Robust parameter estimation and identifiability analysis with Hybrid Neural Ordinary Differential Equations in Computational Biology

Stefano Giampiccolo, Federico Reali, Anna Fochesato, Giovanni Iacca, Luca Marchetti

S1 file: Supplementary notes, tables and figures

Supplementary Note 1: Relationship between different definitions of parameter identifiability

The problem of parameter identifiability is a long-standing challenge in the field of system identification[1,2], and several definitions have been proposed. Here, we focus on the notion of structural identifiability, since practical identifiability is less clearly defined in the literature[3]. In this section, we aim to compare the definition of parameter at-a-point local identifiability used in our work with the established definition of analytical local identifiability[4]. For a comparison between the analytical definition and other definitions, we refer to the work of Anstett-Collin *et al.*[4].

We consider a continuous-time model defined, in the time interval $[0,T]$, by the following ordinary differential equation:

$$\begin{cases} \frac{d\mathbf{y}(\mathbf{t},\theta)}{dt} = f(\mathbf{y}(t),\theta) \\ \mathbf{y}(0,\theta) = \mathbf{y}_0 \end{cases} \tag{1}$$

where $\mathbf{y}(t) \in \mathbb{R}^n$ is the state vector, $\theta \in \mathbb{R}^m$ is the vector of system parameters ranging in the parameter space $\Theta \subseteq \mathbb{R}^m$, $\mathbf{y}_0 \in \mathbb{R}^n$ is the initial state. We denote with θ^i the i-th parameter of the model, i.e. the i-th component of the parameter vector θ. To simplify the mathematical notation compared to the most general case[4], we assume here that the system does not depend on external inputs and that all variables of the system are observable. Furthermore, when defining identifiability, we require that the equalities hold for all the parameter values, rather than almost everywhere with respect to a measure on the parameter space, as in the most general case[4].

In this context, the following is the analytical definition of parameter local structural identifiability:

Definition 1 (**analytical local identifiability**)**.** The parameter θ^i is *locally structurally identifiable* if for all $\theta \in \Theta$

$$\mathbf{y}(t,\hat{\theta}) = \mathbf{y}(t,\theta) \qquad \forall t \in [0,T]$$

implies $\hat{\theta}^i = \theta^i$ for all $\hat{\theta}$ in a suitable neighborhood of θ.

Given a set of observational time points $t_1,\dots,t_k \in [0,T]$, the definition of parameter local identifiability at-a-point used in our work can be stated as follows:

Definition 2 (**local identifiability at-a-point**)**.** The parameter θ^i is *locally identifiable at-a-point* in θ_0 if

$$\mathbf{y}(t_i,\hat{\theta}) = \mathbf{y}(t_i,\theta_0) \quad \forall i \in 1,\dots,k$$

implies $\hat{\theta}^i = \theta^i_0$ for all $\hat{\theta}$ in a suitable neighborhood of θ_0.

If a parameter is analytically locally identifiable, then it is also identifiable at-a-point for every choice of observational time points and at every point in the parameter space. However, the converse is not true: a parameter θ^i could be identifiable at-a-point θ_0 without being structurally locally identifiable. The main reasons for this are the following:

- the observational time points $t_1,\dots,t_k$ could be not sufficient to detect any differences in the trajectories produced by different parameters (that could be visible in other times in the interval $[0,T]$).
- the parameter is locally identifiable at-a-point in some points of Θ, but not in other regions of the parameter search space.

Supplementary Note 2: Lotka-Volterra test case: experimental setting

To simulate in silico observations for the Lotka-Volterra test case, the ODE system has been solved numerically employing the *Vern7* solver from the *DifferentialEquations.jl* library. The solver settings are configured with an absolute tolerance of 10^{-7} and a relative tolerance of 10^{-6}. The parameters and initial states, derived from the values used by Rackauckas et al.[5], are detailed in Supplementary Table 1 and 2.

Parameter	Value	Unit
α	1.3	year^{-1}
β	0.9	year^{-1}
γ	0.8	year^{-1}
δ	1.8	year^{-1}

Supplementary Table 1. Lotka-Volterra test case: parameters employed for generating the *in silico* dataset

Variable	Value
y_1	3.416
y_2	1.537

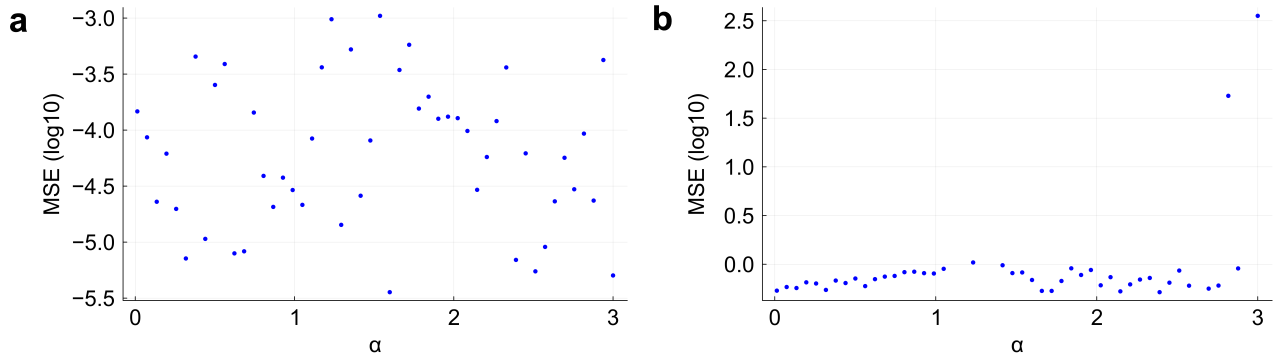
Supplementary Table 2. Lotka-Volterra - initial states employed for generating the *in silico* dataset

Supplementary Note 3: Lotka-Volterra test case: tuned hyper-parameters

Name	Search Space	Tuned value $DS_{0.00}$	Tuned value $DS_{0.05}$
Neural network architecture			
hidden layer number	$\{2,3\}$	2	3
hidden layer width	$\{4,8,16,32\}$	16	8
Starting values for mechanistic parameters			
α	$[1.30\cdot10^{-2}, 1.30\cdot10^{2}]$	$9.85\cdot10^{-1}$	$5.81\cdot10^{-1}$
Training hyper-parameters			
learning rate	$[10^{-5}, 10^{-1}]$	$3.12\cdot10^{-2}$	$7.10\cdot10^{-3}$
λ	$\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$	ND	1.0
k (segmentation)	$\{2,3,4,5,6,7,8,9,10\}$	2	5
ρ	$[10^{-3}, 10^{3}]$	$2.61\cdot10^{-3}$	6.52

Supplementary Table 3. Lotka-Volterra test case: tuned hyperparameters of the HNODE model.

Supplementary Note 4: Lotka-Volterra test case: attempt to retrieve the identifiability of α by regularizing the cost function



Supplementary Figure 1. Mean Squared Error of HNODE models trained with fixed α in the Lotka-Volterra test case. Mean Squared Errors (MSE) of the HNODE model trained with varying fixed α on $DS_{0.00}$ and $DS_{0.05}$ (panels **a** and **b** respectively). On the x-axis, values of α are sampled in the interval $[0.013, 3.0]$. On the y-axis, the MSE of the dynamics predicted by the HNODE models with fixed α with respect to the observed data is depicted. In dataset $DS_{0.05}$, the MSE has been computed normalizing for the uncertainty. The correlation between α and *MSE* is very weak in $DS_{0.00}$ (-0.037) and weak in $DS_{0.05}$ (0.293).

Supplementary Note 5: Lotka-Volterra test case: impact of ε and δ on the identifiability analysis results

To assess the impact of varying hyperparameters ε and δ on the identifiability analysis, we conducted the analysis using different values for these parameters (Supplementary Table 4). Our findings indicate that the results exhibit robustness against changes in ε and δ.

			$\varepsilon = 10^{-6}$			$\varepsilon = 10^{-5}$			$\varepsilon = 10^{-4}$	
Par.	**Dataset**	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.05$	$\delta = 0.07$
α	$DS_{0.00}$								
	$DS_{0.05}$								

Supplementary Table 4. Lotka-Volterra test case: parameter identifiability with different choices of ε and δ. The symbol ✓ denotes the parameters classified as identifiable with the ε and δ threshold specified in the header (the absence of a ✓ in the table indicates that the parameter is always classified as non-identifiable)..

Supplementary Note 6: Cell apoptosis test case: experimental setting

To simulate in silico observations, the cell apoptosis ODE system is solved numerically employing the *TRBDF2* solver from the *DifferentialEquations.jl* library. The solver settings are configured with an absolute tolerance of 10^{-7} and a relative tolerance of 10^{-6}. The parameters and initial states, derived from the values employed by Aldridge at al.[6], are detailed in Supplementary Table 5 and 6.

Parameter	Value	Unit
k_1	0.9612	$\text{cell} \cdot (\text{h} \cdot 10^5 \cdot \text{molecules})^{-1}$
k_{d1}	36.0	h^{-1}
k_{d2}	28.8	h^{-1}
k_3	24.48	$\text{cell} \cdot (\text{h} \cdot 10^5 \cdot \text{molecules})^{-1}$
k_{d3}	180.0	h^{-1}
k_{d4}	3.6	h^{-1}
k_5	25200.0	$\text{cell} \cdot (\text{h} \cdot 10^5 \cdot \text{molecules})^{-1}$
k_{d5}	0.06012	h^{-1}
k_{d6}	0.6012	h^{-1}

Supplementary Table 5. Cell apoptosis test case: parameters employed for generating the *in silico* dataset

Variable	Value	Unit
y_1	1.34	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_2	1.0	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_3	2.67	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_4	0.0	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_5	0.0	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_6	0.0	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_7	0.029	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$
y_8	0.0	$10^5 \cdot \text{molecules} \cdot \text{cell}^{-1}$

Supplementary Table 6. Cell apoptosis test case: initial states employed for generating the *in silico* dataset

The cell apoptosis ODEs, characterized by the previously specified parameters and initial values, is a stiff system within the considered time interval. Supplementary Table 7 presents an analysis of the stiffness index of the ODE system at various time points along the trajectory[7]. The stiffness index S for the interval (t_1, t_2) is calculated as

$$S = \frac{|Re(\overline{\lambda})|}{|Re(\underline{\lambda})|} \cdot (t_2 - t_1),$$

where $\overline{\lambda}$ and $\underline{\lambda}$ represent the eigenvalues of the Jacobian matrix of the ODE at t_1 with the maximum and minimum absolute values of their real parts, respectively.

Time interval	(0.0, 0.8)	(0.8, 1.6)	(1.6, 2.4)	(2.4, 3.2)	(3.2, 4.0)
S	∞	$1.39 \cdot 10^{22}$	$2.75 \cdot 10^{22}$	∞	∞
$\lvert Re(\overline{\lambda})\rvert$	$7.32 \cdot 10^{2}$	$1.51 \cdot 10^{4}$	$2.99 \cdot 10^{4}$	$4.27 \cdot 10^{4}$	$5.21 \cdot 10^{4}$
$\lvert Re(\underline{\lambda})\rvert$	0.0	$8.67 \cdot 10^{-19}$	$8.67 \cdot 10^{-19}$	0.0	0.0

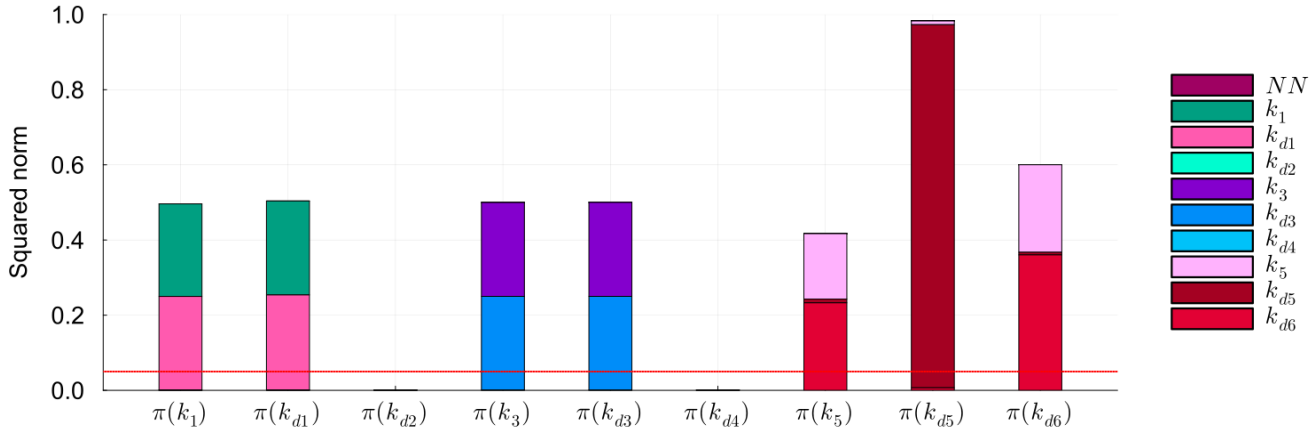
Supplementary Table 7. Analysis of stiffness in the cell apoptosis test case. The table displays the stiffness index (S), along with the absolute values of the eigenvalues $\overline{\lambda}$ and $\underline{\lambda}$ ($|Re(\overline{\lambda})|$ and $|Re(\underline{\lambda})|$, respectively) at various time points along the trajectory.

Supplementary Note 7: Cell apoptosis test case (first scenario): tuned hyper-parameters

Name	Search Space	Tuned value $DS_{0.00}$	Tuned value $DS_{0.05}$
Neural network architecture			
hidden layer number	$\{2,3,4,5,6\}$	2	2
hidden layer width	$\{4,8,16,32\}$	8	8
Starting values for mechanistic parameters			
k_1	$[9.61\cdot10^{-3}, 9.61\cdot10^{1}]$	$3.30\cdot10^{1}$	$7.30\cdot10^{1}$
k_{d1}	$[3.60\cdot10^{-1}, 3.60\cdot10^{3}]$	$1.12\cdot10^{3}$	$2.87\cdot10^{3}$
k_{d2}	$[2.88\cdot10^{-1}, 2.88\cdot10^{3}]$	$4.77\cdot10^{1}$	$1.51\cdot10^{3}$
k_3	$[2.45\cdot10^{-1}, 2.45\cdot10^{3}]$	$1.67\cdot10^{2}$	$2.05\cdot10^{3}$
k_{d3}	$[1.80, 1.80\cdot10^{4}]$	$2.51\cdot10^{3}$	$9.08\cdot10^{3}$
k_{d4}	$[3.60\cdot10^{-2}, 3.60\cdot10^{2}]$	$2.06\cdot10^{2}$	$8.41\cdot10^{1}$
k_5	$[2.52\cdot10^{2}, 2.52\cdot10^{6}]$	$2.37\cdot10^{6}$	$3.92\cdot10^{4}$
k_{d5}	$[6.01\cdot10^{-4}, 6.01]$	$6.30\cdot10^{-1}$	3.58
k_{d6}	$[6.01\cdot10^{-3}, 6.01\cdot10^{1}]$	$4.41\cdot10^{1}$	$1.73\cdot10^{1}$
Training hyper-parameters			
learning rate	$[10^{-5}, 10^{-1}]$	$3.82\cdot10^{-2}$	$2.01\cdot10^{-2}$
λ	$\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$	ND	0.01
k (segmentation)	$\{2,3,4,5,6,7,8,9,10\}$	2	8
ρ	$[10^{-3}, 10^{3}]$	$1.18\cdot10^{-3}$	$1.11\cdot10^{1}$

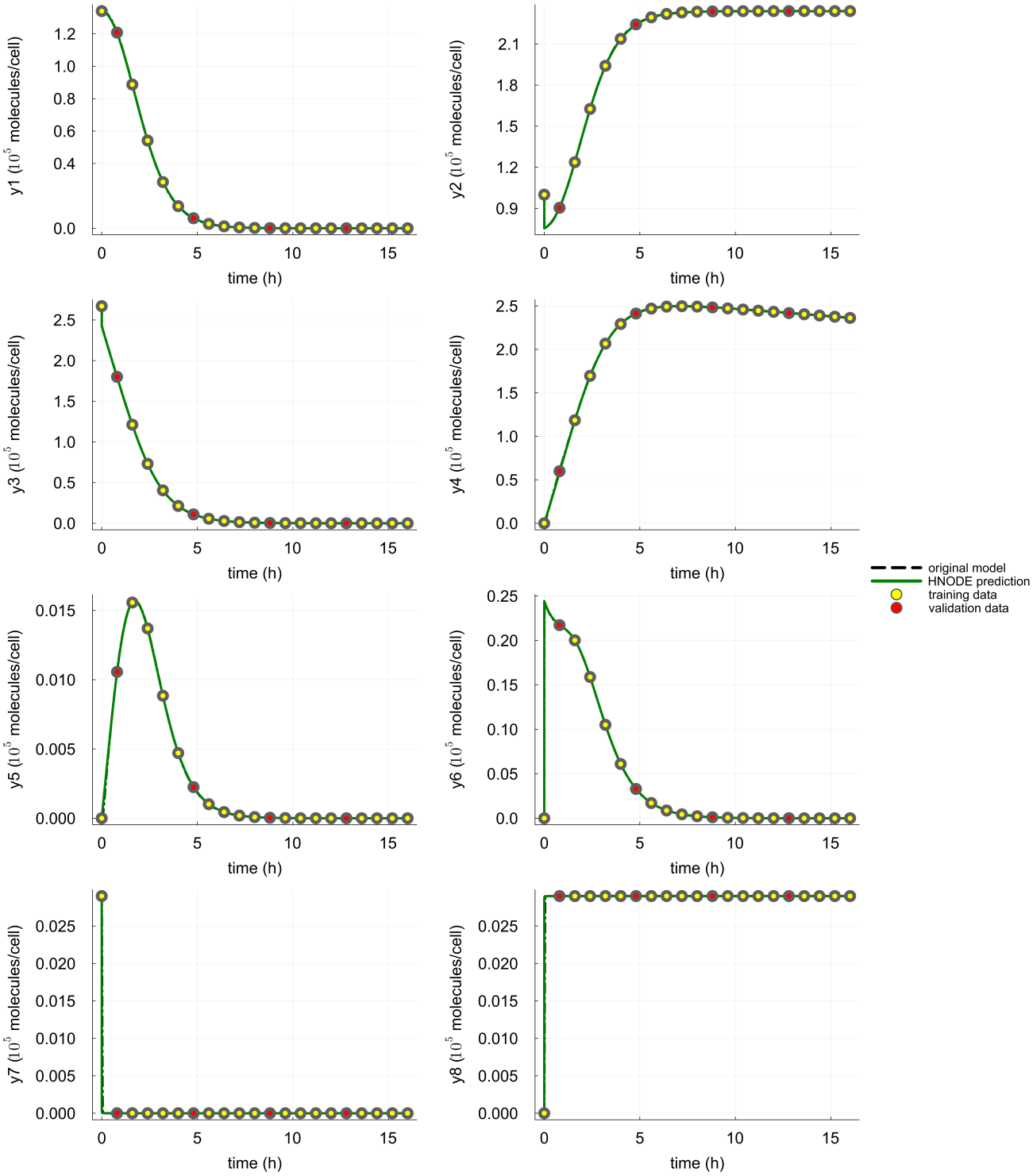
Supplementary Table 8. Cell apoptosis test case (first scenario): tuned hyperparameters of the HNODE model.

Supplementary Note 8: Cell apoptosis test case (first scenario): identifiability analysis

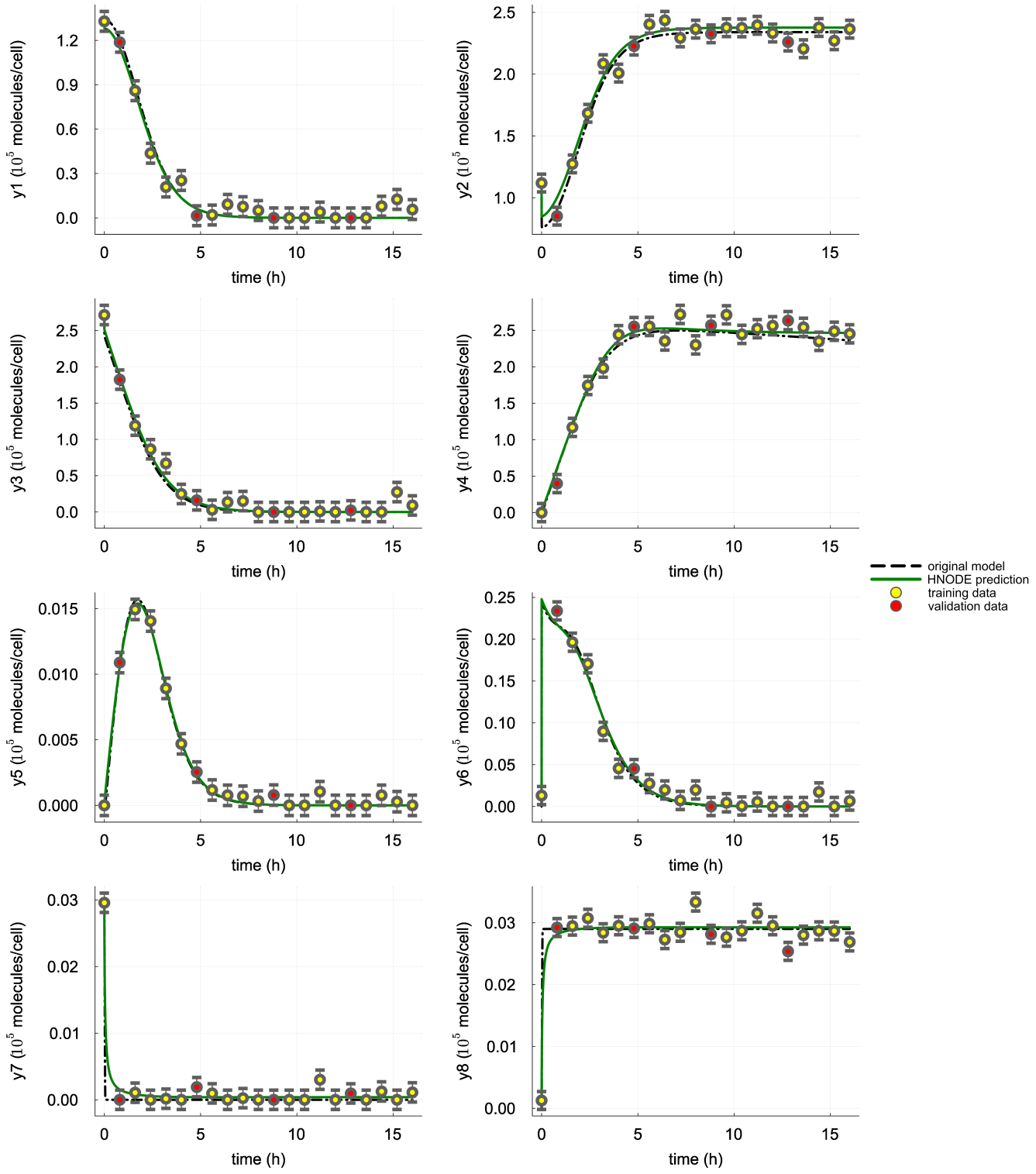


Supplementary Figure 2. Identifiability analysis of mechanistic parameters in the cell apoptosis test case, first scenario ($DS_{0.05}$). Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the model trained on $DS_{0.05}$. The total height of the bar corresponds to the squared norm of the projection, while the different components of the projection are depicted in different colors. The red line indicates the threshold to determine the identifiability of the parameter (0.05). Parameters with a squared norm of the projection exceeding this threshold are classified as locally non-identifiable.

Supplementary Note 9: Cell apoptosis test case (first scenario): predicted dynamics



Supplementary Figure 3. Dynamics predicted by the HNODE model in the cell apoptosis test case, first scenario ($DS_{0.00}$). The dynamics predicted by the HNODE model trained on $DS_{0.00}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.



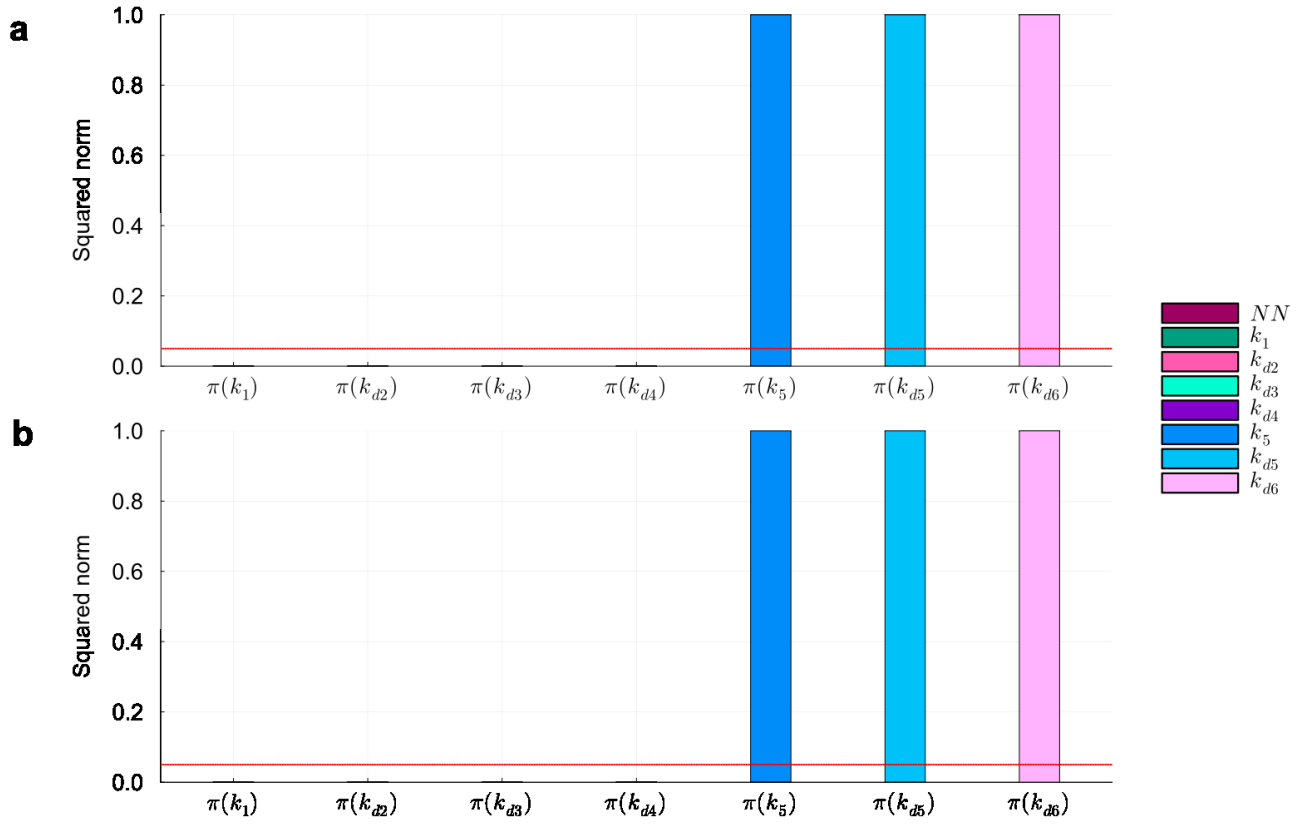
Supplementary Figure 4. Dynamics predicted by the HNODE model in the cell apoptosis test case, first scenario ($DS_{0.05}$). The dynamics predicted by the HNODE model trained on $DS_{0.05}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Supplementary Note 10: Cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed): tuned hyper-parameters

Name	Search Space	Tuned value $DS_{0.00}$	Tuned value $DS_{0.05}$
Neural network architecture			
hidden layer number	$\{2,3,4,5,6\}$	2	4
hidden layer width	$\{4,8,16,32\}$	8	8
Starting values for mechanistic parameters			
k_1	$[9.61\cdot10^{-3}, 9.61\cdot10^{1}]$	$1.53\cdot10^{1}$	8.58
k_{d2}	$[2.88\cdot10^{-1}, 2.88\cdot10^{3}]$	$3.43\cdot10^{1}$	$5.70\cdot10^{2}$
k_{d3}	$[1.80, 1.80\cdot10^{4}]$	$2.74\cdot10^{3}$	$1.28\cdot10^{4}$
k_{d4}	$[3.60\cdot10^{-2}, 3.60\cdot10^{2}]$	$9.04\cdot10^{1}$	$8.42\cdot10^{1}$
k_5	$[2.52\cdot10^{2}, 2.52\cdot10^{6}]$	$2.05\cdot10^{5}$	$1.79\cdot10^{6}$
k_{d5}	$[6.01\cdot10^{-4}, 6.01]$	$7.85\cdot10^{-1}$	2.59
k_{d6}	$[6.01\cdot10^{-3}, 6.01\cdot10^{1}]$	$2.37\cdot10^{1}$	$1.43\cdot10^{1}$
Training hyper-parameters			
learning rate	$[10^{-5}, 10^{-1}]$	$2.57\cdot10^{-2}$	$1.04\cdot10^{-2}$
λ	$\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$	ND	0.01
k (segmentation)	$\{2,3,4,5,6,7,8,9,10\}$	2	8
ρ	$[10^{-3}, 10^{3}]$	$5.24\cdot10^{-3}$	$9.88\cdot10^{2}$

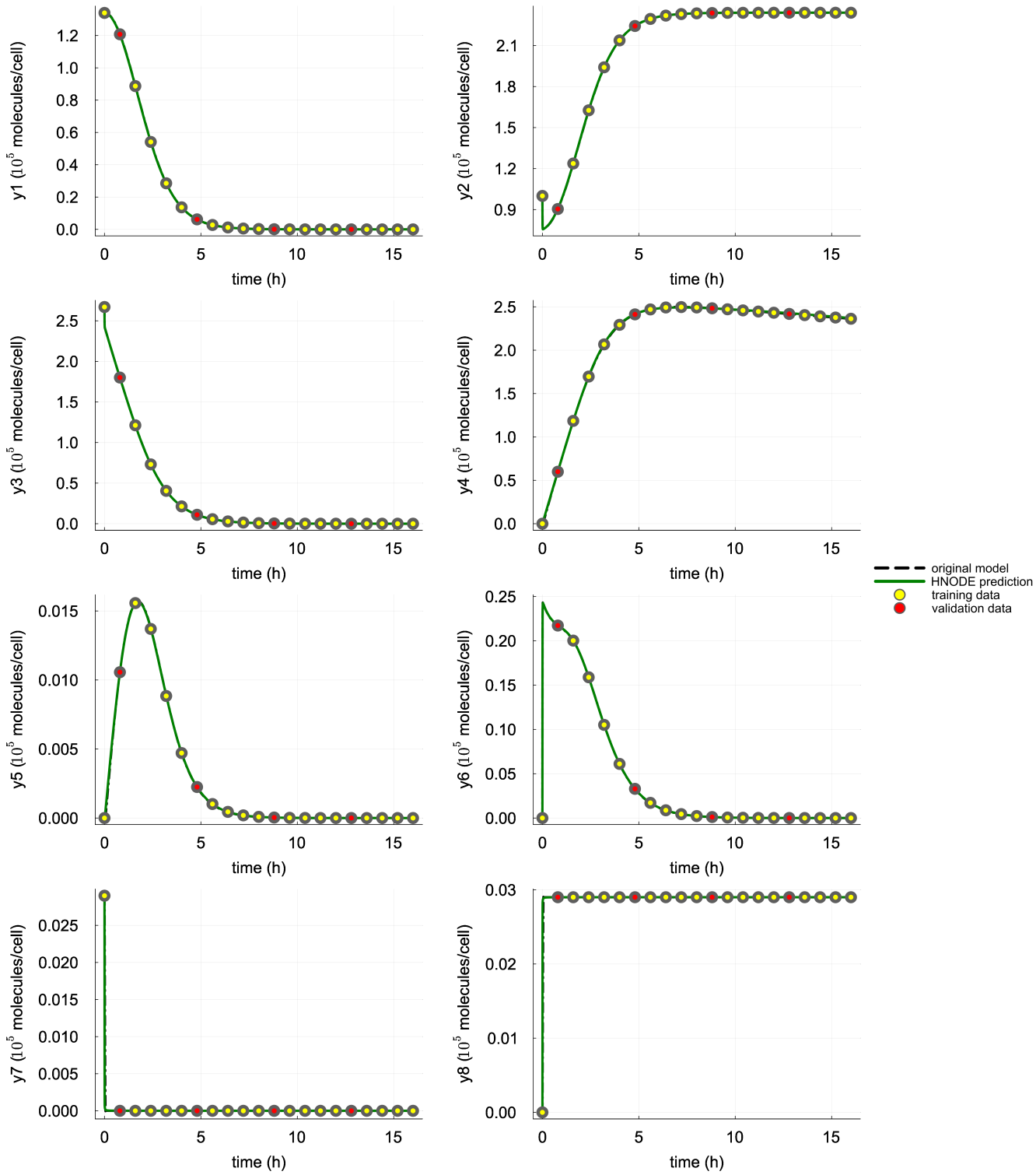
Supplementary Table 9. Cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed): tuned hyperparameters of the HNODE model.

Supplementary Note 11: Cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed): identifiability analysis

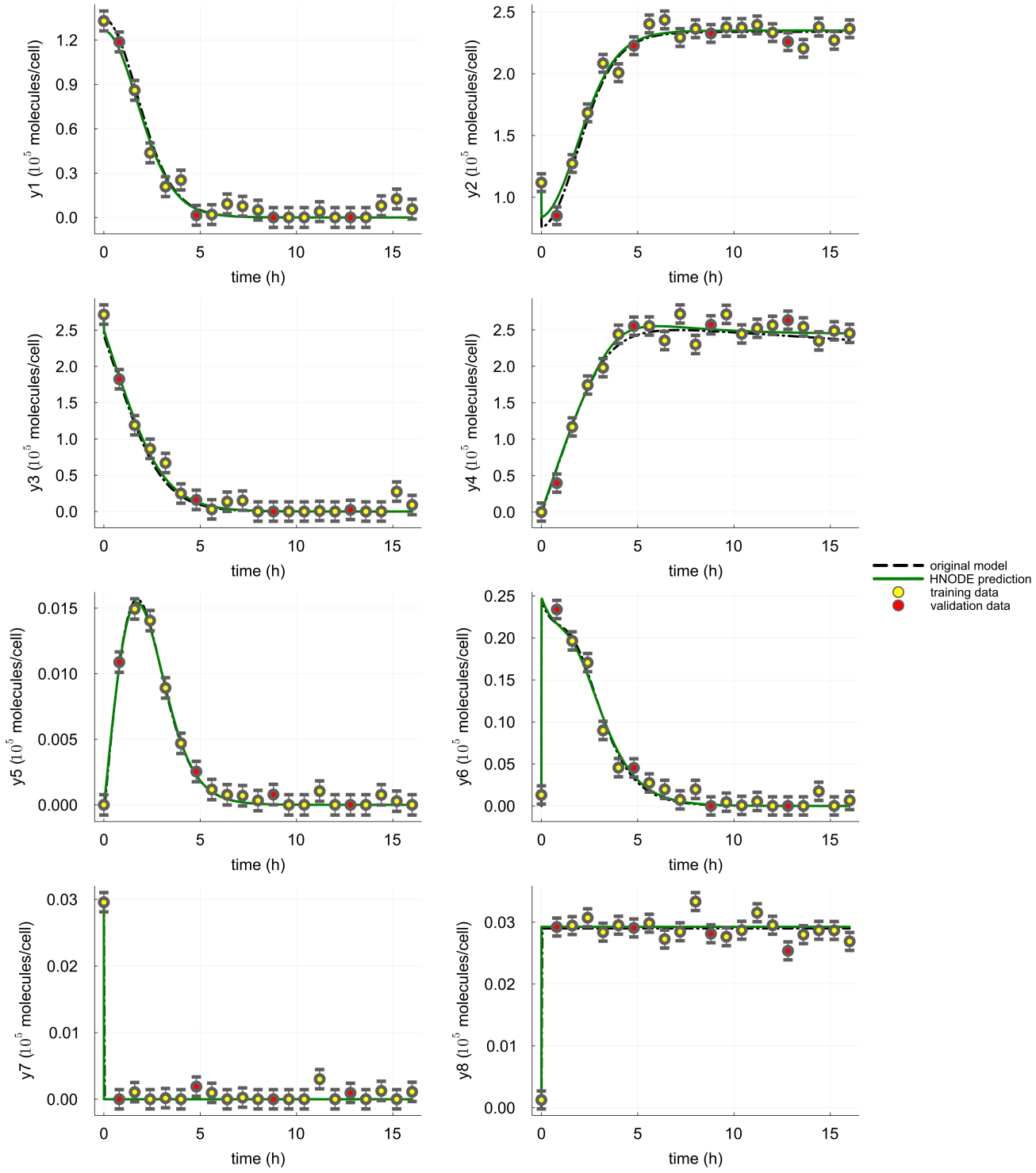


Supplementary Figure 5. Identifiability analysis of mechanistic parameters in the cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed). Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the models trained on $DS_{0.00}$ and $DS_{0.05}$ (panels **a** and **b** respectively). The total height of the bar corresponds to the squared norm of the projection, while the different components of the projection are depicted in different colors. The red line indicates the threshold to determine the identifiability of the parameter (0.05). Parameters with a squared norm of the projection exceeding this threshold are classified as locally non-identifiable.

Supplementary Note 12: Cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed): predicted dynamics



Supplementary Figure 6. Dynamics predicted by the HNODE model in the cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed). The dynamics predicted by the HNODE model trained on $DS_{0.00}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.



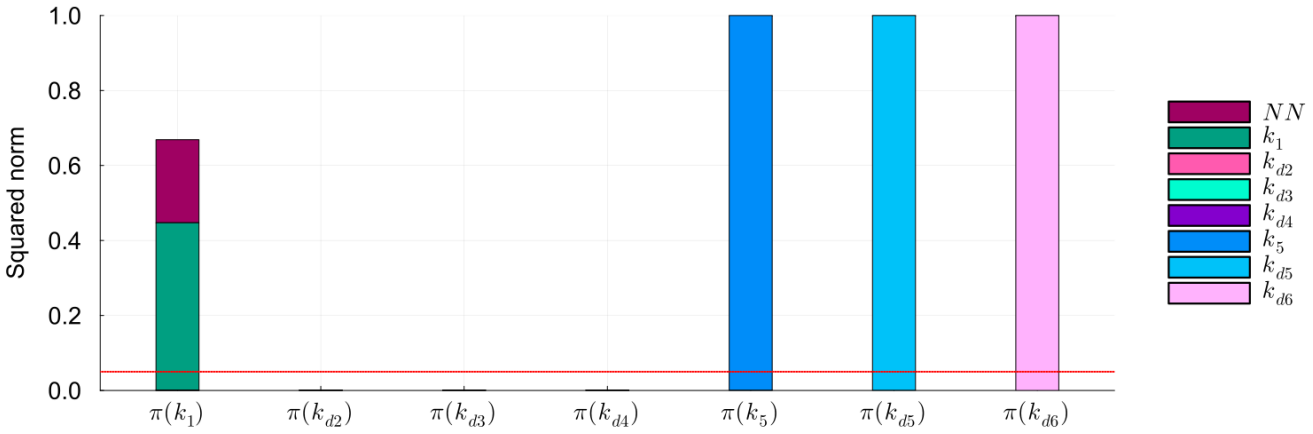
Supplementary Figure 7. Dynamics predicted by the HNODE model in the cell apoptosis test case (first scenario, with k_{d1} and k_3 fixed). The dynamics predicted by the HNODE model trained on $DS_{0.05}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Supplementary Note 13: Cell apoptosis test case (second scenario): tuned hyper-parameters

Name	Search Space	Tuned value $DS_{0.00}$	Tuned value $DS_{0.05}$
Neural network architecture			
hidden layer number	$\{2,3,4,5,6\}$	2	5
hidden layer width	$\{4,8,16,32\}$	8	4
Starting values for mechanistic parameters			
k_1	$[9.61\cdot10^{-3}, 9.61\cdot10^{1}]$	6.73	2.08
k_{d2}	$[2.88\cdot10^{-1}, 2.88\cdot10^{3}]$	$6.93\cdot10^{2}$	$1.24\cdot10^{3}$
k_{d3}	$[1.80, 1.80\cdot10^{4}]$	$8.82\cdot10^{2}$	$1.59\cdot10^{4}$
k_{d4}	$[3.60\cdot10^{-2}, 3.60\cdot10^{2}]$	$4.98\cdot10^{1}$	$5.01\cdot10^{1}$
k_5	$[2.52\cdot10^{2}, 2.52\cdot10^{6}]$	$2.39\cdot10^{6}$	$2.28\cdot10^{6}$
k_{d5}	$[6.01\cdot10^{-4}, 6.01]$	$8.98\cdot10^{-1}$	2.43
k_{d6}	$[6.01\cdot10^{-3}, 6.01\cdot10^{1}]$	$2.70\cdot10^{1}$	$3.14\cdot10^{1}$
Training hyper-parameters			
learning rate	$[10^{-5}, 10^{-1}]$	$4.14\cdot10^{-2}$	$4.12\cdot10^{-3}$
λ	$\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$	ND	0.001
k (segmentation)	$[2, 20]$	19	10
ρ	$[10^{-3}, 10^{3}]$	$4.04\cdot10^{-3}$	$4.97\cdot10^{-3}$

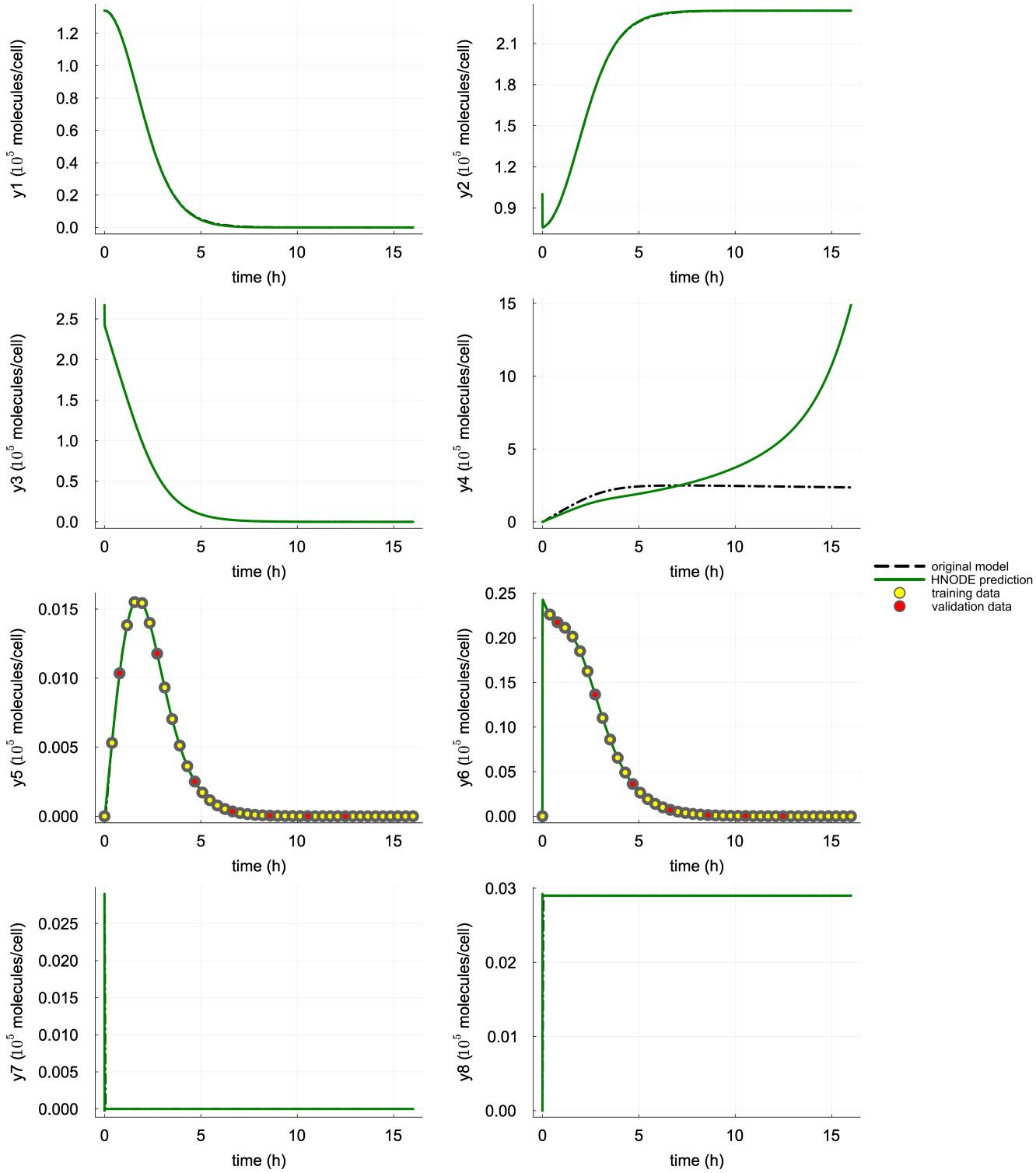
Supplementary Table 10. Cell apoptosis test case (second scenario): tuned hyperparameters of the HNODE model.

Supplementary Note 14: Cell apoptosis test case (second scenario): identifiability analysis

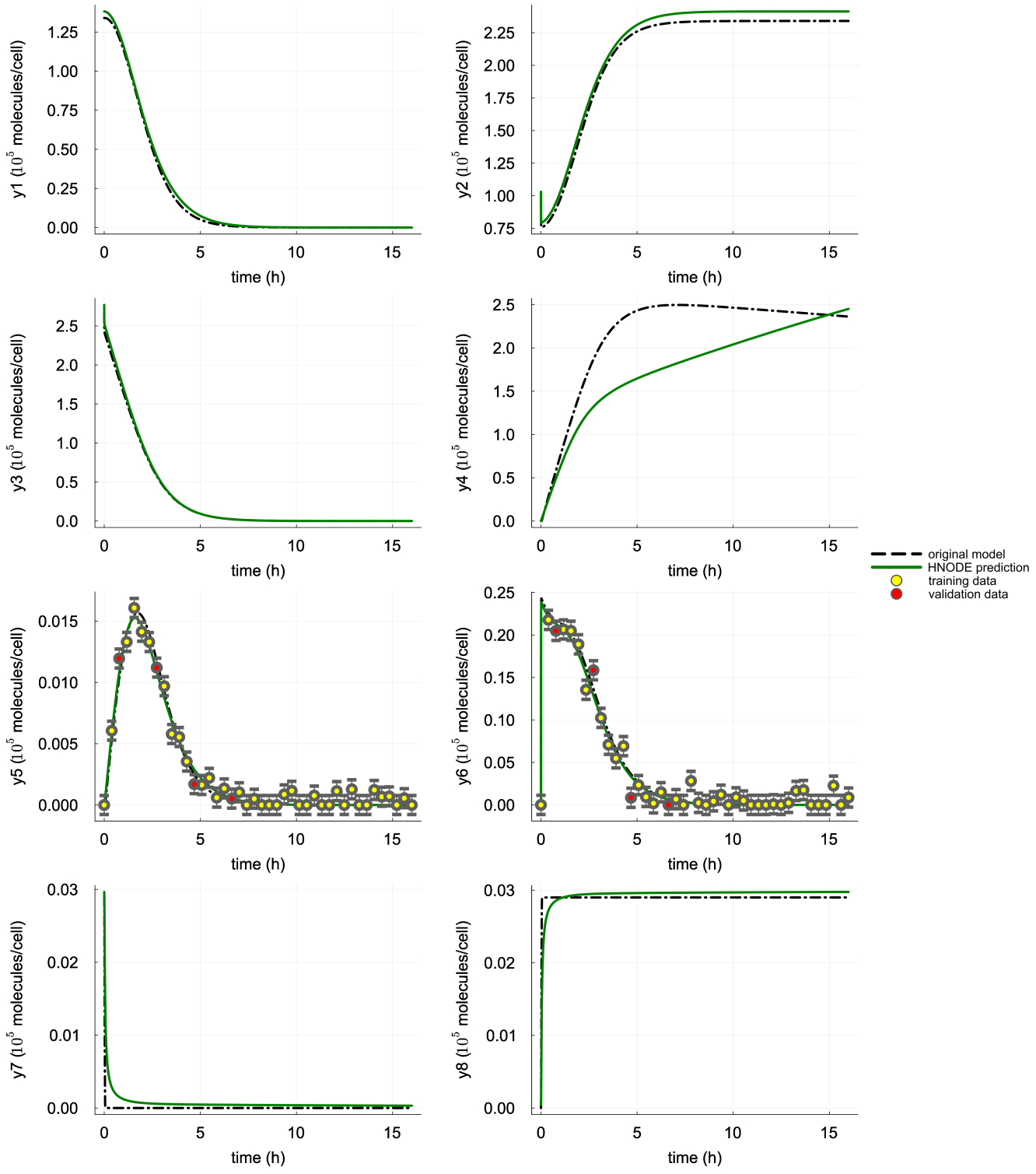


Supplementary Figure 8. Identifiability analysis of mechanistic parameters in the cell apoptosis test case, second scenario ($DS_{0.05}$). Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the model trained on $DS_{0.05}$. The total height of the bar corresponds to the squared norm of the projection, while the different components of the projection are depicted in different colors. The red line indicates the threshold to determine the identifiability of the parameter (0.05). Parameters with a squared norm of the projection exceeding this threshold are classified as locally non-identifiable.

Supplementary Note 15: Cell apoptosis test case (second scenario): predicted dynamics



Supplementary Figure 9. Dynamics predicted by the HNODE model in the cell apoptosis test case, second scenario ($DS_{0.00}$). The dynamics predicted by the HNODE model trained on $DS_{0.00}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.



Supplementary Figure 10. Dynamics predicted by the HNODE model in the cell apoptosis test case, second scenario ($DS_{0.05}$). The dynamics predicted by the HNODE model trained on $DS_{0.05}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Supplementary Note 16: Cell apoptosis test case: identifiability analysis of the original mechanistic model with different combinations of observable variables

The local identifiability of parameters within the original mechanistic model for cell apoptosis was evaluated using the methodology proposed by Quaiser et al.[8]. This evaluation assumed the same observation time-points as the first scenario and explored various combinations of observable variables. Supplementary Table 11 presents the identifiable parameters under different combinations of observable variables. Within the original model, y_5 and y_6 form a minimal subset of observable variables that maximizes the number of identifiable parameters.

Observable variables	k_1	k_{d1}	k_{d2}	k_3	k_{d3}	k_{d4}	k_5	k_{d5}	k_{d6}
y1, y2, y3, y4, y5, y6, y7, y8	✓		✓		✓	✓			✓
y5, y6	✓		✓		✓	✓			✓
y5	✓		✓			✓			
y6	✓				✓				

Supplementary Table 11. Cell apoptosis: identifiability analysis of the original mechanistic model with various combinations of observable variables. The symbol ✓ denotes that the corresponding parameter is classified as identifiable.

Supplementary Note 17: Cell apoptosis test case: impact of ε and δ on the identifiability analysis results

To assess the impact of varying hyperparameters ε and δ on the identifiability analysis, we conducted the analysis using different values for these parameters (Supplementary Tables 12, 13, and 14). Our findings indicate that the results exhibit robustness against changes in ε and δ.

		$\varepsilon = 10^{-6}$			$\varepsilon = 10^{-5}$			$\varepsilon = 10^{-4}$		
Par.	**Dataset**	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.05$	$\delta = 0.07$	
k_1	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d1}	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d2}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_3	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d3}	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d4}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_5	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d5}	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d6}	$DS_{0.00}$									
	$DS_{0.05}$									

Supplementary Table 12. Cell apoptosis test case (first scenario): parameter identifiability with different choices of ε and δ. The symbol ✓ denotes the parameters classified as identifiable with the ε and δ threshold specified in the header.

Par.	Dataset	$\varepsilon = 10^{-6}$			$\varepsilon = 10^{-5}$			$\varepsilon = 10^{-4}$		
		$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.05$	$\delta = 0.07$	
k_1	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_{d2}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_{d3}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_{d4}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_5	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d5}	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d6}	$DS_{0.00}$									
	$DS_{0.05}$									

Supplementary Table 13. Cell apoptosis (first scenario, with k_{d1} and k_3 fixed): parameter identifiability with different choices of ε and δ. The symbol ✓ denotes the parameters classified as identifiable with the ε and δ threshold specified in the header.

Par.	Dataset	$\varepsilon = 10^{-6}$			$\varepsilon = 10^{-5}$			$\varepsilon = 10^{-4}$		
		$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.05$	$\delta = 0.07$	
k_1	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d2}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_{d3}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_{d4}	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_5	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d5}	$DS_{0.00}$									
	$DS_{0.05}$									
k_{d6}	$DS_{0.00}$									
	$DS_{0.05}$									

Supplementary Table 14. Cell apoptosis (second scenario): parameter identifiability with different choices of ε and δ. The symbol ✓ denotes the parameters classified as identifiable with the ε and δ threshold specified in the header.

Supplementary Note 18: Yeast glycolysis test case: experimental setting

To simulate in silico observations, the Yeast glycolysis ODE system is solved numerically employing the *TRBDF2* solver from the *DifferentialEquations.jl* library. The solver settings are configured with an absolute tolerance of 10^{-7} and a relative tolerance of 10^{-6}. The parameters and initial states, derived from the values used by Ruoff et al.[9], are detailed in Supplementary Table 15 and 16.

Parameter	Value	Unit
J_0	2.5	$\mathrm{mM}\cdot\mathrm{min}^{-1}$
k_1	100.0	$\mathrm{mM}^{-1}\cdot\mathrm{min}^{-1}$
k_2	6.0	$\mathrm{mM}^{-1}\cdot\mathrm{min}^{-1}$
k_3	16.0	$\mathrm{mM}^{-1}\cdot\mathrm{min}^{-1}$
k_4	100.0	$\mathrm{mM}^{-1}\cdot\mathrm{min}^{-1}$
k_5	1.28	min^{-1}
k_6	12.0	$\mathrm{mM}^{-1}\cdot\mathrm{min}^{-1}$
k	1.8	min^{-1}
κ	13.0	min^{-1}
q	4.0	
K_1	0.52	mM
ψ	0.1	
N	1.0	mM
A	4.0	mM

Supplementary Table 15. Yeast glycolysis test case: parameters employed for generating the *in silico* dataset

Variable	Value	Unit
y_1	1.6	mM
y_2	2.16	mM
y_3	0.2	mM
y_4	0.35	mM
y_5	0.3	mM
y_6	2.67	mM
y_7	0.1	mM

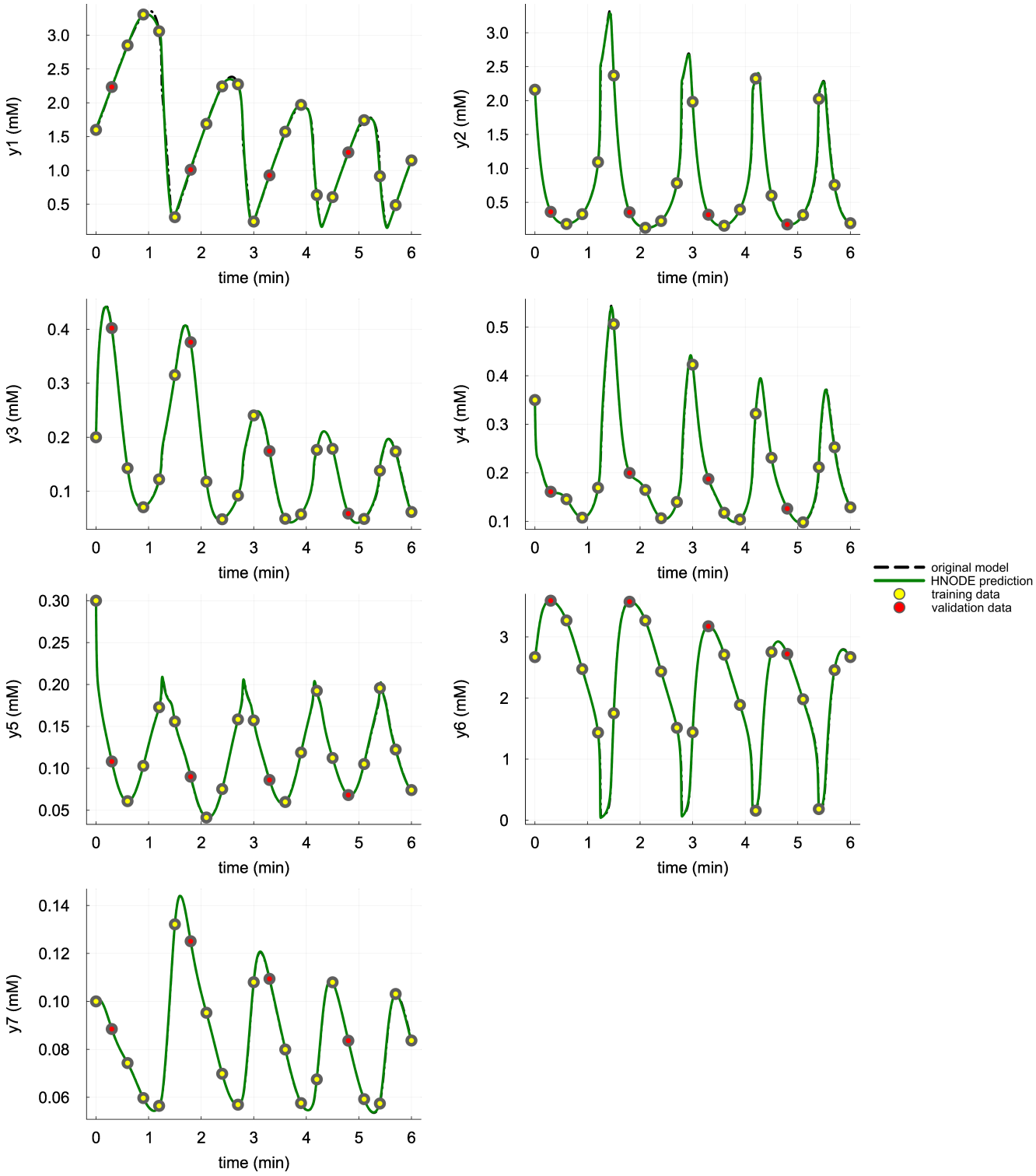
Supplementary Table 16. Yeast glycolysis test case: initial states employed for generating the *in silico* dataset

Supplementary Note 19: Yeast glycolysis test case (first scenario): tuned hyper-parameters

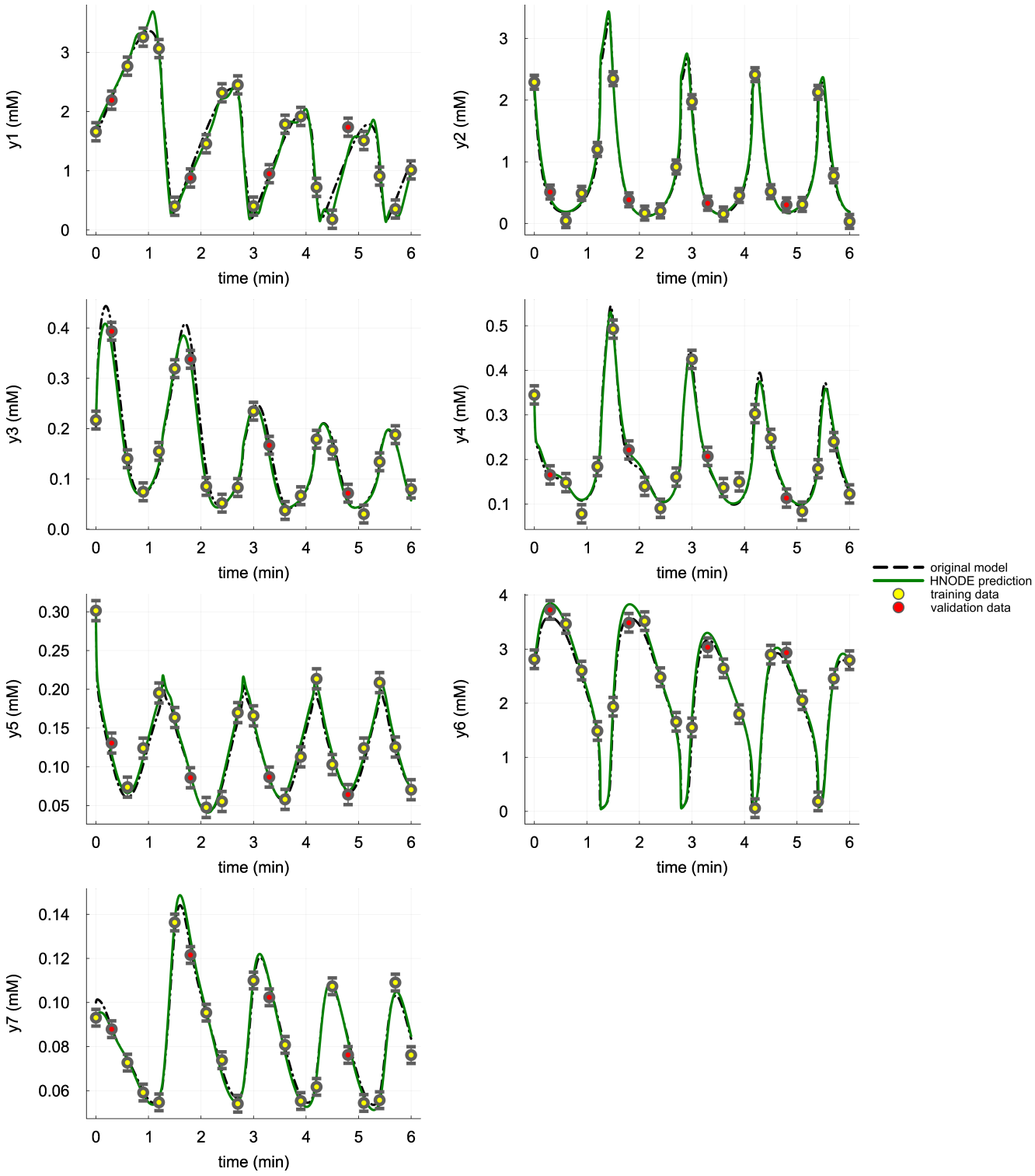
Name	Search Space	Tuned value $DS_{0.00}$	Tuned value $DS_{0.05}$
Neural network architecture			
hidden layer number	$\{2,3,4,5,6\}$	2	5
hidden layer width	$\{4,8,16,32\}$	16	16
Starting values for mechanistic parameters			
k_1	$[1.00\cdot10^{1}, 1.00\cdot10^{3}]$	$5.70\cdot10^{1}$	$1.15\cdot10^{2}$
K_1	$[5.20\cdot10^{-2}, 5.20]$	$1.10\cdot10^{-1}$	$6.39\cdot10^{-2}$
q	$[4.00\cdot10^{-1}, 4.00\cdot10^{1}]$	6.52	5.18
k_2	$[6.00\cdot10^{-1}, 6.00\cdot10^{1}]$	$3.24\cdot10^{1}$	4.98
N	$[1.00\cdot10^{-1}, 1.00\cdot10^{1}]$	2.71	$4.53\cdot10^{-1}$
k_6	$[1.20, 1.20\cdot10^{2}]$	$2.68\cdot10^{1}$	$1.71\cdot10^{1}$
k_3	$[1.60, 1.60\cdot10^{2}]$	$1.21\cdot10^{2}$	$4.92\cdot10^{1}$
A	$[4.00\cdot10^{-1}, 4.00\cdot10^{1}]$	$2.13\cdot10^{1}$	$1.82\cdot10^{1}$
k_4	$[1.00\cdot10^{1}, 1.00\cdot10^{3}]$	$1.35\cdot10^{2}$	$5.30\cdot10^{2}$
κ	$[1.30, 1.30\cdot10^{2}]$	$3.76\cdot10^{1}$	$2.02\cdot10^{1}$
k_5	$[1.28\cdot10^{-1}, 1.28\cdot10^{1}]$	3.81	7.05
ψ	$[1.00\cdot10^{-2}, 1.00]$	$6.24\cdot10^{-1}$	$1.76\cdot10^{-1}$
k	$[1.80\cdot10^{-1}, 1.80\cdot10^{1}]$	2.45	5.93
Training hyper-parameters			
learning rate	$[10^{-5}, 10^{-1}]$	$4.74\cdot10^{-2}$	$1.07\cdot10^{-2}$
λ	$\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$	ND	0.01
k (segmentation)	$\{2,3,4,5,6,7,8,9,10\}$	10	5
ρ	$[10^{-3}, 10^{3}]$	$2.03\cdot10^{-2}$	$5.43\cdot10^{-1}$

Supplementary Table 17. Yeast glycolysis test case (first scenario): tuned hyperparameters of the HNODE model.

Supplementary Note 20: Yeast glycolysis test case (first scenario): predicted dynamics

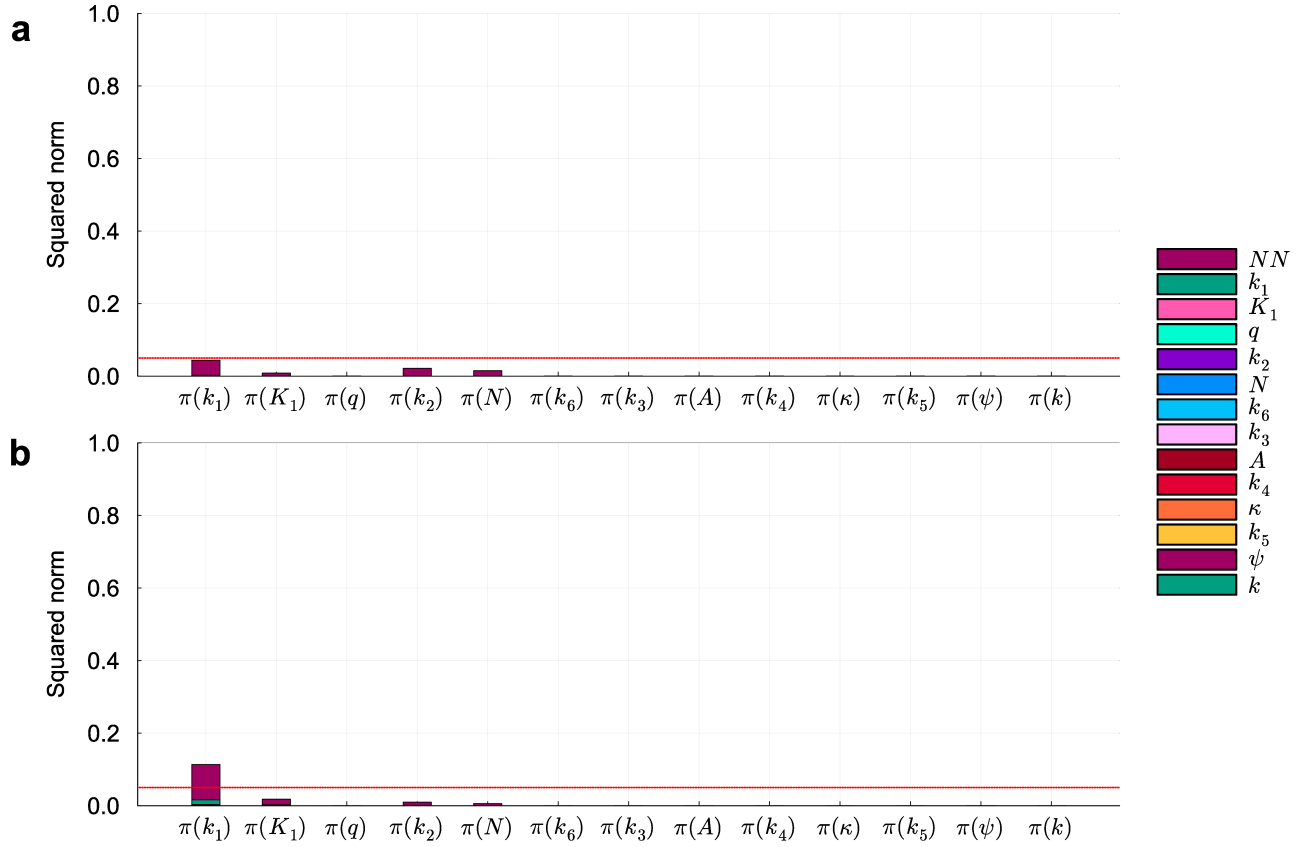


Supplementary Figure 11. Dynamics predicted by the HNODE model in the yeast glycolysis test case, first scenario ($DS_{0.00}$). The dynamics predicted by the HNODE model trained on $DS_{0.00}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.



Supplementary Figure 12. Dynamics predicted by the HNODE model in the yeast glycolysis test case, first scenario ($DS_{0.05}$). The dynamics predicted by the HNODE model trained on $DS_{0.05}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Supplementary Note 21: Yeast glycolysis test case (first scenario): identifiability analysis



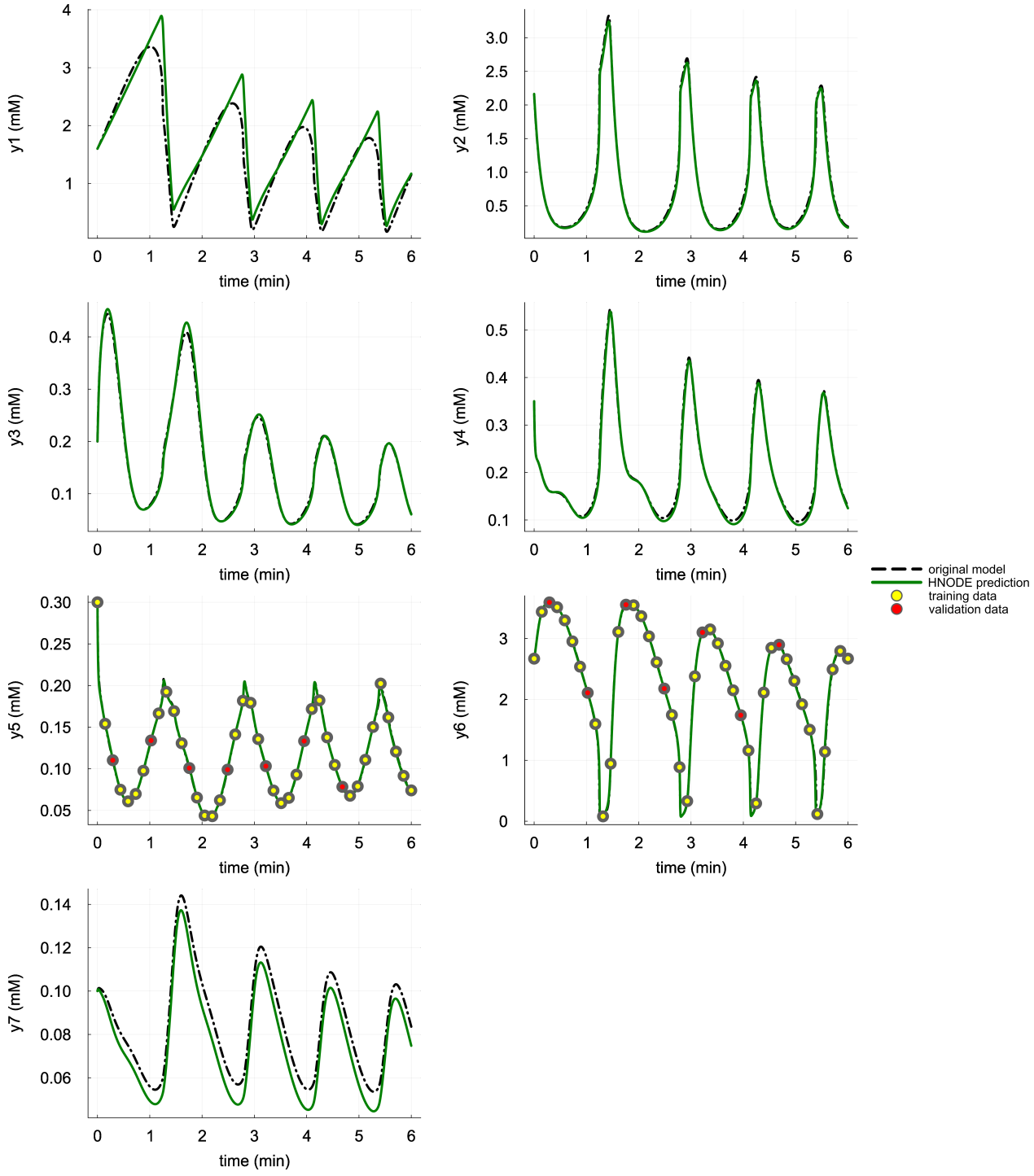
Supplementary Figure 13. Identifiability analysis of mechanistic parameters in the yeast glycolysis test case, first scenario. Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the models trained on $DS_{0.00}$ and $DS_{0.05}$ (panels **a** and **b** respectively). The total height of the bar corresponds to the squared norm of the projection, while the different components of the projection are depicted in different colors. The red line indicates the threshold to determine the identifiability of the parameter (0.05). Parameters with a squared norm of the projection exceeding this threshold are classified as locally non-identifiable.

Supplementary Note 22: Yeast glycolysis test case (second scenario): tuned hyper-parameters

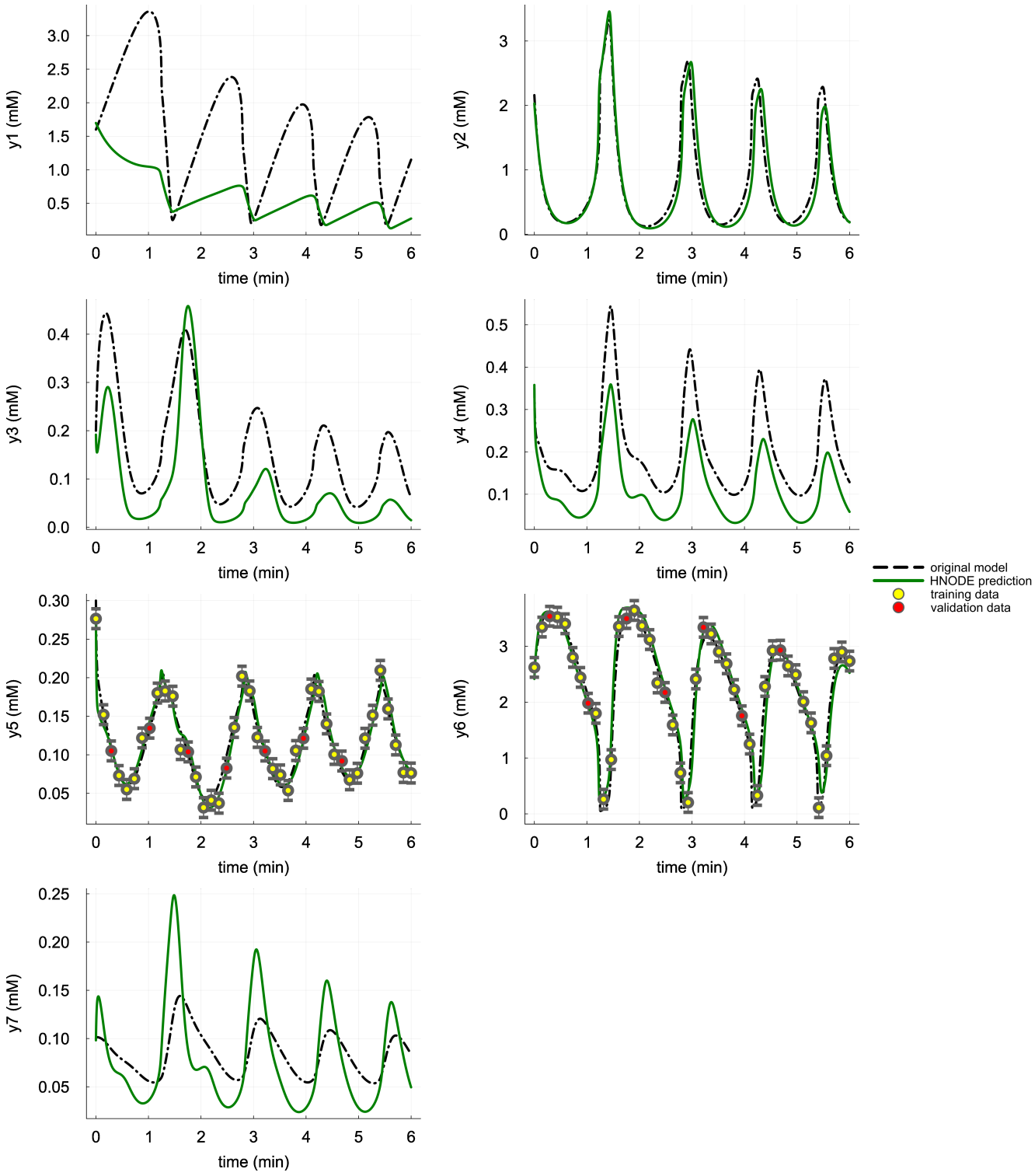
Name	Search Space	Tuned value	
		$DS_{0.00}$	$DS_{0.05}$
Neural network architecture			
hidden layer number	$\{2,3,4,5,6\}$	3	2
hidden layer width	$\{4,8,16,32\}$	16	16
Starting values for mechanistic parameters			
k_1	$[1.00\cdot10^{1}, 1.00\cdot10^{3}]$	$8.96\cdot10^{1}$	$6.50\cdot10^{1}$
K_1	$[5.20\cdot10^{-2}, 5.20]$	$9.09\cdot10^{-2}$	$2.35\cdot10^{-1}$
q	$[4.00\cdot10^{-1}, 4.00\cdot10^{1}]$	5.74	3.34
k_2	$[6.00\cdot10^{-1}, 6.00\cdot10^{1}]$	3.10	1.83
N	$[1.00\cdot10^{-1}, 1.00\cdot10^{1}]$	3.31	1.87
k_6	$[1.20, 1.20\cdot10^{2}]$	1.39	$5.53\cdot10^{1}$
k_3	$[1.60, 1.60\cdot10^{2}]$	4.54	$2.82\cdot10^{1}$
A	$[4.00\cdot10^{-1}, 4.00\cdot10^{1}]$	$1.94\cdot10^{1}$	$1.64\cdot10^{1}$
k_4	$[1.00\cdot10^{1}, 1.00\cdot10^{3}]$	$8.66\cdot10^{2}$	$5.05\cdot10^{2}$
κ	$[1.30, 1.30\cdot10^{2}]$	$7.02\cdot10^{1}$	$1.10\cdot10^{1}$
k_5	$[1.28\cdot10^{-1}, 1.28\cdot10^{1}]$	2.09	1.40
ψ	$[1.00\cdot10^{-2}, 1.00]$	$7.97\cdot10^{-1}$	$2.09\cdot10^{-1}$
k	$[1.80\cdot10^{-1}, 1.80\cdot10^{1}]$	$1.74\cdot10^{1}$	8.71
Training hyper-parameters			
learning rate	$[10^{-5}, 10^{-1}]$	$2.10\cdot10^{-2}$	$7.90\cdot10^{-2}$
λ	$\{10^{-3}, 10^{-2}, 10^{-1}, 1\}$	ND	0.1
k (segmentation)	$[2, 20]$	19	9
ρ	$[10^{-3}, 10^{3}]$	$1.94\cdot10^{-3}$	$7.05\cdot10^{-1}$

Supplementary Table 18. Yeast glycolysis test case (second scenario): tuned hyperparameters of the HNODE model.

Supplementary Note 23: Yeast glycolysis test case (second scenario): predicted dynamics



Supplementary Figure 14. Dynamics predicted by the HNODE model in the yeast glycolysis test case, second scenario ($DS_{0.00}$). The dynamics predicted by the HNODE model trained on $DS_{0.00}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.



Supplementary Figure 15. Dynamics predicted by the HNODE model in the yeast glycolysis test case, second scenario ($DS_{0.05}$). The dynamics predicted by the HNODE model trained on $DS_{0.05}$ are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Supplementary Note 24: Yeast glycolysis test case: parameter estimation with the original mechanistic model on the dataset $DS_{0.05}$

To distinguish between the effects of the lack of mechanistic knowledge and the noise in the datasets on parameter estimation, we conducted parameter estimation of the mechanistic model using the dataset $DS_{0.05}$. The estimation was carried out by minimizing the quadratic cost, weighted with uncertainty, assuming 21 observation time-points. Optimization was performed using the L-BFGS algorithm until convergence, with parameters initialized to their literature values. The results are presented in Supplementary Table 19.

Par.	Target	Estimate	Relative Error
J_0	2.50	2.50	0.14%
k_1	$1.00 \cdot 10^2$	$1.29 \cdot 10^2$	29.23%
K_1	$5.20 \cdot 10^{-1}$	$5.01 \cdot 10^{-1}$	3.65%
q	4.00	4.03	0.82%
k_2	6.00	1.82	69.69%
N	1.00	2.88	188.19%
k_6	$1.20 \cdot 10^1$	$1.11 \cdot 10^1$	7.76%
k_3	$1.60 \cdot 10^1$	$1.53 \cdot 10^1$	4.36%
A	4.00	4.11	2.75%
k_4	$1.00 \cdot 10^2$	$9.34 \cdot 10^1$	6.59%
κ	$1.30 \cdot 10^1$	$1.31 \cdot 10^1$	0.88%
k_5	1.28	1.25	2.69%
ψ	$1.00 \cdot 10^{-1}$	$1.03 \cdot 10^{-1}$	2.93%
k	1.80	1.90	5.44%
		Mean	23.22%
		Max	188.19%

Supplementary Table 19. Yeast glycolysis test case: parameters estimated on the dataset $DS_{0.05}$ using the original model.

Supplementary Note 25: Yeast glycolysis test case: impact of ε and δ on the identifiability analysis results

To assess the impact of varying hyperparameters ε and δ on the identifiability analysis, we conducted the analysis using different values for these parameters. In the first scenario (Supplementary Table 20), the results demonstrate overall robustness, except for parameters k_1, K_1, k_2, and N, which appear to be influenced by compensatory mechanisms, particularly noticeable when considering $\varepsilon = 10^{-4}$. In the second scenario (Supplementary Table 21), there are differences between the HNODE model trained on $DS_{0.00}$ and $DS_{0.05}$. However, overall, only the results related to parameters q, A, k_2, and k_5 appear to maintain robustness against variations in ε. It is important to note that even in the original model, there exist sloppy directions where the model behavior changes only slightly compared to other directions. Choosing a higher value for ε could potentially include these sloppy, albeit not null-change, directions into the null subspace of H_χ, resulting in an overly conservative assessment of parameter identifiability.

		$\varepsilon = 10^{-6}$			$\varepsilon = 10^{-5}$			$\varepsilon = 10^{-4}$		
Par.	**Dataset**	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.05$	$\delta = 0.07$	
k_1	$DS_{0.00}$	✓	✓	✓		✓	✓			
	$DS_{0.05}$	✓	✓	✓						
K_1	$DS_{0.00}$	✓	✓	✓	✓	✓	✓			
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓			
q	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_2	$DS_{0.00}$	✓	✓	✓	✓	✓	✓			
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓			
N	$DS_{0.00}$	✓	✓	✓	✓	✓	✓			
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓			
k_6	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_3	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
A	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_4	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
κ	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_5	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
ψ	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓		✓	✓
k	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓

Supplementary Table 20. Yeast glycolysis, first scenario - parameter identifiability with different choices of ε and δ. The symbol ✓ denotes the parameters classified as identifiable with the ε and δ threshold specified in the header.

Par.	Dataset	$\varepsilon = 10^{-6}$			$\varepsilon = 10^{-5}$			$\varepsilon = 10^{-4}$		
		$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.03$	$\delta = 0.05$	$\delta = 0.07$	$\delta = 0.05$	$\delta = 0.07$	
k_1	$DS_{0.00}$	✓	✓	✓						
	$DS_{0.05}$	✓	✓	✓						
K_1	$DS_{0.00}$	✓	✓	✓	✓	✓	✓			
	$DS_{0.05}$	✓	✓	✓			✓			
q	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓		✓	✓
k_2	$DS_{0.00}$									
	$DS_{0.05}$									
N	$DS_{0.00}$		✓	✓						
	$DS_{0.05}$									
k_6	$DS_{0.00}$	✓	✓	✓	✓	✓	✓			
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓			
k_3	$DS_{0.00}$	✓	✓	✓		✓	✓			
	$DS_{0.05}$	✓	✓	✓						
A	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
k_4	$DS_{0.00}$		✓	✓			✓			
	$DS_{0.05}$	✓	✓	✓						
κ	$DS_{0.00}$									
	$DS_{0.05}$	✓	✓	✓						
k_5	$DS_{0.00}$	✓	✓	✓	✓	✓	✓	✓	✓	✓
	$DS_{0.05}$	✓	✓	✓	✓	✓	✓		✓	✓
ψ	$DS_{0.00}$									
	$DS_{0.05}$	✓	✓	✓						
k	$DS_{0.00}$									
	$DS_{0.05}$	✓	✓	✓						

Supplementary Table 21. Yeast glycolysis, second scenario - parameter identifiability with different choices of ε and δ. The symbol ✓ denotes the parameters classified as identifiable with the ε and δ threshold specified in the header.

Supplementary Note 26: Analysis of computational time required to run the pipeline

Although performance optimization was not the primary focus of our work, this section provides an analysis of the computational time required to run the pipeline, along with a discussion of future directions to reduce its computational complexity. We report the computational times for executing the pipeline on the test cases described in the main text, with a breakdown for each step of the process (see Supplementary Table 22 for a recap description of the pipeline steps).

Step 1	Splitting of the dataset into training and validation subsets
Step 2a - first stage	Exploration of hyperparameters and mechanistic parameter space
Step 2a - second stage	Adjustment of the L2 regularization factor (applied only to noisy datasets)
Step 2b	Training of the model
Step 3	Conducting identifiability analysis
Step 4	Estimation of confidence intervals

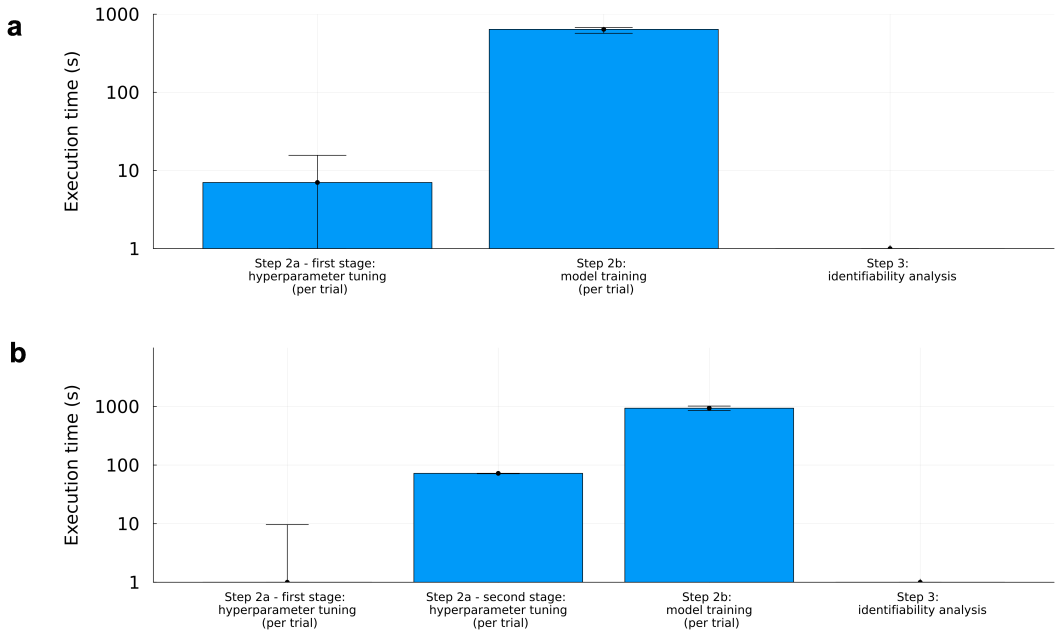
Supplementary Table 22. Description of the pipeline steps.

The results for HNODE models trained on datasets with and without noise are presented separately, as the second stage of hyperparameter tuning (tuning the L_2 regularization factor) is performed only for the noisy training datasets. Additionally, *Step 1* is not considered in this analysis, as it is not computationally significant. Supplementary Figs. 16, 17, and 19 illustrate the results for the Lotka-Volterra, cell apoptosis, and yeast glycolysis test cases, respectively. The computational times were measured under the conditions outlined in the *Implementation* subsection of the *Methods* section in the main text. While identifiability analysis (*Step 3*) and confidence interval computations (*Step 4*) are executed once per test case, hyperparameter tuning (*Step 2a*, first and second stage) and model training (*Step 2b*) involve multiple trials, and the time per trial is reported.

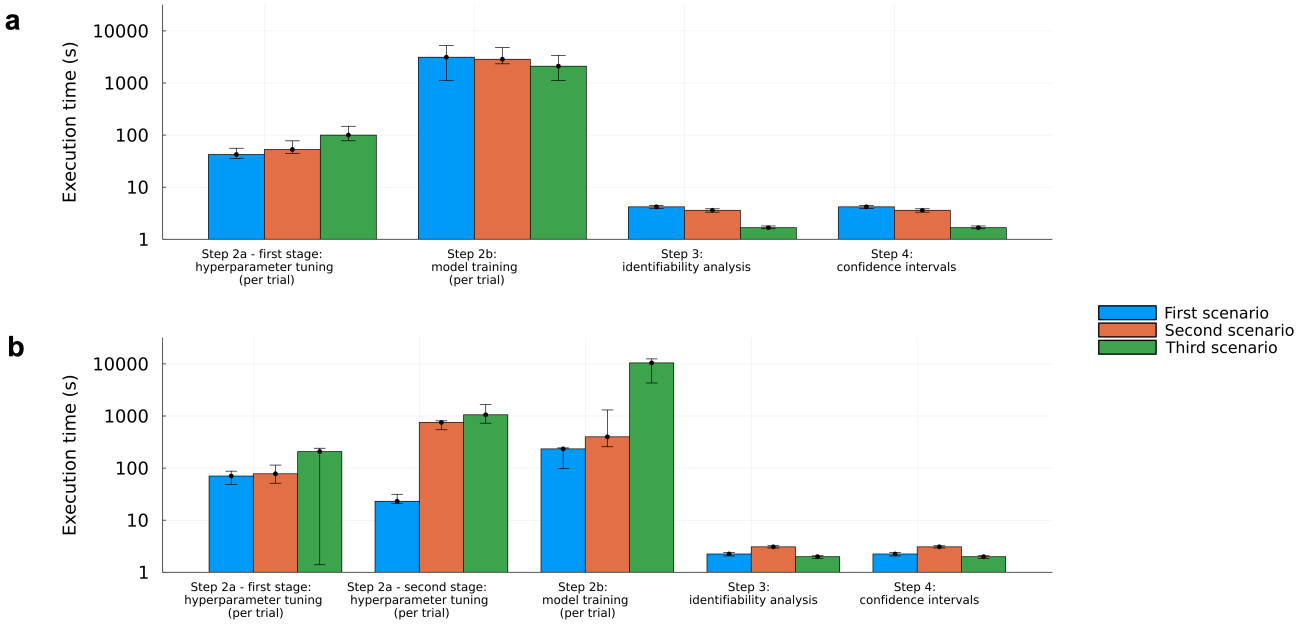
The times required to run the steps of the pipeline in the Lotka-Volterra test case are significantly lower than the times required in the other two cases. This can be explained (1) by the lower dimensionality of the Lotka-Volterra model, and (2) by the stiff configurations used to tune the pipeline in the cell apoptosis and yeast glycolysis test cases, which involve the use of a stiff numerical integrator and *ad hoc* adjoint sensitivity algorithms.

Within each test case, the time required to execute *Step 3* and *Step 4* is always negligible compared to the time needed for the other steps of the pipeline. This aligns with our expectations, as the identifiability analysis and confidence interval computations primarily require only one computation of the Jacobian of the HNODE system at the measured time points with respect to the model parameters.

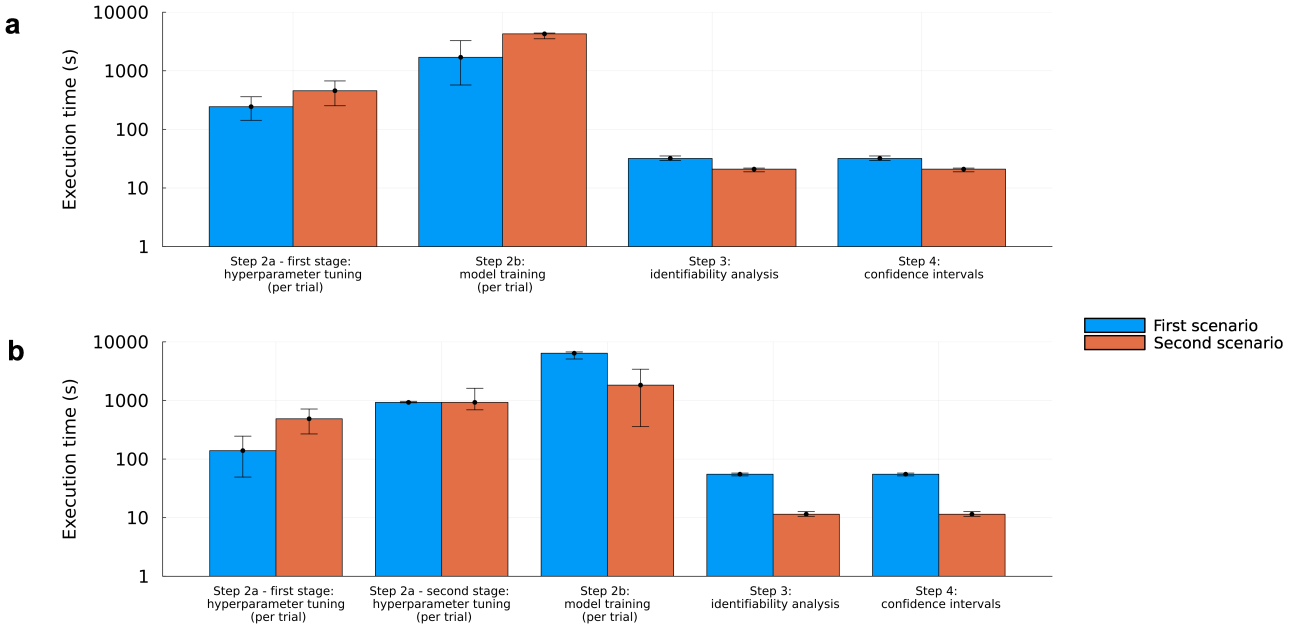
The times required to run a trial in *Step 2a - first stage*, *Step 2a - second stage*, and *Step 2b* reflect the number of epochs of the Adam optimizer used in each step: 5000, 2000, and 10000, respectively (plus up to 5000 epochs of the LBFGS optimizer on $DS_{0.00}$). Even though the time per single trial required for *Step 2a - first stage* is less than that of *Step 2a - second stage* and *Step 2b*, it is important to note that the first stage involves a significantly higher number of trials (500 vs. 5 and 10, respectively), and the trials in the Tree Structured Parzen Estimator (TPE) cannot be parallelized, whereas the trials in the other two steps can be executed in parallel. Therefore, the *Step 2a - first stage* could be the most computationally expensive part of the pipeline.



Supplementary Figure 16. Times required to run the pipeline in the Lotka-Volterra test case, considering the datasets $DS_{0.00}$ (panel **a**) and $DS_{0.05}$ (panel **b**). For *Step 2a - first stage*, *Step 2a - second stage*, and *Step 2b*, the times represent the median duration for a single trial, with error bars corresponding to the first and third quartiles. For *Step 3*, the uncertainty in execution time was determined by performing 100 runs of the step, with the median and interquartile range reported in the plot. In this test case, *Step 4* was not performed due to the non-identifiability of α.



Supplementary Figure 17. Times required to run the pipeline in the cell apoptosis test case, considering the datasets $DS_{0.00}$ (panel **a**) and $DS_{0.05}$ (panel **b**). For *Step 2a - first stage*, *Step 2a - second stage*, and *Step 2b*, the times represent the median duration for a single trial, with error bars indicating the first and third quartiles. For *Step 3* and *Step 4*, the uncertainty in execution time was assessed by performing 100 runs of each step, with the median and interquartile range shown in the plot. The various scenarios reported are described in the main text.

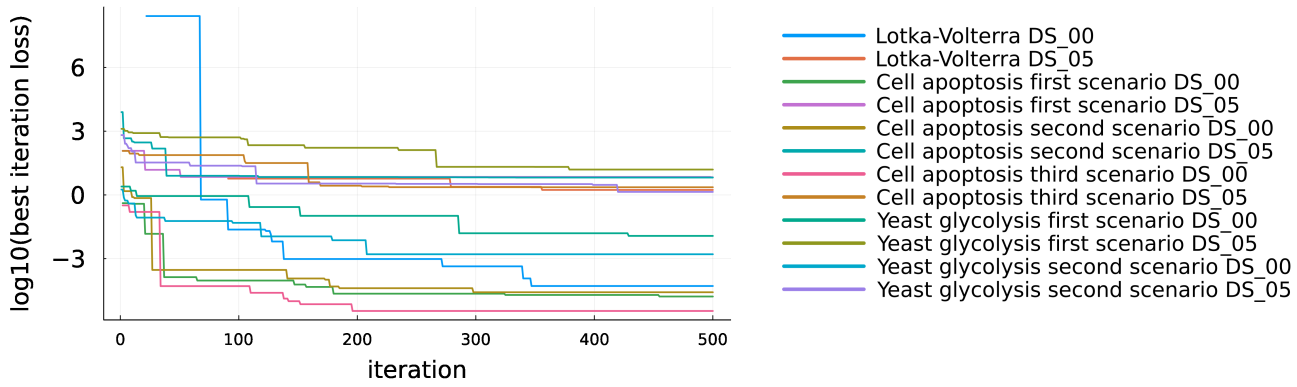


Supplementary Figure 18. Times required to run the pipeline in the yeast glycolysis test case, considering the datasets $DS_{0.00}$ (panel **a**) and $DS_{0.05}$ (panel **b**). For *Step 2a - first stage*, *Step 2a - second stage*, and *Step 2b*, the times represent the median duration for a single trial, with error bars indicating the first and third quartiles. For *Step 3* and *Step 4*, the uncertainty in execution time was assessed by performing 100 runs of each step, with the median and interquartile range shown in the plot. The various scenarios reported are described in the main text.

To mitigate the computational time required to run the pipeline, we propose two directions: (1) reducing the total number of training epochs for each model, and (2) lowering the computational cost associated with the TPE during hyperparameter tuning.

In the current setup, we fixed the number of training epochs across all test cases to ensure convergence in every scenario, prioritizing the convergence over computational efficiency. However, a potential improvement could involve the introduction of early stopping during the HNODE model training. Early stopping could serve a dual purpose by acting also as a regularizer, thereby avoiding the need to tune the $L2$ regularization factor (*Step 2a - second stage*).

As discussed previously, the execution of the TPE in *Step 2a - First Stage* could become a bottleneck in the pipeline, since trials in this stage are required to run sequentially. Currently, we fix the total number of trials to 500 across all test cases to ensure convergence in each scenario. However, in many cases, the optimal trial is achieved before reaching 500 iterations (see Supplementary Fig. 19). To address this, we could employ a stopping criterion to terminate execution once convergence is achieved. Additionally, within each trial, pruning strategies already available in the *Optuna* framework[10] could be used to discard non-promising trials, further reducing computational overhead.



Supplementary Figure 19. Profile of the best loss achieved during the execution of the TPE across the different scenarios considered in this work. The flat profiles indicate that the optimal hyperparameters for the HNODE models are typically found before reaching the full 500 iterations.

Supplementary Note 27: Analysis of the neural network impact on the system dynamics in the considered test-cases

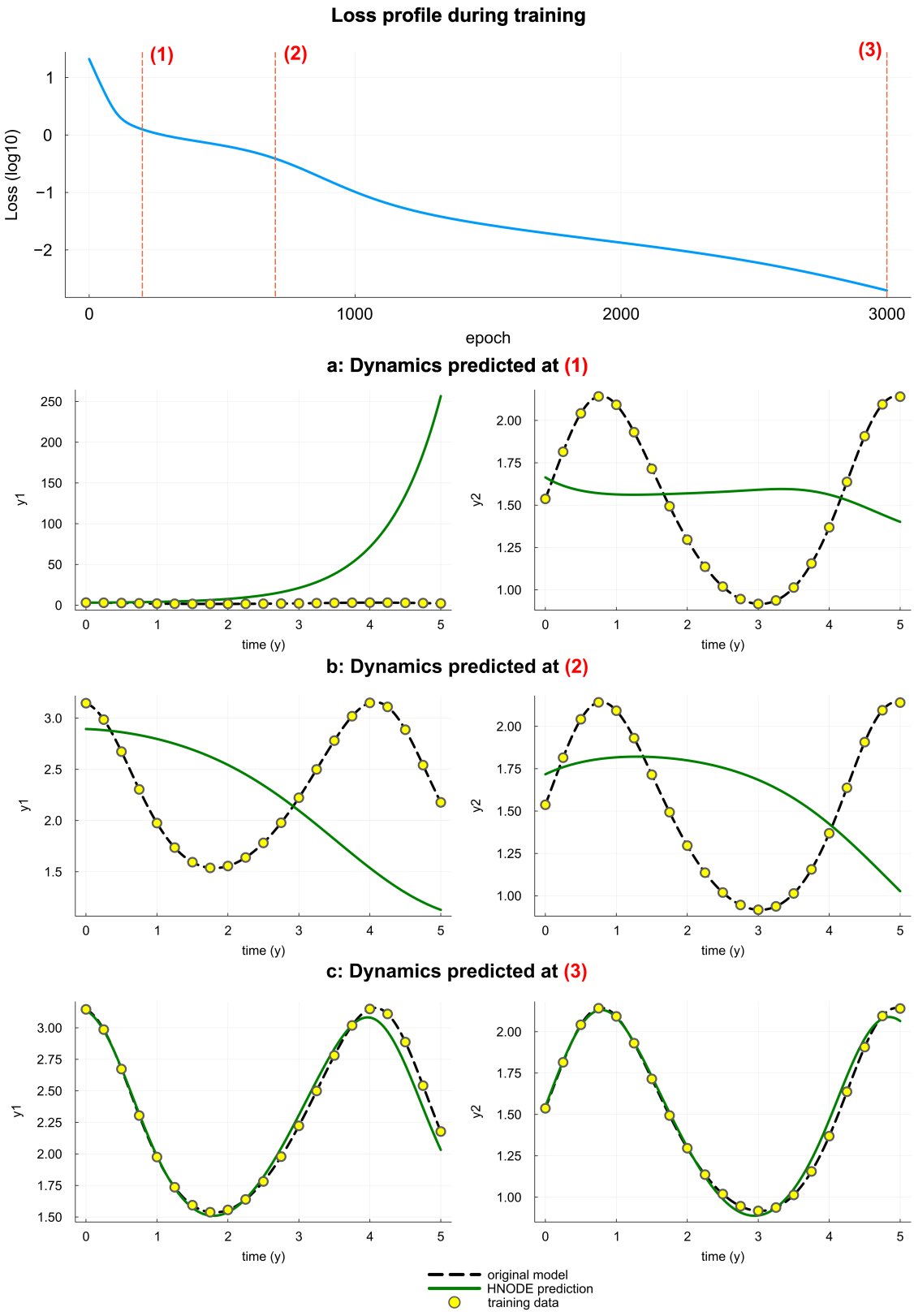
The goal of this section is to show that, in the test cases considered in this work, the neural network used to replace the assumed unknown parts of the system affects the dynamics of the entire system. This impact is therefore not limited to the variables whose equations are replaced, either fully or partially, by the neural network. To demonstrate this, we fix the mechanistic parameters of the HNODE model to their ground truth values and train only the neural network, showing that the training process impacts all the dynamics of the system variables. For simplicity, this analysis is performed for all test cases on the dataset without error ($DS_{0.0}$) in the first scenario, where it is assumed that all variables are observable.

The neural network architecture hyperparameters are fixed to their tuned values (see Supplementary Table 3, 8, and 17 for Lotka-Volterra, cell apoptosis, and yeast glycolysis respectively). The other hyperparameters have been manually tuned and are described in Supplementary Table 23. Since the aim of the analysis is to demonstrate how the neural network influences the dynamics of all system variables, we do not focus on the convergence of the training and perform 3000 epochs with Adam optimizer for all cases.

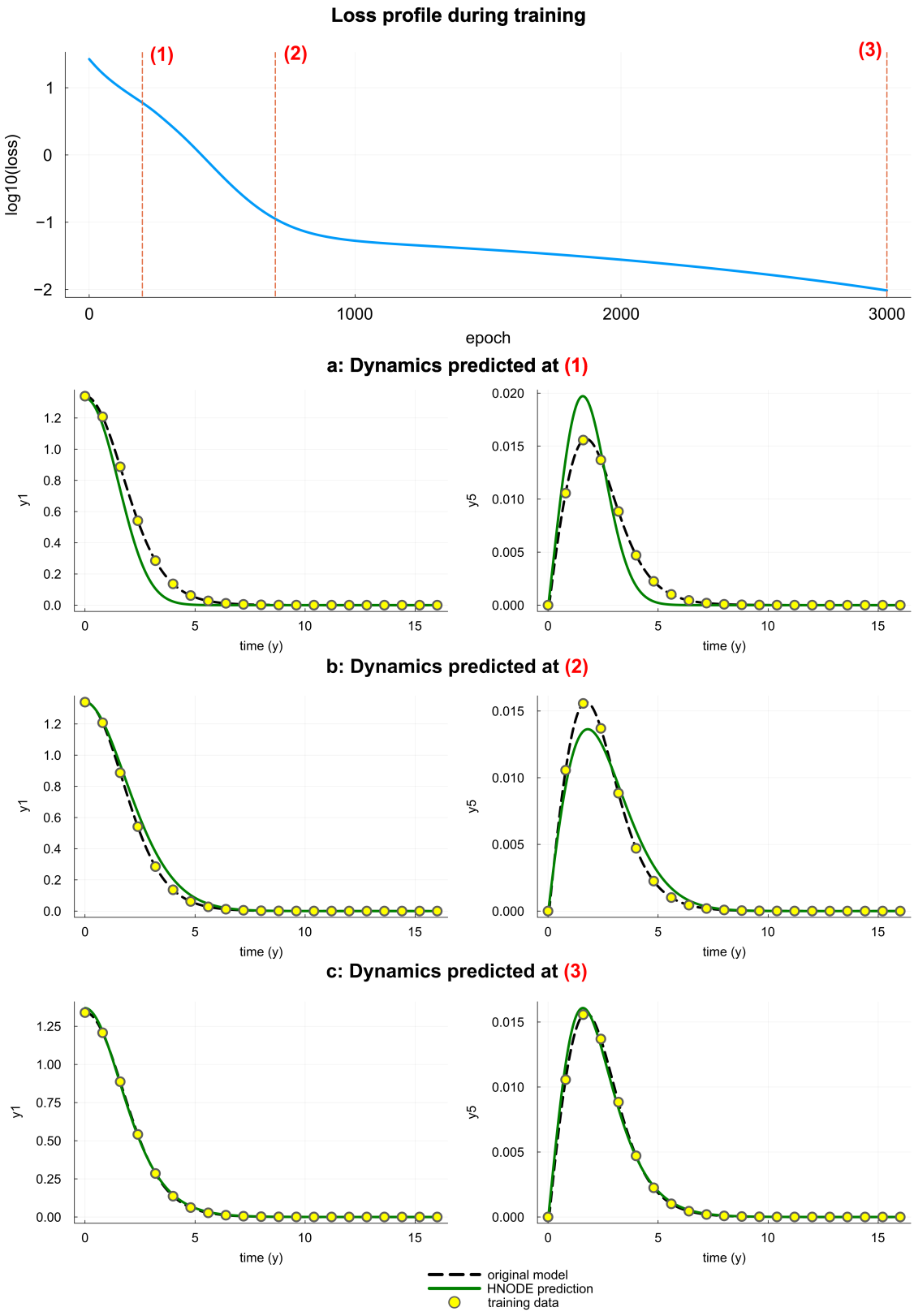
	Test-case		
Name	**Lotka-Volterra**	**Cell apoptosis**	**Yeast glycolysis**
learning rate	$1.0 \cdot 10^{-3}$	$5.0 \cdot 10^{-4}$	$1.0 \cdot 10^{-5}$
k (segmentation)	2	5	5
ρ	$3.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	0.0

Supplementary Table 23. Hyperparameter values used for training the HNODE models with mechanistic parameters fixed to their ground truth values.

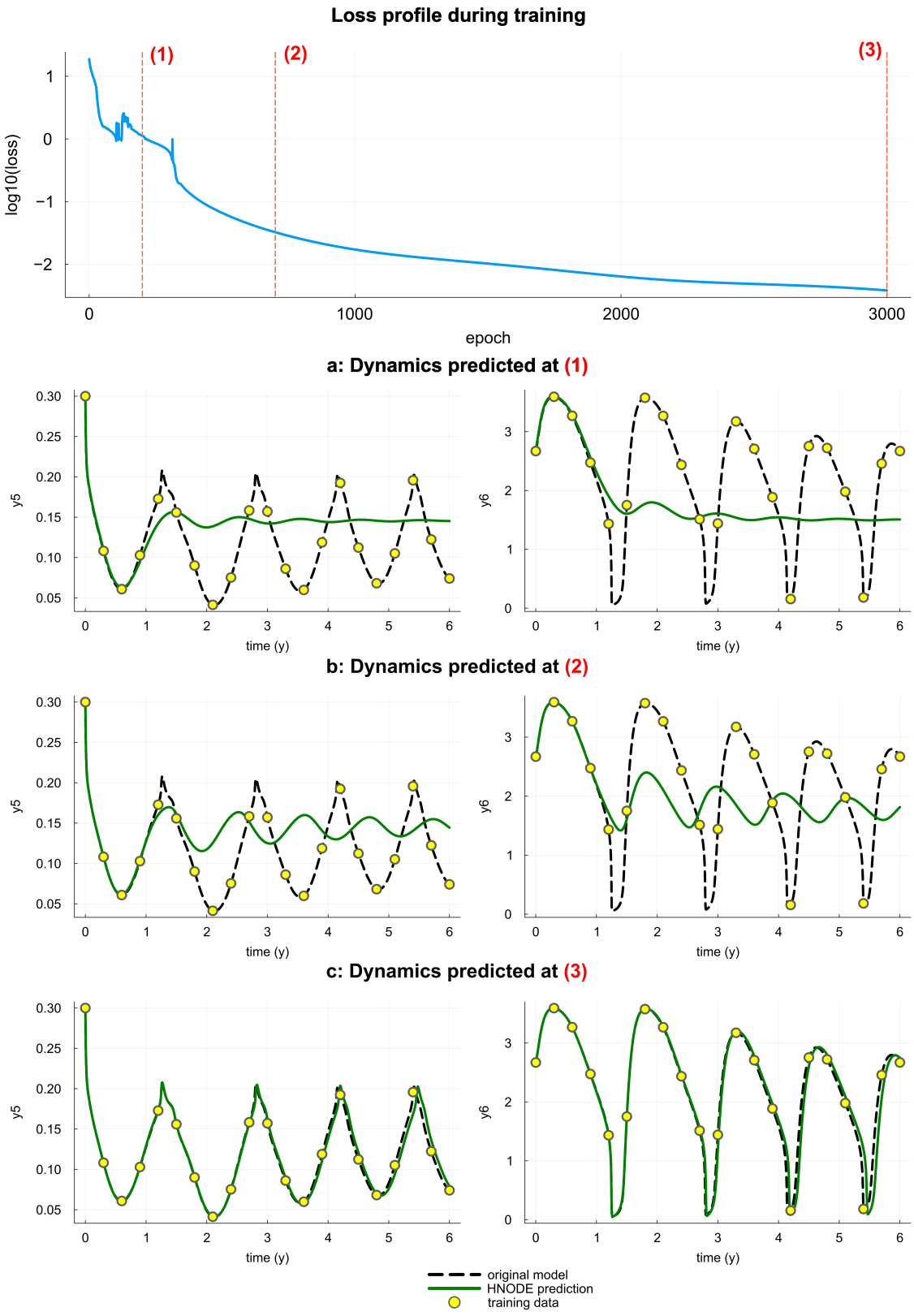
The results of the analysis for selected variables in the Lotka-Volterra, cell apoptosis, and yeast glycolysis test cases are presented in Supplementary Figs. 20, 21, and 22, respectively. In the Lotka-Volterra case, the neural network is involved in the equations governing both system variables, so it is expected that the training process impacts the dynamics of both variables. However, the results indicate that in the other two test cases as well, the training process affects the dynamics of variables whose equations are not directly modeled by the neural network.



Supplementary Figure 20. The training process of the HNODE model with mechanistic parameters fixed to their ground truth values in the Lotka-Volterra case is conducted on $DS_{0.00}$. The dynamics of the variables y_1 and y_2 predicted by the HNODE model after 200, 700, and 3000 training epochs are shown in panels **a**, **b**, and **c**, respectively.



Supplementary Figure 21. The training process of the HNODE model with mechanistic parameters fixed to their ground truth values in the cell apoptosis test case is conducted on $DS_{0.00}$. The neural network in the HNODE model describes the derivative of y_4. The dynamics of the variables y_1 and y_5 predicted by the HNODE model after 200, 700, and 3000 training epochs are shown in panels **a**, **b**, and **c**, respectively.



Supplementary Figure 22. The training process of the HNODE model with mechanistic parameters fixed to their ground truth values in the Yeast glycolysis test case is conducted on $DS_{0.00}$. The neural network in the HNODE model describes the derivative of y_1. The dynamics of the variables y_5 and y_6 predicted by the HNODE model after 200, 700, and 3000 training epochs are shown in panels **a**, **b**, and **c**, respectively.

References

1. Raue, A., Karlsson, J., Saccomani, M. P., Jirstrand, M. & Timmer, J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics* **30**, 1440–1448 (2014).
2. Chis, O.-T., Banga, J. R. & Balsa-Canto, E. Structural identifiability of systems biology models: a critical comparison of methods. *PloS one* **6**, e27755 (2011).
3. Raue, A. *et al.* Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929 (2009).
4. Anstett-Collin, F., Denis-Vidal, L. & Millérioux, G. A priori identifiability: An overview on definitions and approaches. *Annu. Rev. Control.* **50**, 139–149 (2020).
5. Rackauckas, C. *et al.* Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385* (2020).
6. Aldridge, B. B., Haller, G., Sorger, P. K. & Lauffenburger, D. A. Direct lyapunov exponent analysis enables parametric study of transient signalling governing cell behaviour. *IEE Proceedings-Systems Biol.* **153**, 425–432 (2006).
7. Kim, S., Ji, W., Deng, S., Ma, Y. & Rackauckas, C. Stiff neural ordinary differential equations. *Chaos: An Interdiscip. J. Nonlinear Sci.* **31**, 093122 (2021).
8. Quaiser, T. & Mönnigmann, M. Systematic identifiability testing for unambiguous mechanistic modeling–application to jak-stat, map kinase, and nf-κ b signaling pathway models. *BMC systems biology* **3**, 1–21 (2009).
9. Ruoff, P., Christensen, M. K., Wolf, J. & Heinrich, R. Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations. *Biophys. chemistry* **106**, 179–192 (2003).
10. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2623–2631 (2019).