This method leverages the data efficiency of mathematical models and the accuracy of neural networks as universal function approximators, resulting in a **data efficient** tool for forecasting.

4. UDE Application

The UDE approach was applied to the Lotka-Volterra equations. As in section 3, the quadratic terms are assumed to be unknown and are replaced with neural networks.

The resulting grey-box model is trained on only 30 data points (highlighting the data efficiency of this method) and an **extrapolation** is simulated.

The trained neural networks are then regressed back to mechanistic form using SINDy, and the true quadratic terms are discovered. The figure below shows the improved extrapolation of this learned model.



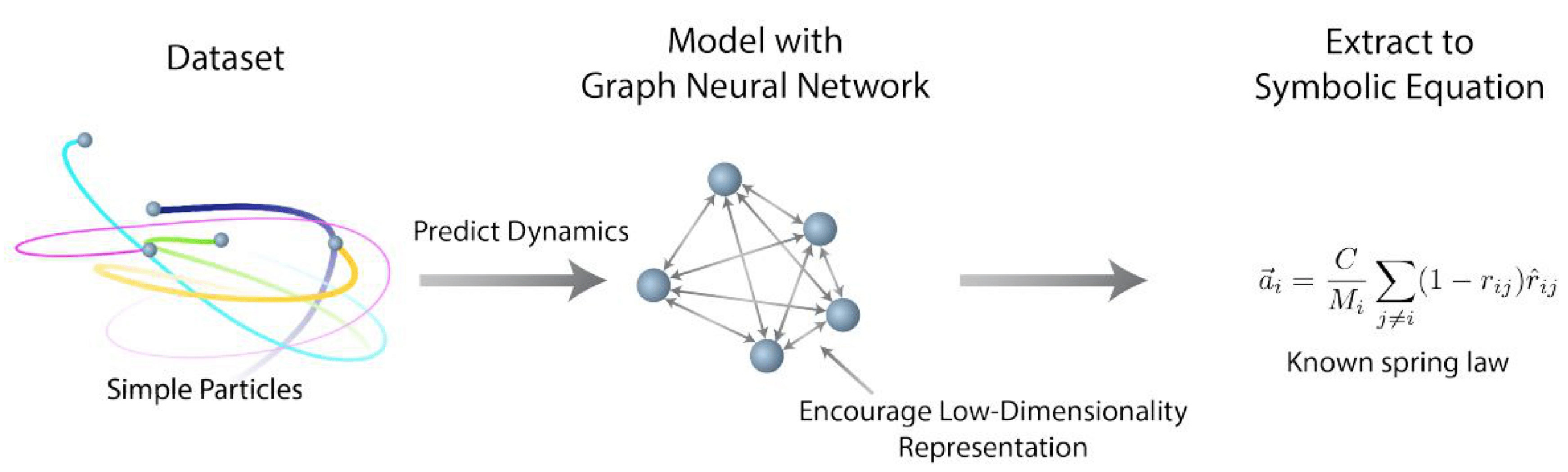
5. Deep Symbolic Regression

Symbolic regression is a machine learning technique that fits analytic expressions to data (through genetic programming). However, this method is intractable in high dimensions.

Deep learning models are great at high dimensional problems but don't generalise well, which precisely complements symbolic regression.

By using a neural network to learn a **low-dimensional representation** of high dimensional data, it then becomes tractable to apply symbolic regression to the output of the neural network.

The figure below shows the outline of the method.



6. References

1. **Chris Rackauckas et al.** - Universal Differential Equations for Scientific Machine Learning (2020).
2. **Miles Cranmer et al.** - Discovering Symbolic Models from Deep Learning with Inductive Biases (2020).

7. Future Work

- Apply one or both of these methods to the cardiovascular model or another physiological model.
- Apply deep symbolic regression to clinical data.