Data-Driven Dynamics:

Learning Vector Fields for Epidemic Forecasting with Neural ODEs

Vasilis Tsilidis, Department of Mathematics, University of Patras

PCMA-2025

20 December, 2025

Mechanistic

First principles $\Rightarrow \begin{cases} \frac{\mathrm{d}S}{\mathrm{d}t} &= -\frac{\beta S I}{N} \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta S I}{N} \end{cases}$

β, $\frac{\beta S I}{N}$ correspond to real processes $\Rightarrow$ Allows for explanation and verification of assumptions

Unknown mechansims or modeling simplifications $\Rightarrow$ High bias

Data Driven

Data $\Rightarrow$ $f(\mathbf{x}) = \sigma\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right)$

Availability of high-quality data $\Rightarrow$ Capable of capturing complex, nonlinear relations

Overreliance on data $\Rightarrow$ No intrinsic way of explaing mechanisms

Hybrid

$$\frac{\mathrm{d}z}{\mathrm{d}t} = f_\theta(z(t), t), \quad \text{where} \quad f_\theta(\mathbf{x}) = \sigma\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right)$$

Agenda Overview

A Theory

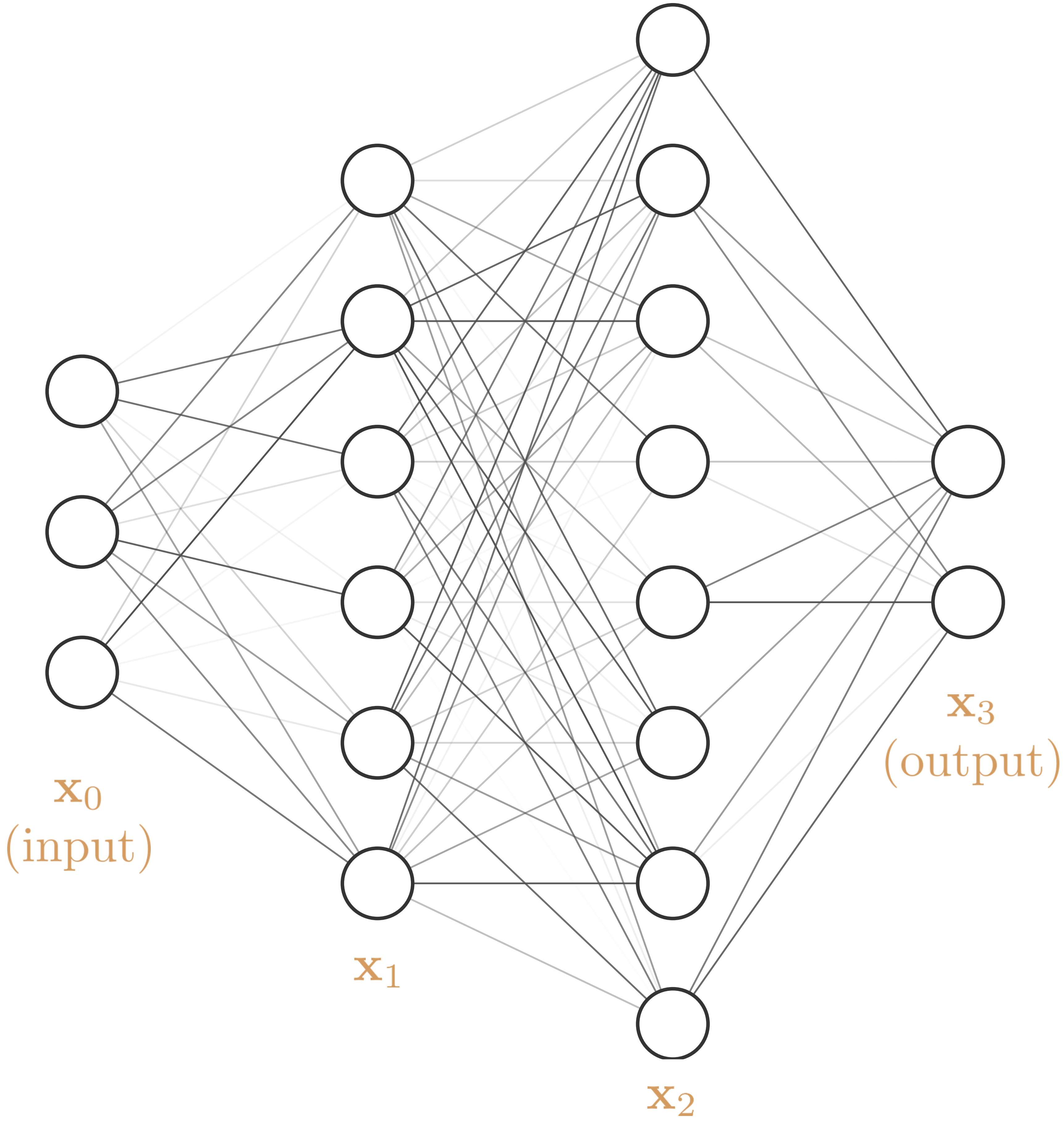
- A1 MLP
- A2 ResNet
- A3 The Bridge
- A4 Neural ODEs
- A5 Intuition

B Epidemic Forecasting

- B1 Data
- B2 Forecasting
- B3 Interpolation
- B4 Results
- B5 Takeaways

A Theory

A1 The Multilayer Perceptron (MLP)



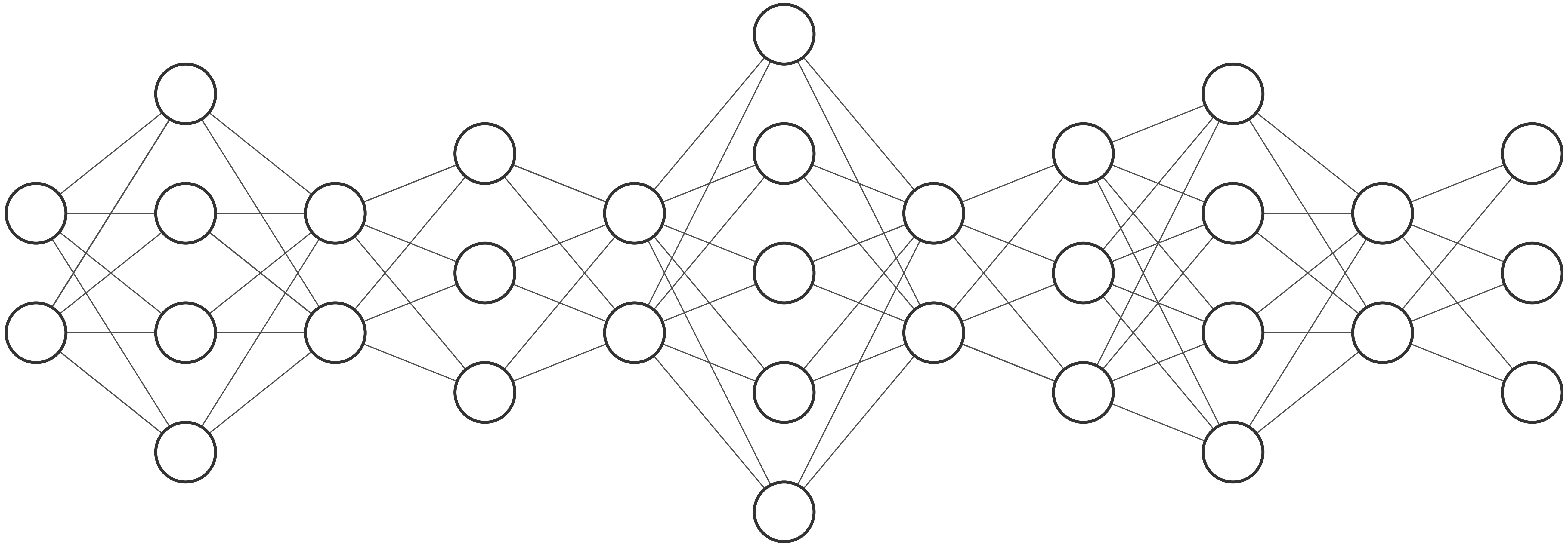
$$\mathbf{x}_{t+1} = \underbrace{\sigma\left(\mathbf{W}_t\mathbf{x}_t + \mathbf{b}_t\right)}_{\Vert}, \; 0 \leq t \leq L-1$$

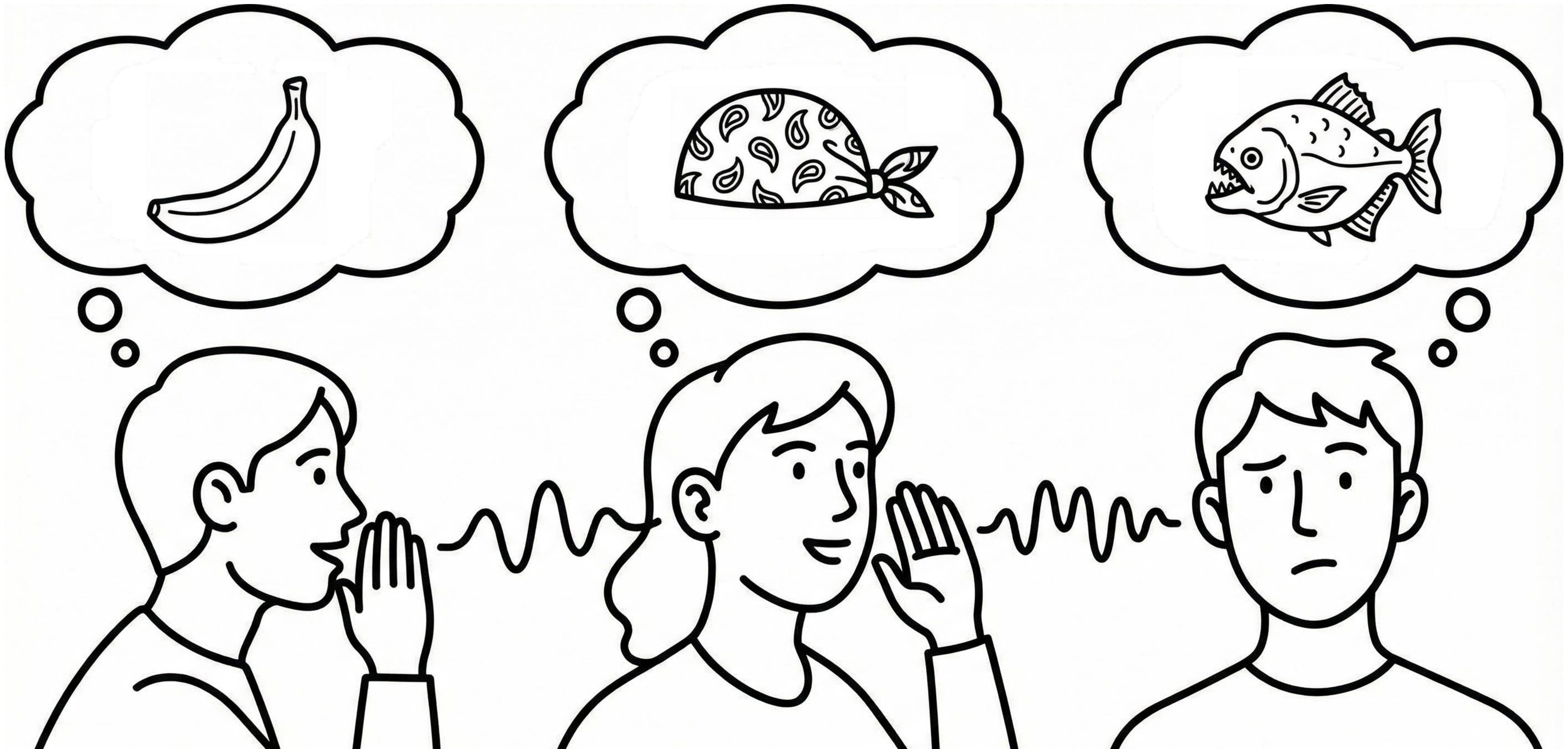
Where: $g_t(\mathbf{x}_t)$

- $L+1$ is the number of layers
- $\mathbf{x}_t = \begin{cases} \text{input vector}, & t = 0 \\ \text{hidden vector}, & 0 < t < L\,, \\ \text{output vector}, & x = L \end{cases}$
- $\mathbf{W}_t, \mathbf{b}_t$ are learnable parameters
- σ is a nonlinear function

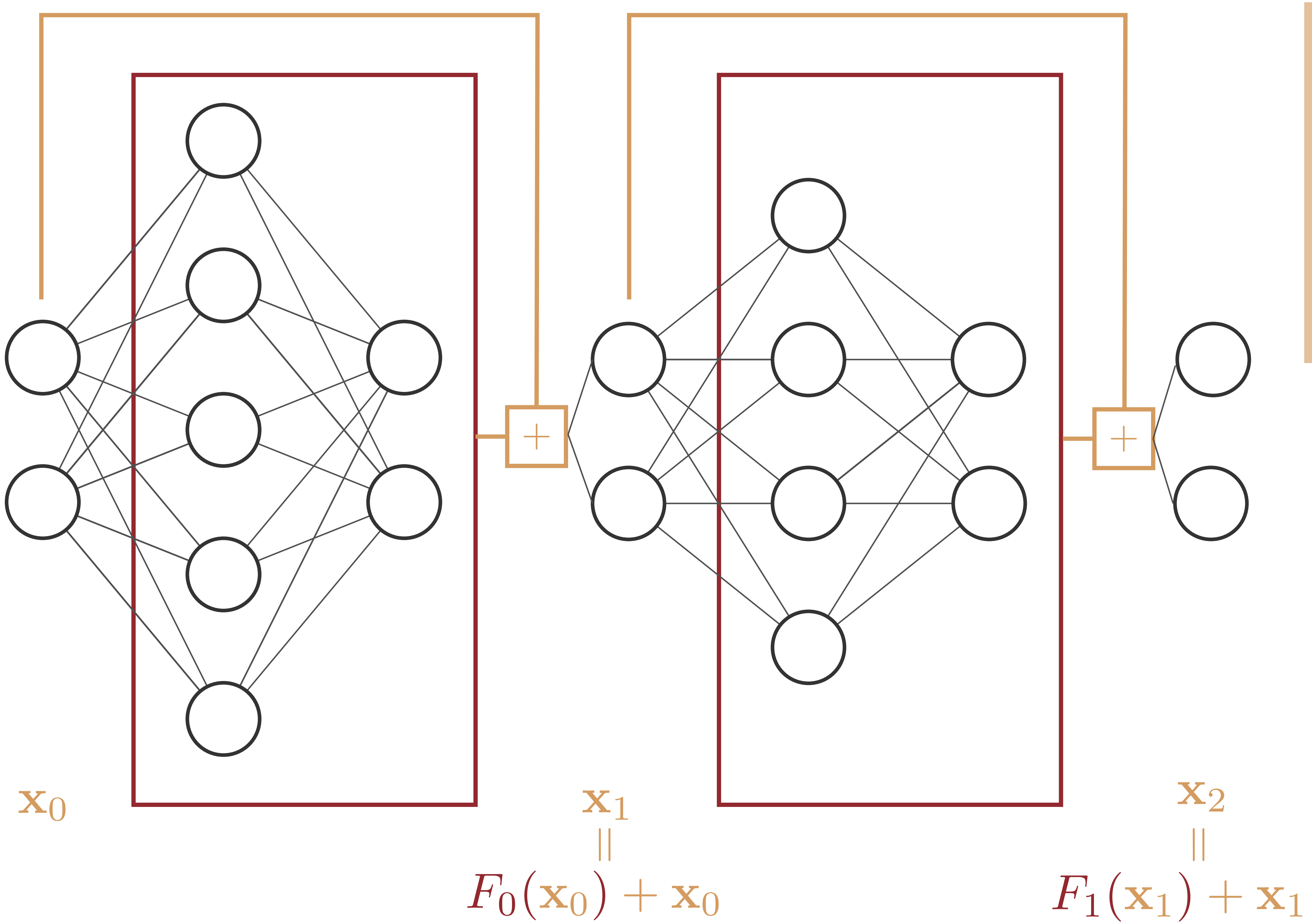
For the output to be produced, an input vector is passed through a sequence of discrete layers.
The output of each layer is recursively produced, using the output of the previous layer as input.
Each layer applies an affine transformation followed by a nonlinear transformation.
Hence, an MLP is a composition of functions:

$$f(\mathbf{x}) = g_{L-1} \circ g_{L-2} \circ \cdots \circ g_0(\mathbf{x})$$





A2 The Residual Network (ResNet)



$$\mathbf{x}_{t+1} = F_t(\mathbf{x}_t) + \mathbf{x}_t, \; 0 \le t \le L-1$$

Where:

- $F_t : \mathbb{R}^d \mapsto \mathbb{R}^d$ are MLPs
- $\mathbf{x}_t \in \mathbb{R}^d$ are state vectors

The skip connection is modeled through a small algebraic change in the recurrence relation. Instead of learning the new state from scratch, ResNet learns the additive change to the current state. This allows the information to propagate easier.

A3 The Bridge

Euler Method

To solve the IVP $\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,t),\ x(0) = x_0$ we discretize using Euler's method:

$$x_{t+1} = x_t + h \cdot f(x_t, t),$$

- y_t: state at step t.
- h: step size.
- f: vector field.

$$\overset{h=1}{\Longleftrightarrow}$$

ResNet

To transform input x_0, we apply layers:

$$x_{t+1} = x_t + F(x_t),$$

- x_t: state at layer t.
- F: MLP.

$$\xrightarrow{h \to 0}$$

$$\xrightarrow{\text{more layers}}$$

$$x(t) = x_0 + \int_0^T f\left(x(s), s\right)\,\mathrm{d}s$$

A4 Neural ODEs

$$\begin{cases} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = f(\mathbf{x}(t), t) \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases},$$

Where:

- $\mathbf{x}_0$ is the input data
- f is an MLP
- t is a continious depth variable

The Forward Pass

The output of the model is the trajectory of the solution of the IVP. The computation is performed using any standard numerical ODE solver

Adaptive Computation

Unlike ResNets the solver can adapt the step size based on the complexity of the trajectory (error tolerance)

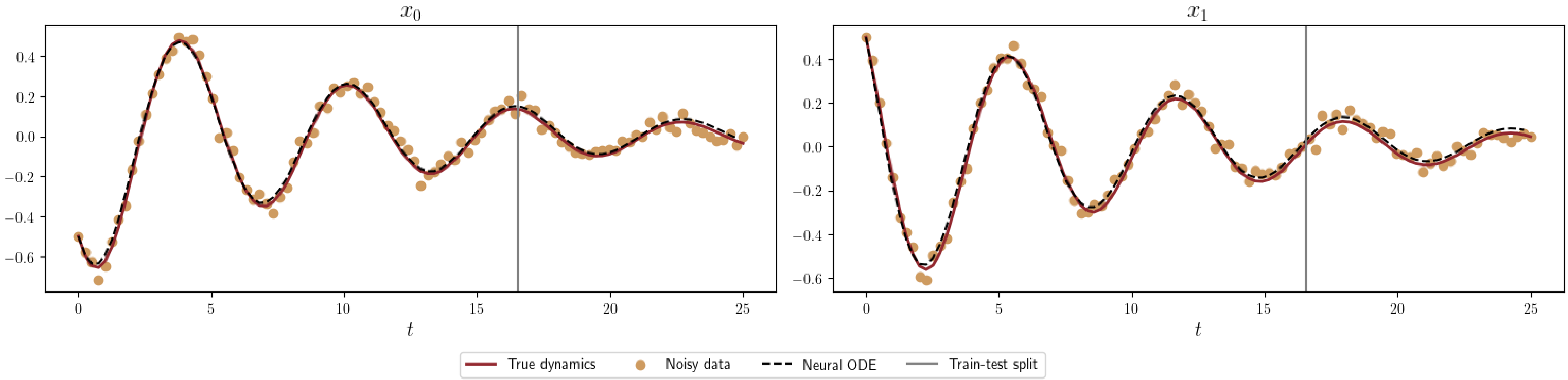
Topology Preservation

Assuming the MLP is Lipschitz continuous, the trajectories cannot cross (uniqueness), preserving the topology of the input space

A5

Intuition - The Stable Spiral

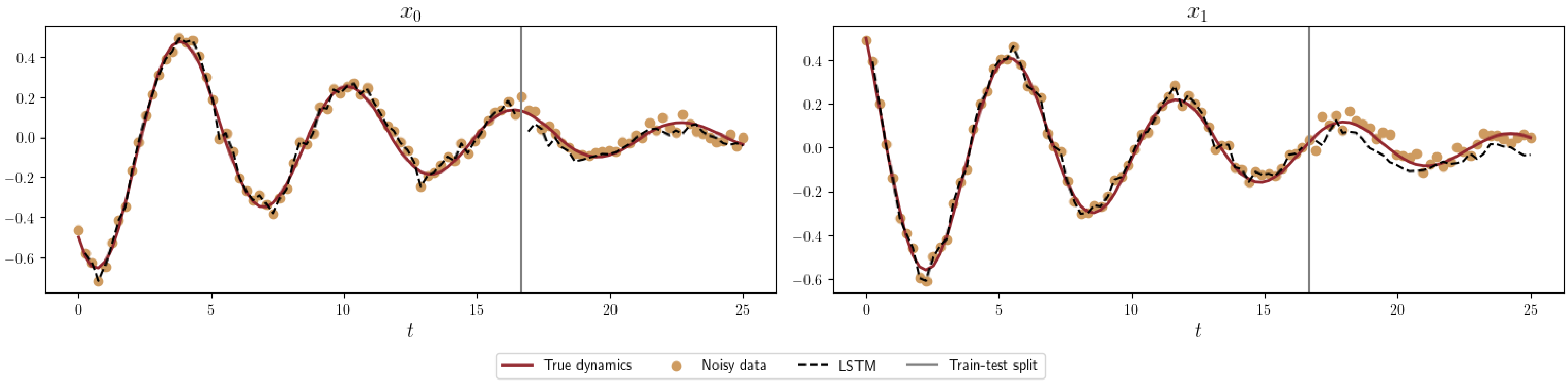
$$\begin{cases} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \begin{pmatrix} -0.1 & 1 \\ 1 & -1 \end{pmatrix} \mathbf{x} \\ \mathbf{x}(0) = \begin{pmatrix} -0.5 & .5 \end{pmatrix}^\top \end{cases}$$



The neural ODE learns the vector field of a stable spiral.

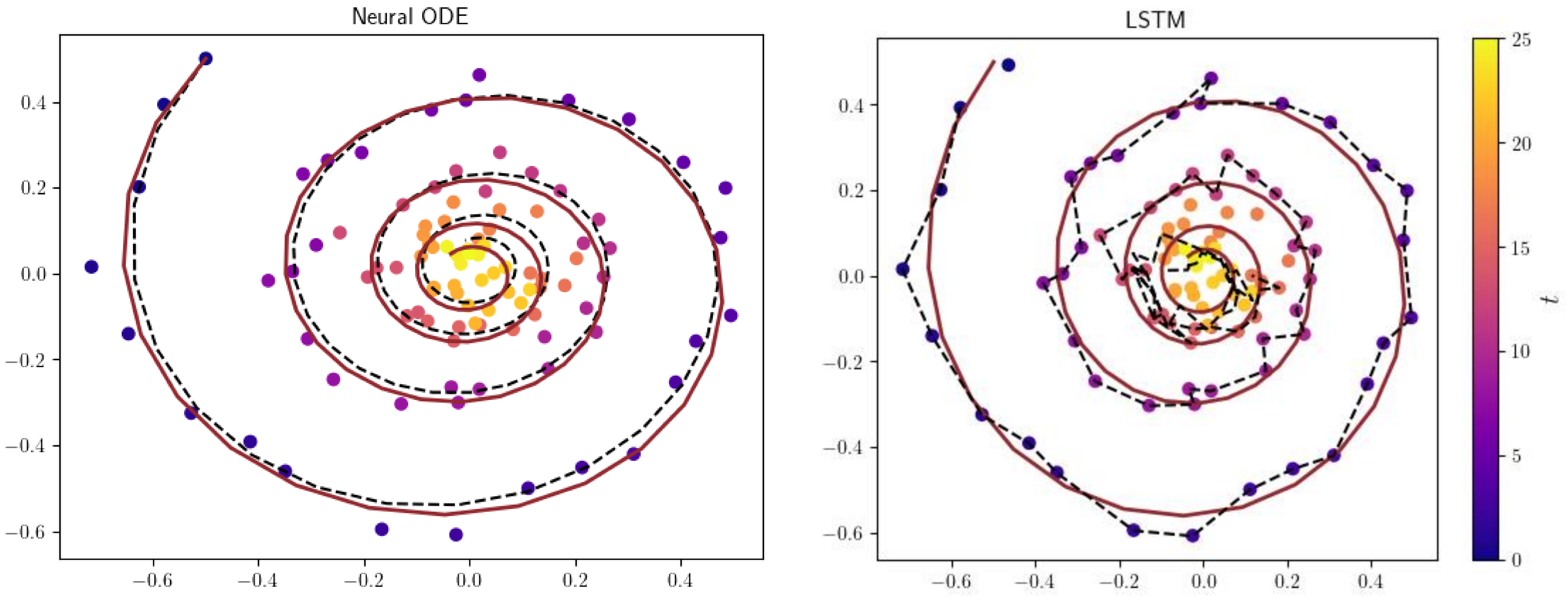
The trajectory is smooth, continuous, and physically plausible.

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \begin{pmatrix} -0.1 & 1 \\ 1 & -1 \end{pmatrix} \mathbf{x} \\ \mathbf{x}(0) = \begin{pmatrix} -0.5 & .5 \end{pmatrix}^\top \end{cases}$$



The LSTM tries to predict the next point of a stable spiral.

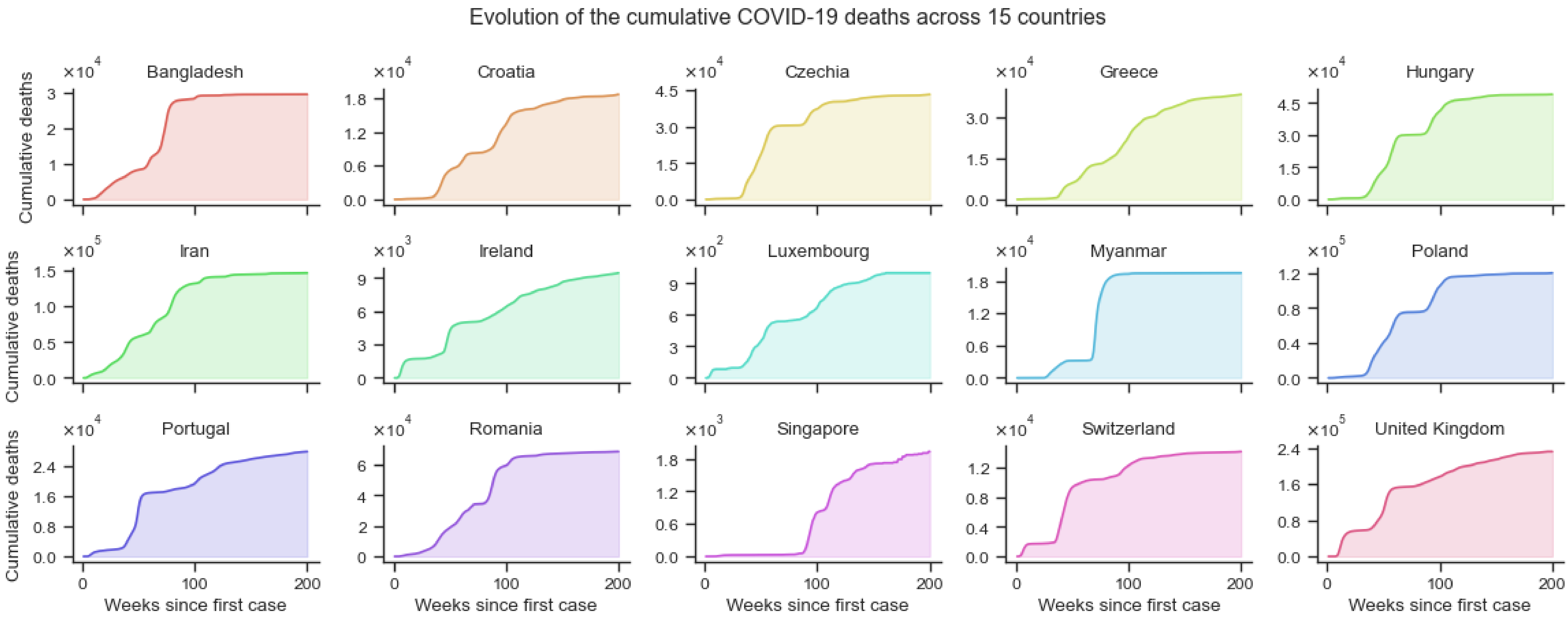
The trajectory is “jagged” and fails to capture the smooth dynamics.



Neural ODE
LSTM
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.4
−0.2
0.0
0.2
0.4
25
20
15
10
5
0
t

B Epidemic Forecasting

B1 **Data**



The time series for each country has been divided into four datasets, each containing 25%, 50%, 75%, or 100% of the full time series.

B2 Models

ANODE

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = f_\theta(\mathbf{z}(t))$$

Where:

$$\mathbf{z}(t) = [D(t), \mathbf{h}(t)]$$

Seasonal ANODE

$$\frac{\mathrm{d}\mathbf{z}}{\mathrm{d}t} = f_\theta(\mathbf{z}(t), \mathbf{s}(t))$$

Where:

$$\mathbf{s}(t) = \bigoplus_{k=1}^{K} \begin{bmatrix} \sin(\omega_k t + \phi_k) \\ \cos(\omega_k t + \phi_k) \end{bmatrix}$$

SIRD

$$\begin{cases} \frac{\mathrm{d}S}{\mathrm{d}t} & = -\beta \frac{IS}{N} \\ \frac{\mathrm{d}I}{\mathrm{d}t} & = \beta \frac{IS}{N} - \gamma I - \mu I \\ \frac{\mathrm{d}R}{\mathrm{d}t} & = \gamma I \\ \frac{\mathrm{d}D}{\mathrm{d}t} & = \mu I \\ N & = S + I + R + D \end{cases}$$

SIRD w/ reinfection

$$\begin{cases} \frac{\mathrm{d}S}{\mathrm{d}t} & = -\beta \frac{IS}{N} + \delta R \\ \frac{\mathrm{d}I}{\mathrm{d}t} & = \beta \frac{IS}{N} - \gamma I - \mu I \\ \frac{\mathrm{d}R}{\mathrm{d}t} & = \gamma I - \delta R \\ \frac{\mathrm{d}D}{\mathrm{d}t} & = \mu I \\ N & = S + I + R + D \end{cases}$$

SEIRD

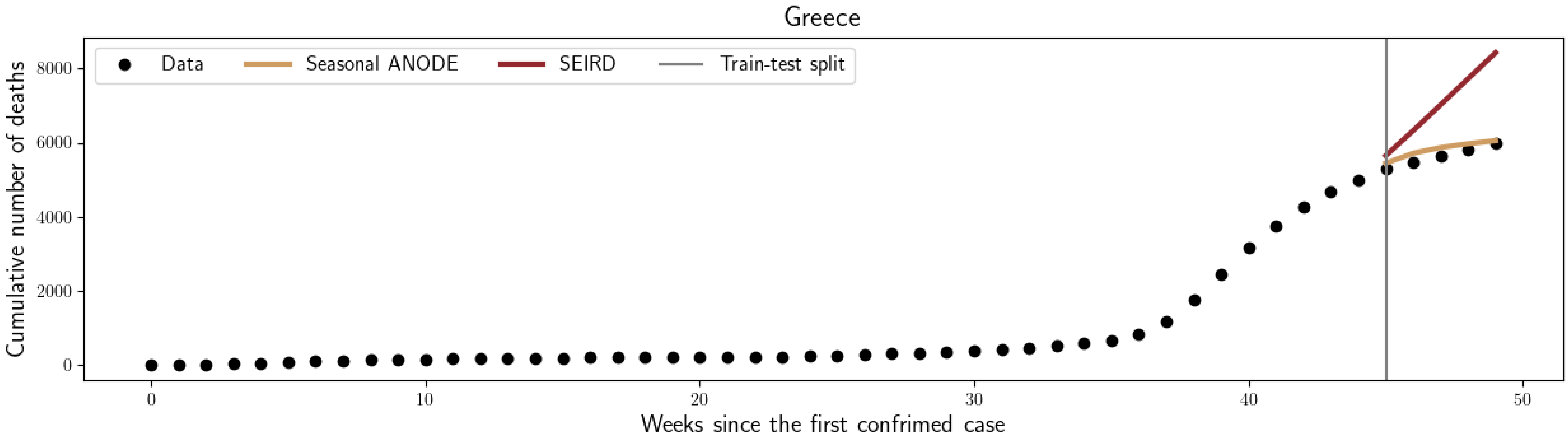
$$\begin{cases} \frac{\mathrm{d}S}{\mathrm{d}t} & = -\beta \frac{IS}{N} \\ \frac{\mathrm{d}E}{\mathrm{d}t} & = \beta \frac{IS}{N} - \sigma E \\ \frac{\mathrm{d}I}{\mathrm{d}t} & = \sigma E - \gamma I - \mu I \\ \frac{\mathrm{d}R}{\mathrm{d}t} & = \gamma I \\ \frac{\mathrm{d}D}{\mathrm{d}t} & = \mu I \\ N & = S + E + I + R + D \end{cases}$$

SEIRD w/ vital dynamics

$$\begin{cases} \frac{\mathrm{d}S}{\mathrm{d}t} = \Lambda - \dfrac{\beta SI}{N} - \mu S \\ \frac{\mathrm{d}E}{\mathrm{d}t} = \dfrac{\beta SI}{N} - \sigma E - \mu E \\ \frac{\mathrm{d}I}{\mathrm{d}t} = \sigma E - (\gamma + \delta + \mu) I \\ \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I - \mu R \\ \frac{\mathrm{d}D}{\mathrm{d}t} = \delta I \end{cases}$$

B3

Forecasting

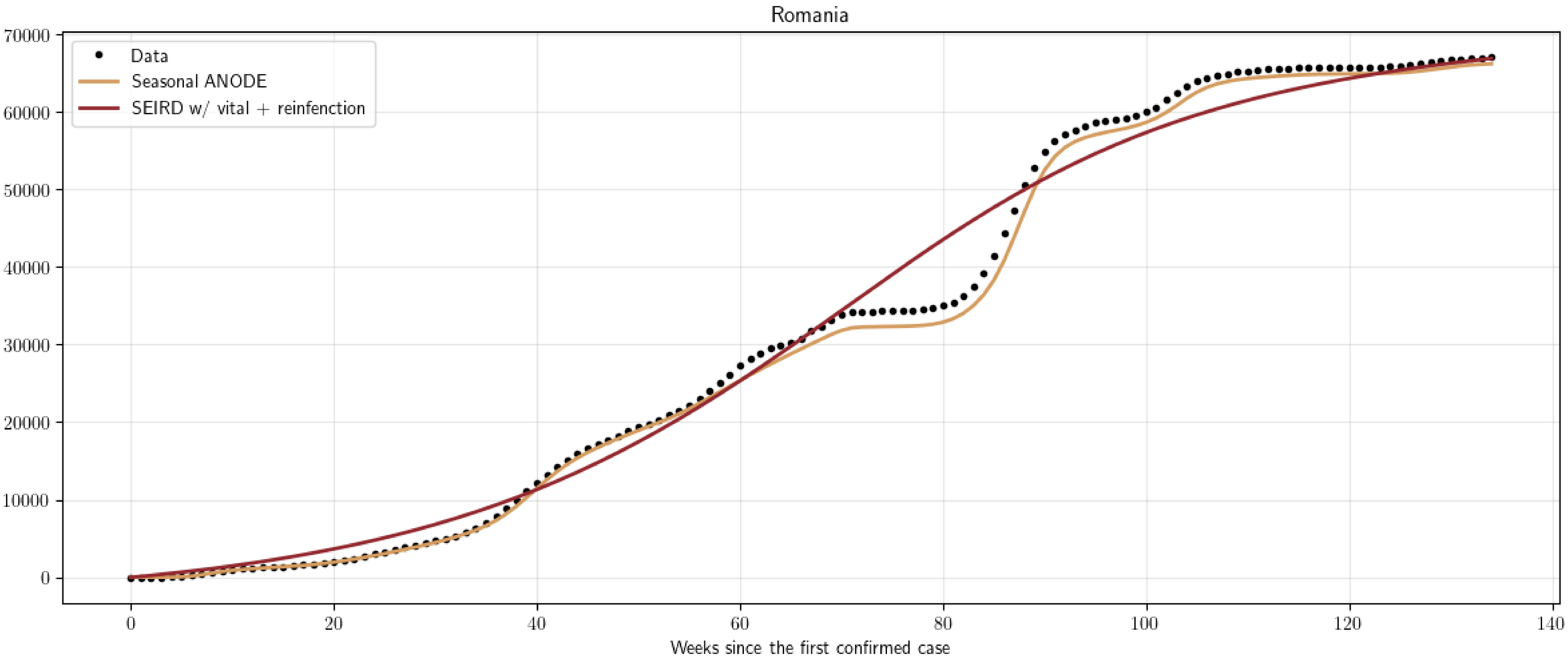


The SEIRD model thinks the number of deaths is still increasing.

The Neural ODEs model successfully captures the slow-down in deaths.

B4

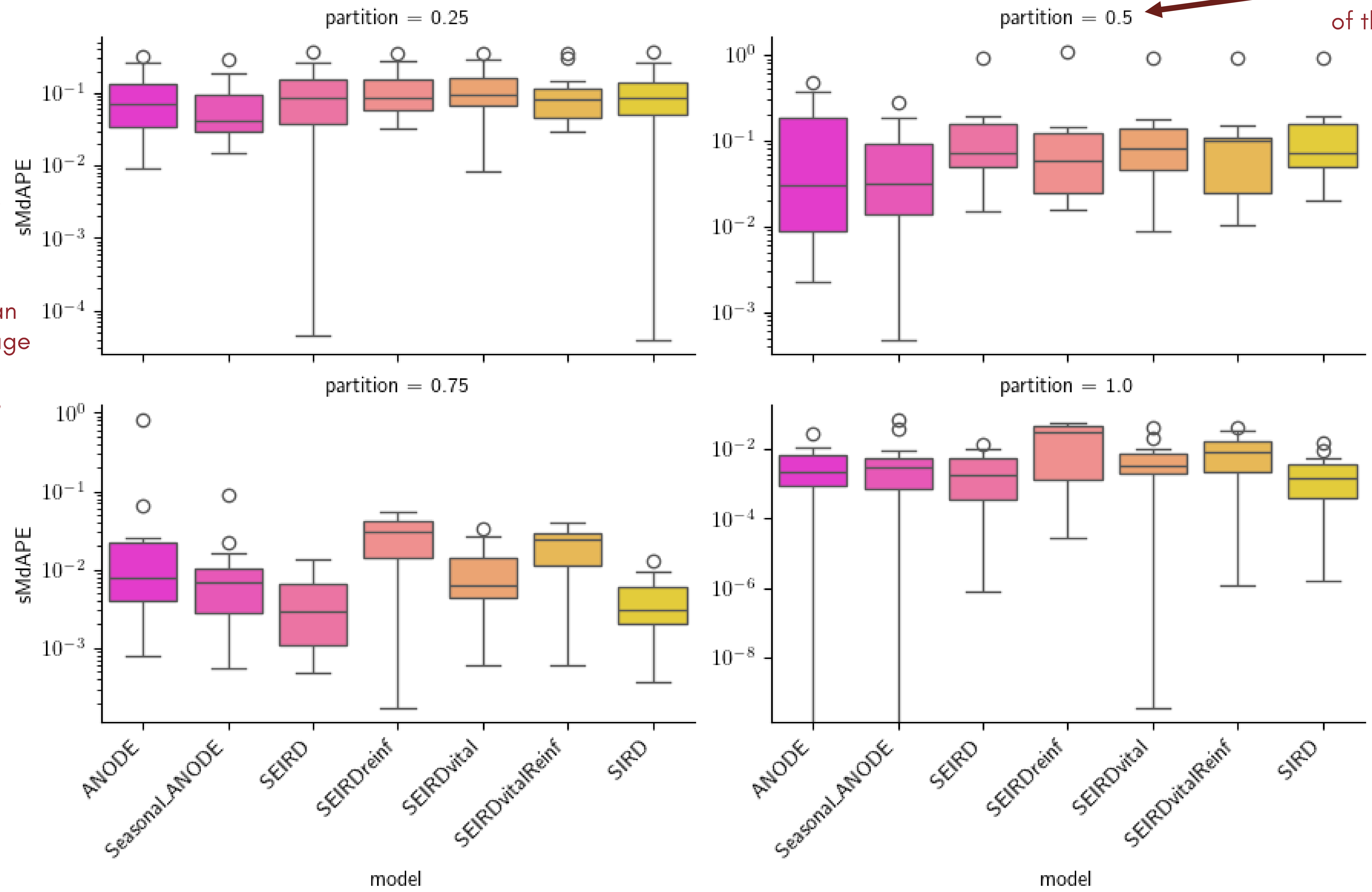
Interpolation



B5 Results

Percentage of the partition of the full time series

Symmetric median absolute percentage error for all countries



B6 Takeaways

	Classical ODEs	Neural ODEs
Forecasting	Fail to capture complex waves or policy shifts without manual tweaking	Can easily learn time-varying dynamics and complex interactions
Computational time	Simple functional forms means fast integration	Requires many function evaluations (NFE) during the solver steps and possibly hyperparameter tuning

UNIVERSITY OF PATRAS
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

Thank You!