npj | systems biology and applications

Article

Published in partnership with the Systems Biology Institute

https://doi.org/10.1038/s41540-024-00460-3

Robust parameter estimation and identifiability analysis with hybrid neural ordinary differential equations in computational biology

Stefano Giampiccolo[1,2], Federico Reali[1], Anna Fochesato[1,3,5], Giovanni Iacca[2] & Luca Marchetti[1,4]

Parameter estimation is one of the central challenges in computational biology. In this paper, we present an approach to estimate model parameters and assess their identifiability in cases where only partial knowledge of the system structure is available. The partially known model is embedded into a system of hybrid neural ordinary differential equations, with neural networks capturing unknown system components. Integrating neural networks into the model presents two main challenges: global exploration of the mechanistic parameter space during optimization and potential loss of parameter identifiability due to the neural network flexibility. To tackle these challenges, we treat biological parameters as hyperparameters, allowing for global search during hyperparameter tuning. We then conduct *a posteriori* identifiability analysis, extending a well-established method for mechanistic models. The pipeline performance is evaluated on three test cases designed to replicate real-world conditions, including noisy data and limited system observability.

Mathematical and computational models are increasingly employed in the study of biological systems[1]. These models not only facilitate the creation of predictive and explanatory tools[2] but also offer a means to understand the interactions among the variables of the system[3].

The well-established approach to develop a mathematical model for biological systems, i.e., the mechanistic modeling approach[4], relies on encoding known biological mechanisms into systems of partial or ordinary differential equations (PDE or ODE) using suitable kinetic laws, e.g., the law of mass action or Michaelis-Menten kinetics[5]. These equations typically incorporate several unknown model parameters, making parameter estimation a central challenge in model development[6]. The prevailing approach to parameter estimation involves optimizing the model dynamics to align with experimental data[7]. Various optimization techniques are employed for this purpose, including linear and nonlinear least squares methods[8,9], genetic and evolutionary algorithms[7], Bayesian Optimization[10,11], control theory-derived approaches[12] and, more recently, physics-informed neural networks[6]. The scarcity of experimental data and the measurement noise often lead to non-identifiability issues, where the optimization problem lacks a unique solution[13]. To address this challenge, various methods exist to analyze the identifiability of model parameters. *Structural identifiability* analysis[14], performed before parameter estimation, analyzes the structure of the model to determine if parameters can be uniquely estimated. On the other hand, *practical identifiability* analysis[15], conducted after parameter estimation, evaluates how uncertainties in experimental measurements affect parameter estimation.

Developing mathematical models and estimating model parameters with the mechanistic modeling approach presents significant challenges due to the need for a detailed understanding of the interactions between (and within) biological systems. Complex biological systems often involve processes at different scales, such as genetic, molecular, tissue, organ, or whole-body, and such intricate mechanistic details are only partially known[16,17]. To overcome the limits of mechanistic modeling, hybrid models that combine mechanistic ODE-based dynamics with neural network components have grown in popularity in various scientific domains[18,19]. The integration between neural networks and ODE systems is known by different names, such as *Hybrid Neural Ordinary Differential Equations* (HNODEs)[20], *graybox modeling*[21,22], or *universal differential equations*[23]. In a HNODE, neural networks are used as universal approximators to represent unknown

[1]Fondazione The Microsoft Research-University of Trento Centre for Computational and Systems Biology (COSBI), Rovereto, Trento, Italy. [2]Department of Information Engineering and Computer Science (DISI), University of Trento, Povo, Trento, Italy. [3]Department of Mathematics, University of Trento, Povo, Trento, Italy. [4]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Povo, Trento, Italy. [5]Present address: Roche Pharma Research and Early Development, Roche Innovation Center Basel, Basel, Switzerland. e-mail: marchetti@cosbi.eu; luca.marchetti@unitn.it

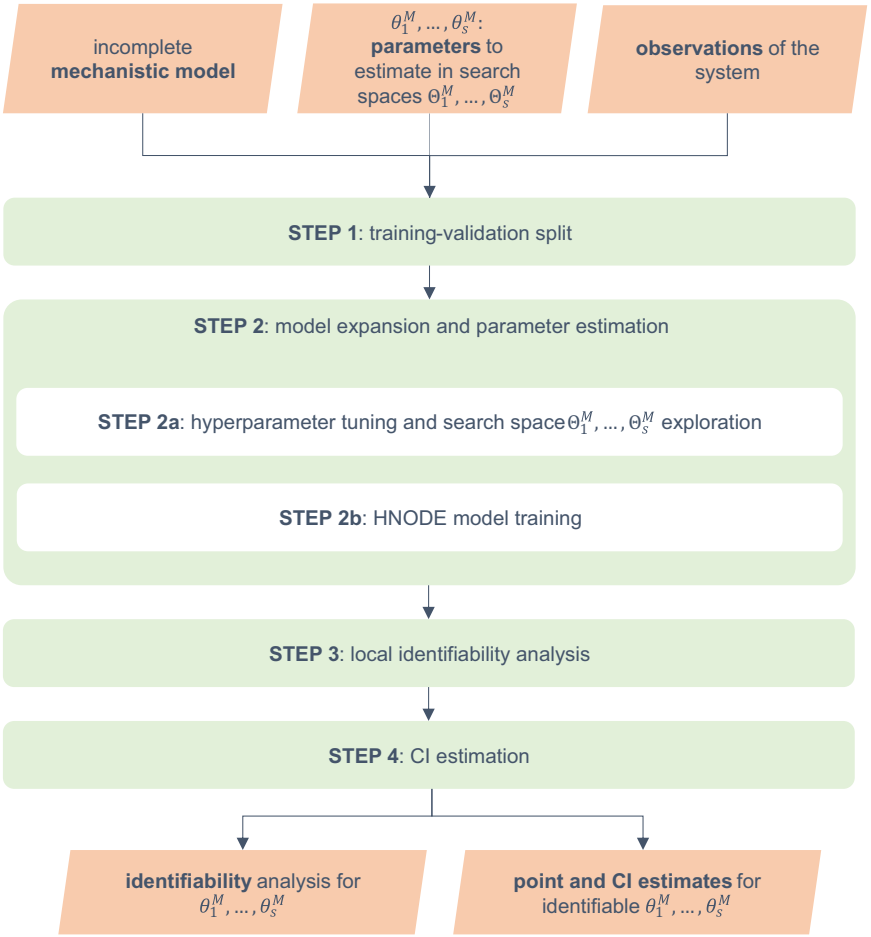

Fig. 1 | Schematic representation of the workflow. In the workflow schema, $\theta_1^M, \ldots, \theta_s^M$ denote the mechanistic parameters to estimate in the corresponding search spaces $\Theta_1^M, \ldots, \Theta_s^M$. HNODE: Hybrid Neural Ordinary Differential Equation, CI: Confidence Interval.

portions of the system. Mathematically, HNODE can be formulated as the following ODE:

$$\frac{d\mathbf{y}}{dt}(t) = f(\mathbf{y}, NN(\mathbf{y}), t, \boldsymbol{\theta}) \qquad \mathbf{y}(0) = \mathbf{y_0}, \tag{1}$$

where *NN* denotes the neural network component of the model, f encodes the mechanistic knowledge of the system, and $\boldsymbol{\theta}$ represents a vector of unknown mechanistic parameters. This approach has proven successful in several fields, and early results in isolated and relatively simple scenarios have shown promise in computational biology as well[24,25]. However, modeling with HNODEs remains an active area of research, and best practices for estimating mechanistic parameters and assessing their identifiability within this framework remain to be outlined.

Mechanistic parameter estimation within the HNODE framework presents significant challenges. Firstly, while model calibration in mechanistic models usually relies on global optimization techniques to explore the parameter search space[9], training HNODE models necessitates the use of local and gradient-based methods[23]. Secondly, incorporating a universal approximator, such as a neural network, into a dynamical model may compromise the identifiability of the HNODE mechanistic components[26,27]. In this sense, the existing literature has focused on trying to enforce the identifiability of the mechanistic parameters within a HNODE by integrating a regularization term into the cost function[28]. Common choices for this regularization term include minimizing the impact of the neural network on the model[28] or ensuring that the outputs of the neural network and mechanistic part are uncorrelated[26]. However, these approaches do not always guarantee the correct identification of the mechanistic part, and the outcomes depend on the specific regularization term used[26]. To the best of our knowledge, the identifiability analysis of the mechanistic parameters in a HNODE model has not been investigated in the literature so far.

In this contribution, we present an end-to-end approach for mechanistic parameter estimation and identifiability analysis in scenarios where mechanistic knowledge about the system is incomplete. Initially, we focus on tuning hyperparameters to embed our incomplete mechanistic model into a HNODE model that is capable of effectively capturing the experimental dynamics. Subsequently, we proceed to compute mechanistic parameter estimates by training the HNODE model. Following parameter estimation, we extend a well-established approach for mechanistic models to assess the parameter identifiability *a posteriori*. For identifiable parameters, we finally estimate asymptotic confidence intervals (CIs).

The proposed approach has been tested in three different in silico scenarios, that have been constructed by assuming a lack of information about some portions of known mechanistic models. We aimed to replicate typical conditions found in real-world scenarios, with a noisy and scattered training set describing the time evolution of a subset of the model variables. Firstly, we consider the traditional Lotka-Volterra model for predator-prey interactions. Despite being a relatively simple model, it shows the ability of the approach to identify compensations between the neural network and the mechanistic component of the HNODE model. Secondly, we evaluate the performances of our approach on a model for cell apoptosis[29], known for the non-identifiability of part of its parameters. Thirdly, we consider a model of oscillations in yeast glycolysis[30], which has been frequently employed as a benchmark for inference in computational systems biology due to its nonlinear oscillatory dynamics.

Results

A schematic representation of the workflow is presented in Fig. 1. The input of the pipeline consists of an incomplete mechanistic model containing parameters to be estimated, along with a time series dataset containing experimental observations of some of the system variables. By embedding the incomplete model into a HNODE model, our approach enables parameter estimation and identifiability analysis.

The workflow starts by splitting the observation time points into training and validation sets (Step 1). In the second step, using this partition, we expand the incomplete mechanistic model into a HNODE model. We employ Bayesian Optimization to simultaneously tune the model hyperparameters and explore the mechanistic parameter search space (Step 2a). The model is then fully trained (Step 2b), yielding mechanistic parameter estimates. In the next step (Step 3), we assess the local identifiability at-a-point of the parameters. For the locally identifiable ones, we proceed to estimate confidence intervals (Step 4).

In the rest of the section, we use the following notation. Let $\mathbf{y}(t, \boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}, \mathbf{y_0}) \in \mathbb{R}^n$, with $n \in \mathbb{N}^+$, be the HNODE model defined by the following differential equation:

$$\begin{cases} \dfrac{d\mathbf{y}}{dt} = f^M(\mathbf{y}, t, \boldsymbol{\theta}^M) + NN(\mathbf{y}, t, \boldsymbol{\theta}^{NN}) \\ \mathbf{y}(t_0) = \mathbf{y_0} \end{cases} \tag{2}$$

where $\mathbf{y_0} \in \mathbb{R}^n$ stands for the initial conditions, f^M and $\boldsymbol{\theta}^M = (\theta_1^M, \ldots, \theta_s^M)$ denote the mechanistic part of the model and the $s \in \mathbb{N}^+$ mechanistic parameters to be estimated, respectively, while $\boldsymbol{\theta}^{NN}$ represents the neural network parameters. We denote the i-th component of the model as $y^i(t, \boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}, \mathbf{y_0})$, with $i = 1, \ldots, n$. Typically, only a subset of the system variables is observable; we indicate this subset with $O \subseteq \{1, \ldots, n\}$. We assume to have access to the experimental measurements of the observable variables at $m+1$ time points $t_0, \ldots, t_m$; $\hat{y}_j^i \pm \hat{\sigma}_j^i$ indicates the observations of the i-th variable at time t_j, with the uncertainty reflecting the variability of the measurement. Additionally, we assume to have access to the initial conditions $\hat{y}_0^i \pm \hat{\sigma}_0^i$ of all the variables.

The remainder of this section is organized as follows. First, we present a detailed description of each step in the workflow. In the subsequent subsections, we analyze the results of the pipeline across various in silico test cases. The implementation details and algorithmic setup for these test cases are outlined in the Methods Section. Given that the training of HNODE extends the techniques used for Neural Ordinary Differential Equations (NODE)[31], we begin with an overview of the methods employed for training NODE, which precedes the description of the pipeline.

Background: NODE training

In this section, to discuss the NODE case, we assume that $f_M = \mathbf{0}$ in Eq. (2), indicating no mechanistic knowledge about the system. Under this hypothesis, the vector field is entirely parameterized by the neural network NN, and the ODE:

$$\begin{cases} \frac{d\mathbf{y}}{dt} = NN(\mathbf{y}, t, \boldsymbol{\theta}^{NN}) \\ \mathbf{y}(t_0) = \mathbf{y_0} \end{cases} \tag{3}$$

represents a NODE. Given a loss function $\mathcal{L}$, the training of the NODE can be achieved through the minimization of the cost function:

$$C(\boldsymbol{\theta}^{NN}) = \frac{1}{m}\sum_{j=1}^{m}\sum_{i\in O}\mathcal{L}(y^i(t_j, \theta^{NN}, \mathbf{y_0}), \hat{y}_j^i), \tag{4}$$

where y^i is computed numerically by integrating Eq. (3). Minimizing $C(\boldsymbol{\theta}^{NN})$ requires to back-propagate the error through the ODE solver algorithm used for the numerical integration of y. By the chain rule, for a differentiable $\mathcal{L}$, this amounts to compute the gradients

$$\frac{\partial y^i(t_j, \boldsymbol{\theta}^{NN}, \mathbf{y_0})}{\partial \boldsymbol{\theta}^{NN}}.$$

Chen et al.[31] demonstrated that these gradients can be efficiently computed using adjoint sensitivity[32,33], treating the ODE solver as a black box. Adjoint sensitivity requires a backward integration of the system and different numerical methods have been proposed to efficiently calculate it[23].

As in the traditional ODE case, stiffness constitutes a significant challenge in the training of NODE[34]. However, as shown by Kim et al.[35], there are specific ways to overcome this issue. These include employing deep neural network architectures, ad hoc methods for computing adjoint sensitivity, and a normalized loss function.

Step 1: training-validation split

In the first step of the workflow, the observation time points are divided into training and validation sets, indicated with $\{t_j\}_{j\in T}$ and $\{t_j\}_{j\in V}$ respectively. The validation time points are chosen from $t_1, \ldots, t_m$ to ensure a homogeneous distribution along the observed trajectory.

Step 2: model expansion and parameter estimation

In this step, the incompletely specified mechanistic model is expanded into a HNODE model through the use of neural networks to replace unknown portions of the system. The estimates for mechanistic parameters are derived by training the HNODE model.

The training of a HNODE model is analogous to that of NODE. Adopting a loss function $\mathcal{L}$, the most straightforward approach would be minimizing the cost function:

$$C(\boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}) = \frac{1}{|T|\cdot|O|}\sum_{j\in T}\sum_{i\in O}\mathcal{L}(y^i(t_j, \boldsymbol{\theta}^M, \theta^{NN}, \mathbf{y_0}), \hat{y}_j^i). \tag{5}$$

Here, $|T|$ and $|O|$ denote the cardinality of T and O respectively. Gradients of $C(\boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN})$ with respect to mechanistic and neural network parameters are calculated via adjoint sensitivity methods. If the cost function defined by Eq. (5) is employed, a single trajectory of $\mathbf{y}$ spanning from t_0 to t_m is computed at each epoch of the training. This approach, known as single shooting, is often suboptimal due to the potential risks of the training getting stuck in local minima[27].

Therefore, we adopt the multiple shooting technique (MS)[36]. In MS, the time interval (t_0, t_m) is partitioned into different segments. Rather than computing a single trajectory of $\mathbf{y}$ on the entire training interval, a potentially discontinuous trajectory $\mathbf{y}_{PW}(t, \boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN})$ is reconstructed piece-wise by solving an initial value problem on each segment, adding the states of the system at the initial points of each segment as additional parameters to be optimized. The cost function is then composed of two terms:

$$C(\boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}) = C_{FIT}(\boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}) + \rho \cdot C_{DIS}(\boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}). \tag{6}$$

In Eq. (6), C_{FIT} measures the difference between the piecewise-defined trajectory and the training data, C_{DIS} measures the discontinuity of the trajectory, and $\rho \in \mathbb{R}$ is a hyperparameter. For C_{FIT}, we use the cost function defined in Eq. (5), adopting the loss function

$$\mathcal{L}\left(y_{PW}^i\left(t_j, \boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}\right), \hat{y}_j^i\right) = \begin{cases} \left(y_{PW}^i(t_j, \boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}) - \hat{y}_j^i\right)^2 / \left(\hat{\sigma}_j^i\right)^2 & \text{if } \hat{\sigma}_j^i \neq 0 \\ \left(y_{PW}^i(t_j, \boldsymbol{\theta}^M, \boldsymbol{\theta}^{NN}) - \hat{y}_j^i\right)^2 & \text{if } \hat{\sigma}_j^i = 0 \end{cases} \tag{7}$$

in which the quadratic loss is weighted for the uncertainty measure if it is not zero. C_{DIS} is computed by summing the squared values of the discontinuities at the extremes of the training interval partition. To prevent overfitting, in the case of noisy training data, we add an L2 regularization term. Given this cost function, the training is performed with a gradient-based optimizer,

and gradients are computed using adjoint sensitivity methods. In our test cases, we use the Adam optimizer[37], and we refine the results with L-BFGS[38] on training datasets without noise.

To define the segmentation of the time interval, we partition the training set of time points into subsets containing the same number k of items (excluding the last interval), where k is a hyperparameter. When using MS to train the model, we introduce as additional parameters to optimize the state of the system at the initial points of each segment. These parameters are initialized with the measured states of the system at the time point when the corresponding system variable is observable; otherwise, they are initialized with the initial state of the variable.

How to tune the hyperparameters and explore the parameter search space. The hyperparameters requiring tuning can be grouped into three different macro-categories: the hyperparameters of the architecture of the neural network, the starting values of the mechanistic parameters to initiate the gradient-based optimization, and the hyperparameters related to the MS training. This latter category includes the segmentation of the time interval (t_0, t_m) used to define the piecewise trajectory, the weight factor ρ used in Eq. (6), the weight λ of the $L2$ regularization term, and the learning rate of the gradient-based method. Incorporating the initial values of $\theta_1^M, \ldots, \theta_s^M$ into hyperparameter tuning enables the integration of local gradient-based search with Bayesian Optimization techniques, as proposed by Gao et al.[39] The objective is to globally explore the mechanistic parameter search space during this step.

The hyperparameters are tuned in two stages, described below.

Hyperparameter tuning - Stage 1. We simultaneously tune all hyperparameters, except for the $L2$ regularization factor λ, using the multivariate Tree-Structured Parzen Estimator (TPE)[40,41]. In each TPE trial, the HNODE model is trained for a limited number of MS iterations. The optimization metric during the TPE algorithm is the mean squared error on the validation time points of the trajectory predicted by the trained model:

$$\frac{1}{|V|\cdot|O|}\sum_{j\in V}\sum_{i\in O}\left(\frac{y^i(t_j,\boldsymbol{\theta}^M,\boldsymbol{\theta}^{NN},\mathbf{y_0})-\hat{y}_j^i}{\hat{\sigma}_j^i}\right)^2.$$

As before, we employ a pure quadratic error when uncertainty indications on measurements are absent.

Hyperparameter tuning - Stage 2. In the second stage, we fix the hyperparameters tuned in the first step and focus on tuning the $L2$ regularization factor, λ, using a grid-search approach. In each trial during this step, the HNODE model is trained for an extended number of MS iterations, and we continue to employ the same metric as in the previous step to evaluate each trial.

The decision to split the tuning process into two stages is motivated by the observation that the effects of overfitting may not become apparent with a low number of iterations. Therefore, tuning λ in the initial stage might not be as effective. Conversely, conducting a large number of MS iterations in the first stage would significantly increase computational costs.

How to deal with stiffness. Before starting the hyperparameter tuning phase, the behavior of the HNODE model, whether stiff or non-stiff, is unknown. To train a stiff HNODE model, we follow the guidelines outlined for stiff NODE. This involves employing specialized techniques such as ad hoc adjoint sensitivity methods, deeper neural network architectures, stiff ODE solvers, and normalized loss functions. We do not include all configurations used for stiff systems as hyperparameters to tune, to prevent an excessive increase in the search space dimension. Instead, we propose initiating the hyperparameter tuning with non-stiff configurations and monitoring the integration of the HNODE model during the first trials of the process, such as observing the number of steps taken by the ODE solver to complete system integration. If stiffness is detected, we finalize the tuning process and we train the HNODE model using stiff configurations.

Step 3: local identifiability analysis

In this step, we investigate *a posteriori* the local identifiability of the mechanistic parameters $\theta_1^M, \ldots, \theta_s^M$ estimated with the model training. This is particularly important in HNODE models, as incorporating a universal approximator, such as a neural network, into a dynamical model may compromise the identifiability of mechanistic parameters. The following subsections introduce the foundational concepts for our approach and the description of the workflow.

Local identifiability at-a-point and sloppy directions. We refer to the combined neural network and mechanistic parameter estimates, obtained through the training in Step 2, as the single vector:

$$\overline{\boldsymbol{\theta}} = (\overline{\boldsymbol{\theta}}^M, \overline{\boldsymbol{\theta}}^{NN}) = (\bar{\theta}_1, \ldots, \bar{\theta}_p).$$

We recall that the k-th parameter of the model is *locally identifiable at-a-point*[42] in $\overline{\boldsymbol{\theta}}$ if

$$y^i(t_j,\boldsymbol{\theta},\mathbf{y_0}) = y^i(t_j,\overline{\boldsymbol{\theta}},\mathbf{y_0}) \qquad \forall\, j \in \{1,\ldots,\mathrm{m}\}, i \in O$$

implies $\theta_k = \bar{\theta}_k$ for all $\boldsymbol{\theta}$ in a suitable neighborhood of $\overline{\boldsymbol{\theta}}$ (for a discussion of the relationship between this and other definitions of parameter identifiability, we refer to Supplementary Note 1).

To study the local identifiability, we quantify the change in the HNODE model behavior, as parameters $\boldsymbol{\theta}$ vary from $\overline{\boldsymbol{\theta}}$, with the function:

$$\chi(\boldsymbol{\theta}) = \frac{1}{m}\frac{1}{|O|}\sum_{j=1}^{m}\sum_{i\in O}\left(\frac{y^i(t_j,\boldsymbol{\theta},\mathbf{y_0})-y^i(t_j,\overline{\boldsymbol{\theta}},\mathbf{y_0})}{\gamma_i}\right)^2, \tag{8}$$

where $\gamma_i = max_j\{|y^i(t_j,\overline{\boldsymbol{\theta}},\mathbf{y_0})|\}$ is introduced to normalize the contribution to χ of each variable. In a neighborhood of $\overline{\boldsymbol{\theta}}$, the function χ can be approximated as:

$$\chi(\boldsymbol{\theta}) \approx \frac{1}{2}\cdot(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}})^t\cdot H_\chi(\overline{\boldsymbol{\theta}})\cdot(\boldsymbol{\theta}-\overline{\boldsymbol{\theta}}),$$

where $H_\chi(\overline{\boldsymbol{\theta}})$ denotes the Hessian matrix of χ in $\overline{\boldsymbol{\theta}}$. If $\mathbf{v}$ is an eigenvector of $H_\chi(\overline{\boldsymbol{\theta}})$ and μ is the corresponding eigenvalue, for small values of α it holds:

$$\chi(\bar{\boldsymbol{\theta}}+\alpha\cdot\mathbf{v}) \approx \frac{1}{2}\cdot\alpha^2\cdot\mu. \tag{9}$$

Eq. (9) implies that the HNODE model behavior does not change significantly when the parameters move along the direction identified by $\mathbf{v}$ if μ is zero. In mechanistic models, the eigenvectors related to zero or almost-zero eigenvalues are commonly called sloppy directions[43–45]. Denoting with V_0 the null space of H_χ, spanned by the eigenvectors related to zero eigenvalues, and with V_S the subspace spanned by the remaining eigenvectors, each parameter θ_k of the model can be decomposed as:

$$\theta_k = \pi_0(\theta_k) + \pi_S(\theta_k) \qquad \text{with} \qquad ||\pi_0(\theta_k)||^2 + ||\pi_S(\theta_k)||^2 = 1$$

where $\pi_0(\theta_k)$ and $\pi_S(\theta_k)$ denote the projection of θ_k onto V_0 and V_S respectively. $\pi_0(\theta_k)$ represents the local direction that maximizes the parameter perturbation without affecting the model behavior. Although one might expect identifiability of θ_k to yield $\|\pi_0(\theta_k)\|^2 = 0$ and $\|\pi_S(\theta_k)\|^2 = 1$, strict equalities generally do not hold. This is because, as observed by Gutenkunst et al.[43], each sloppy eigenvector typically encompasses mixed

components of nearly all model parameters. Hence, $0 < \|\pi_0(\theta_k)\|^2 < 1$ applies to all model parameters.

How to assess the local identifiability at-a-point. Our approach involves identifying mechanistic parameters $\theta_1^M, \ldots, \theta_s^M$ having a significant projection onto the null subspace of the HNODE model. We proceed as follows. First, we compute $H_\chi(\overline{\boldsymbol{\theta}})$ with the Gauss-Newton method:

$$H_\chi^{k,l}(\overline{\boldsymbol{\theta}}) = \frac{2}{m \cdot |O|} \sum_{j=1}^{m} \sum_{i \in O} \frac{1}{(\gamma_i)^2} \frac{\partial y^i}{\partial \theta_k}(t_j, \overline{\boldsymbol{\theta}}, \mathbf{y_0}) \cdot \overline{\theta}_k \cdot \frac{\partial y^i}{\partial \theta_l}(t_j, \overline{\boldsymbol{\theta}}, y_0) \cdot \overline{\theta}_l. \tag{10}$$

In Eq. (10), we rescale the sensitivity coefficients to make them independent of the absolute values of the parameters[46]; this consideration is important when simultaneously considering both neural network and mechanistic parameters. $H_\chi(\overline{\boldsymbol{\theta}})$ is a symmetric and positive semi-definite matrix. We proceed to determine the null subspace of $H_\chi(\overline{\boldsymbol{\theta}})$ spanned by all eigenvectors corresponding to eigenvalues $\mu \leq \epsilon$. Here, ϵ is a hyperparameter embodying the threshold used to distinguish zero eigenvalues in a numerical environment. To discriminate whether there exists a direction in the parameter search space along which we can significantly perturb the parameter θ_i^M without affecting the model simulation, we analyze the norm of the projection of θ_i^M onto V_0. If it exceeds a predefined threshold δ, which is a second hyperparameter of our approach, we classify the parameter as non-identifiable.

Step 4: CI estimation

In the final step of our pipeline, we estimate CIs for the locally identifiable parameters. Several methods for estimating parameter CIs in mechanistic models have been proposed. We employ the Fisher Information Matrix (FIM)-based approach[47]. We assume that our observed data follow independent Gaussian distributions $N(y_j^i, \sigma_j^i)$, and that our parameter estimator is an approximation of the maximum likelihood estimator, as proposed in Yazdani et al.[6]. Therefore, we can approximate the observed FIM as:

$$\mathrm{FIM}^{k,l}(\overline{\theta}) = \sum_j \sum_{i \in O} \frac{1}{(\hat{\sigma}_i)^2} \frac{\partial y^i}{\partial \theta_k}(t_j, \overline{\boldsymbol{\theta}}, \mathbf{y_0}) \cdot \frac{\partial y^i}{\partial \theta_l}(t_j, \overline{\boldsymbol{\theta}}, \mathbf{y_0}). \tag{11}$$

The pseudo-inverse of the FIM provides a lower bound for the covariance matrix of our estimators[48]. Consequently, we can define a lower bound $\overline{\sigma}_i^M$ for the standard deviation of the estimator of the parameter θ_i^M by taking the square roots of the elements on the diagonal of FIM^{-1}. We can then compute the lower bound for the α-level confidence intervals of θ_i^M as:

$$\left(\overline{\theta}_i^M - \Phi_{\frac{1+\alpha}{2}} \cdot \overline{\sigma}_i^M, \overline{\theta}_i^M + \Phi_{\frac{1+\alpha}{2}} \cdot \overline{\sigma}_i^M\right),$$

where Φ_x represents the x quantile of a standard Gaussian distribution.

Lotka-Volterra test case

In this section, we analyze the performance of the pipeline using a test case based on the Lotka-Volterra predator-prey model, which describes the temporal dynamics between two species, where one preys upon the other[49]. Although it is based on a relatively simple system, this test case offers the opportunity to evaluate the effectiveness of our approach in detecting compensations between the neural network and the mechanistic parameters of the model. Additionally, it allows for a discussion on why we opt not to enforce the identifiability of the mechanistic parameters through the use of a regularizer in the cost function, as proposed by Yin et al.[28].

The full mechanistic model is defined by the following system of ODE:

$$\begin{cases} \frac{dy_1(t)}{dt} = \alpha y_1 - \beta y_1 y_2 \\ \frac{dy_2(t)}{dt} = \gamma y_1 y_2 - \delta y_2, \end{cases} \tag{12}$$

where y_1 represents the prey population, and y_2 represents the predator population. In this context, we assume a lack of information regarding the interaction terms between the two species ($\beta y_1 y_2$ and $\gamma y_1 y_2$). By replacing them with a neural network $NN : \mathbb{R}^2 \to \mathbb{R}^2$, we have the following HNODE model:

$$\begin{cases} \frac{dy_1(t)}{dt} = \alpha y_1 + NN([y_1, y_2], \boldsymbol{\theta}^{NN})[1] \\ \frac{dy_2(t)}{dt} = NN([y_1, y_2], \boldsymbol{\theta}^{NN})[2] + \delta y_2. \end{cases} \tag{13}$$

Our objective is to estimate and assess the identifiability of the prey birth rate α, assuming we know the predator decay rate δ. To generate in silico the observation datasets, we numerically integrate Eq. (12) over the interval from $t = 0$ to $t = 5$ years (the initial conditions and parameters are provided in Supplementary Note 2) and sample the states every 0.25 years (resulting in 21 observed time points). We test our pipeline on a noiseless dataset (denoted with $DS_{0.00}$), and on a dataset in which each time series is perturbed with a zero-mean Gaussian noise with a standard deviation equal to 5% of its min-max variation (denoted with $DS_{0.05}$). The pipeline is run independently on the two datasets. Our search space for the mechanistic parameter is $[\overline{\alpha} \cdot 10^{-2}, \overline{\alpha} \cdot 10^{2}]$, where $\overline{\alpha}$ is the ground truth value used for generating the in silico observations.

We run the pipeline assuming non-stiff configurations, employing an explicit Runge-Kutta ODE solver[50] and the *Interpolating Adjoint* method[23] to compute adjoint sensitivity. Additionally, the search space for the neural network architecture is restricted to shallow networks (at most 3 hidden layers) with *tanh* as the activation function. The tuned hyperparameters, together with the search spaces, are reported in Supplementary Table 3. The dynamics predicted by the fully trained HNODE models are illustrated in Fig. 2, demonstrating the effective fitting to the observation data points. The resulting α estimates are reported in Table 1.

To assess the identifiability of α, we analyze its projection onto the null subspace of H_χ (Fig. 3). In both HNODE models, the norm of the projection exceeds δ, indicating the non-identifiability of the mechanistic parameter. These projections consist of both α and neural network parameters, suggesting a compensation between the neural network and the mechanistic parameter. This aligns with our expectations, as the first equation of Eq. (13) consists of a sum between a neural network and a mechanistic term involving α. It is thus plausible that the neural network, with its approximation properties, could adapt to changes in the mechanistic parameter α.

In this simple scenario, we aim to demonstrate qualitatively that the projection of the mechanistic parameter onto the null subspace of H_χ effectively embodies a compensation between the mechanistic parameter and the neural network. To achieve this, we analyze the model behavior when the parameters of the trained HNODE model are perturbed along the direction identified by the projection (Fig. 4). The analysis reveals that, despite the sensitivity of the model dynamics to changes solely in the parameter α, moving along the projection allows a significant alteration of the value of α without impacting the overall model dynamics.

Such compensation hinders the identifiability of α, and our workflow terminates at this point. We conclude by showing that, in this scenario, the regularization of the cost function proposed by Yin et al.[28] does not ensure an accurate estimation of α. Their proposed regularization involves minimizing the contribution of the neural network in the HNODE model. Mathematically, this approach consists of adding the following regularizer to the cost

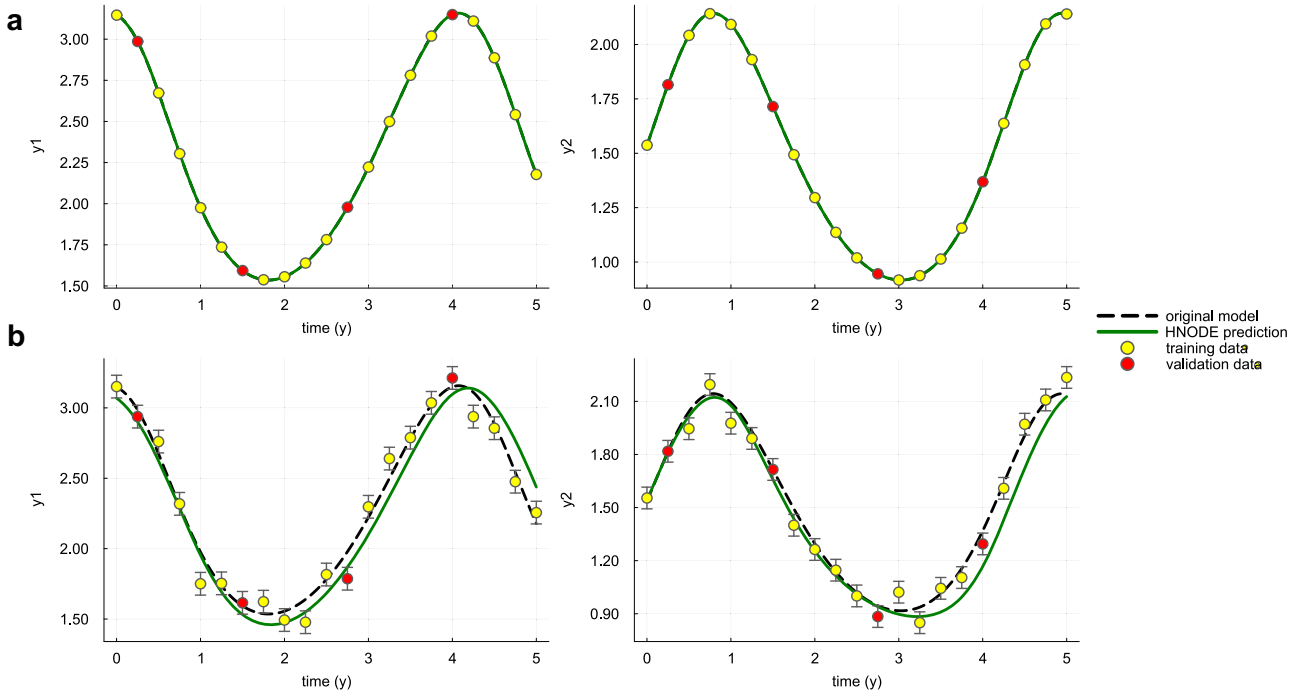

Fig. 2 | Dynamics predicted by the HNODE model in the Lotka-Volterra test case. The dynamics predicted by the HNODE model trained on $DS_{0.00}$ and $DS_{0.05}$ (shown in panels **a** and **b** respectively) are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Table 1 | Lotka-Volterra, estimated values of the prey birth rate α

Target	Dataset	Estimate
1.3	$DS_{0.00}$	0.804
	$DS_{0.05}$	0.744

function:

$$R(\boldsymbol{\theta}^{NN}) = \sum_{t \in X} \frac{||NN([y_1(t), y_2(t)], \boldsymbol{\theta}^{NN})||^2,}{||(\frac{dy_1}{dt}, \frac{dy_2}{dt})||^2}$$

where X is a grid of time points on the integration interval. To demonstrate that in this case this approach would not be effective, we sampled 50 values for α within the interval [0.013, 3] and trained the HNODE model keeping α fixed. This resulted in obtaining a neural network parameterization $\boldsymbol{\theta}_{\alpha}^{NN}$ for each α, yielding a profile for the regularizer:

$$R : \alpha \rightarrow R(\boldsymbol{\theta}_{\alpha}^{NN}).$$

For $\alpha \leq 2$, the value of α is not correlated with the ability of the HNODE model to fit the data (Supplementary Fig. 1), suggesting that the compensation holds in this region. If the inclusion of the regularizer enables the accurate estimation of α, the ground truth value $\overline{\alpha}$ should correspond to the minimum of the regularizer profile R. However, our analysis indicates that this is not the case (Fig. 5). Therefore, in this scenario, using the regularizer would result in a biased estimation of the parameter.

Cell apoptosis test case

The second test case is based on a model for cell death in apoptosis[29], which constitutes a core sub-network within the signal transduction cascade that regulates programmed cell death. This model is known for the structural and practical non-identifiability of part of its parameters[6]. The equations

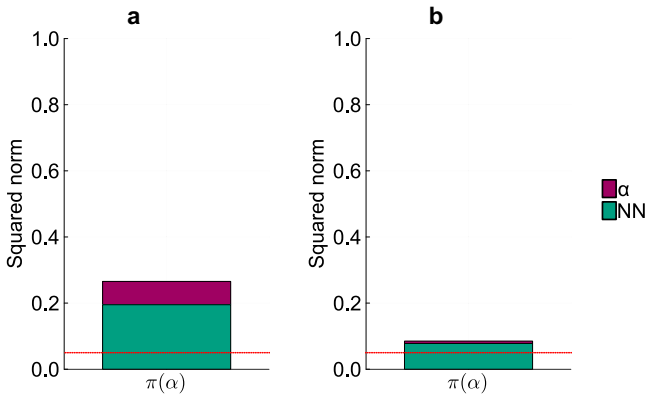

Fig. 3 | Identifiability analysis of α in the Lotka-Volterra test case. Squared norm of the projections of the parameter α onto the null subspace of H_χ for the models trained on $DS_{0.00}$ and $DS_{0.05}$ (panels **a** and **b** respectively). The total height of the bar corresponds to the squared 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.

defining the model are as follows:

$$\begin{cases} \frac{dy_1(t)}{dt} = -k_1 y_4 y_1 + k_{d1} y_5 \\ \frac{dy_2(t)}{dt} = k_{d2} y_5 - k_3 y_2 y_3 + k_{d3} y_6 + k_{d4} y_6 \\ \frac{dy_3(t)}{dt} = -k_3 y_2 y_3 + k_{d3} y_6 \\ \frac{dy_4(t)}{dt} = k_{d4} y_6 - k_1 y_4 y_1 + k_{d1} y_5 - k_5 y_7 y_4 + k_{d5} y_8 + k_{d2} y_5 \\ \frac{dy_5(t)}{dt} = -k_{d2} y_5 + k_1 y_4 y_1 - k_{d1} y_5 \\ \frac{dy_6(t)}{dt} = -k_{d4} y_6 + k_3 y_2 y_3 - k_{d3} y_6 \\ \frac{dy_7(t)}{dt} = -k_5 y_7 y_4 + k_{d5} y_8 + k_{d6} y_8 \\ \frac{dy_8(t)}{dt} = k_5 y_7 y_4 - k_{d5} y_8 - k_{d6} y_8. \end{cases} \quad (14)$$

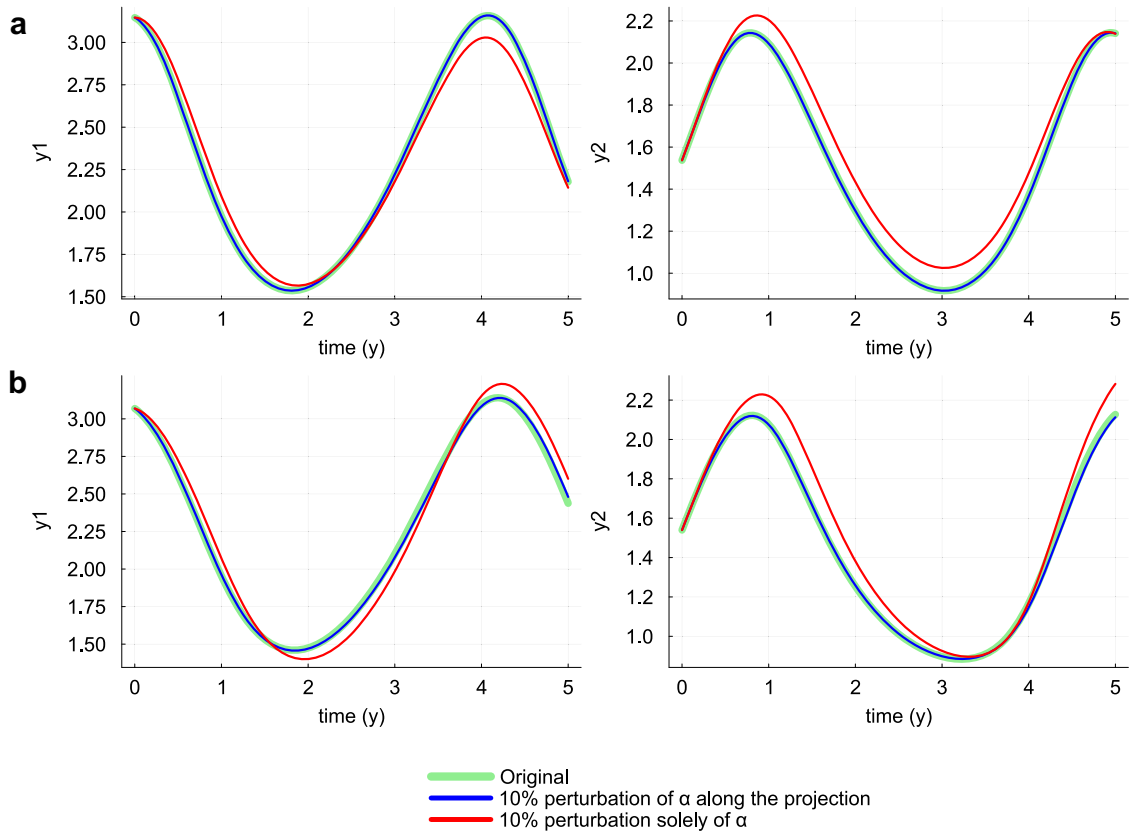

Fig. 4 | Qualitative inspection of the compensation in the Lotka-Volterra test case. The dynamics predicted by the HNODE models trained on $DS_{0.00}$ and $DS_{0.05}$ (panels **a** and **b** respectively) are compared with the dynamics obtained perturbing the α value by 10% along the projection onto the null subspace of H_χ, and with the dynamics obtained perturbing solely α by 10%.

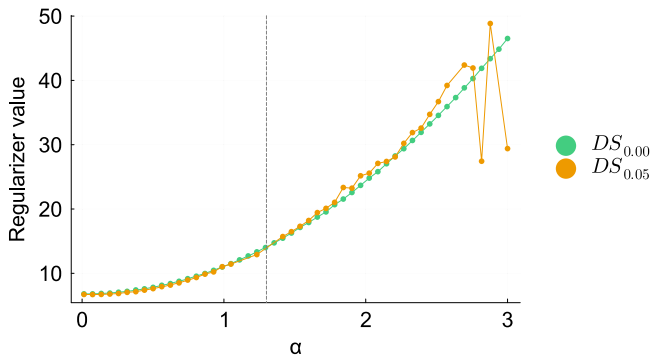

Fig. 5 | Profile of the regularizer in the Lotka-Volterra test case. The profiles of $R(\alpha)$ for the models trained on $DS_{0.00}$ and $DS_{0.05}$ are plotted in the interval [0, 3]. The vertical dashed line denotes the ground truth value of α.

In this context, we assume a lack of mechanistic knowledge about the entire equation related to y_4. We selected this variable to maximize the challenge, as the corresponding equation contains the highest number of non-linear terms in the system. Our objective is to estimate all the parameters of the model. We assume knowledge of the variables directly influencing the dynamics of y_4; thus, we model the unknown derivative of y_4 with the neural network:

$$\frac{dy_4(t)}{dt} = NN([y_1, y_4, y_5, y_6, y_7, y_8], \boldsymbol{\theta}^{NN}).$$

As in Lotka-Volterra, we assume a search space for the mechanistic parameters ranging from 10^{-2} to 10^{2} times the ground truth value for each parameter. These ground truth values, derived from the values used by Aldridge et al.[29], are specified in Supplementary Note 6, together with an analysis of the stiffness of the ODE system.

In the first scenario, we assume that all the system variables are observable. The observations are generated by numerically integrating Eq. (14) from 0 to 16 hours, with the parameters and initial conditions outlined in Supplementary Note 6. The trajectory is sampled every 0.8 hours, resulting in 21 observed time points. Similar to the previous test case, we consider a dataset $DS_{0.00}$ without noise, and a dataset $DS_{0.05}$ in which each time series is perturbed with zero-mean Gaussian noise with a standard deviation equal to 5% of its min-max variation. In this test case, after observing the integration in the first trials of the hyperparameter tuning, we manually switched to stiff configurations. These configurations include utilizing the *TRBDF2*[51] implicit Runge-Kutta ODE solver and the *QuadratureAdjoint*[35] method for adjoint sensitivity computation. Additionally, the neural network architecture search spaces are expanded to include a deeper neural network (up to 6 hidden layers), with *gelu* as the activation function. The hyperparameters selected by the pipeline are reported in Supplementary Table 8, and the dynamics for y_4, y_5, y_6, and y_7 predicted by the trained HNODE models are presented in Fig. 6 (Supplementary Figs. 3 and 4 for the dynamics of all the variables). The trained HNODE model effectively fits the data.

With the trained HNODE models, we proceed to the identifiability analysis and the CI estimations of the mechanistic parameters of the model (Table 2). Two of the model parameters, k_{d2} and k_{d4}, are classified as identifiable, and this is empirically confirmed by the accuracy of their

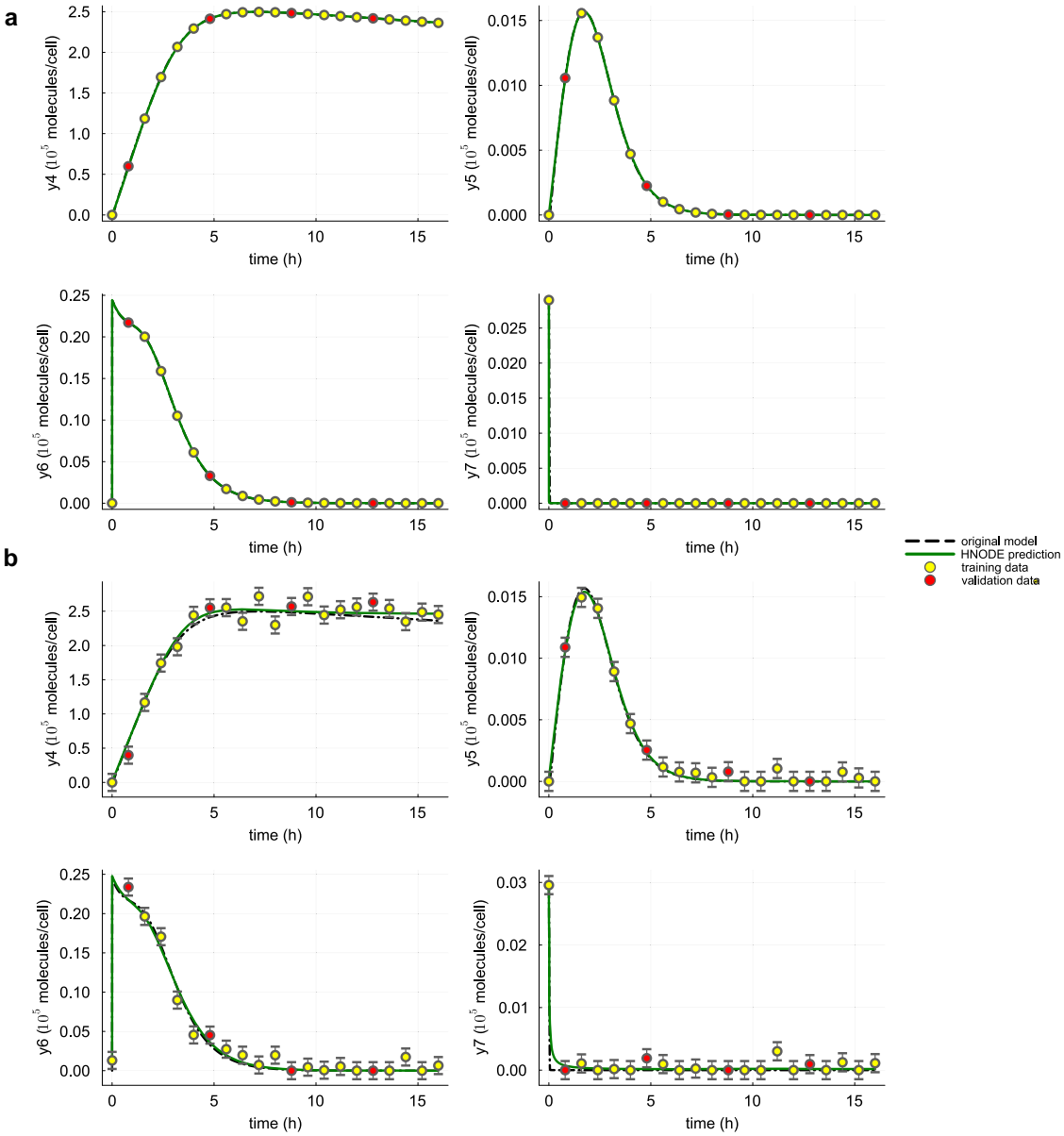

Fig. 6 | Dynamics predicted by the HNODE model in the cell apoptosis test case, first scenario. The dynamics of y_4, y_5, y_6, and y_7 predicted by the HNODE model trained on $DS_{0.00}$ and $DS_{0.05}$ (shown in panels **a** and **b** respectively) are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

estimates (maximum relative error of 0.38% when estimated on the dataset without noise and 5.08% on the dataset with noise).

To understand why the other mechanistic parameters are classified as not identifiable, we analyze the components of the projections on the null subspace of H_χ. The analysis reveals the absence of neural network parameter components in both the HNODE models (Fig. 7 for the model trained on $DS_{0.00}$, Supplementary Fig. 2 for analogous results obtained with the model trained on $DS_{0.05}$). This implies that the parameter non-identifiability is not caused by compensations due to the neural network. Notably, the projections of parameters k_1 and k_{d1} show almost equal contributions from both the parameters, and the same holds for the parameters k_3 and k_{d3}. This suggests the existence of local compensations between k_1 and k_{d1}, as well as between k_3 and k_{d3}.

To assess the existence of such compensations and to evaluate the performance of our approach with a larger number of identifiable parameters, we fix k_{d1} and k_3 to their ground truth values ($36.0\,\mathrm{h}^{-1}$ and $24.48\,\mathrm{cell}\cdot\mathrm{h}^{-1}\cdot 10^{-5}\cdot\mathrm{molecules}^{-1}$ respectively[29]) and execute the pipeline. The selected hyperparameters are listed in Supplementary Table 9; the dynamics predicted by the HNODE models effectively fit the experimental data (Supplementary Figs. 6 and 7). The results show that in this condition the identifiability of k_{d1} and k_{d3} is restored, allowing for the estimation of four model parameters with low relative error (Table 3).

To mimic more realistic conditions in the second scenario, keeping k_{d1} and k_3 fixed to their ground truth values, we suppose that only the variables y_5 and y_6 are observable. We selected these variables because, within the mechanistic model, they represent the minimal subset that maximizes the number of identifiable parameters (Supplementary Note 16). The in silico observations are generated as in the first scenario. However, since only 2 out of 8 model variables are observable, trajectories are sampled every 0.4 hour to maintain a comparable amount of information in the training set, resulting in 42 time points. The hyperparameters tuned in the pipeline are listed in Supplementary Table 10, and the dynamics of the observable variables predicted by the HNODE models are shown in Supplementary Figs. 9 and 10.

Table 2 | Cell apoptosis, first scenario - estimated values of mechanistic parameters

Par.	Target	Dataset	Estimate	Rel. err.	Id.	CI
k_1	$9.61 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$2.76 \cdot 10^{1}$ $5.28 \cdot 10^{1}$	2768.72% 5397.02%		
k_{d1}	$3.60 \cdot 10^{1}$	$DS_{0.00}$ $DS_{0.05}$	$1.82 \cdot 10^{3}$ $3.45 \cdot 10^{3}$	4963.72% 9496.00%		
k_{d2}	$2.88 \cdot 10^{1}$	$DS_{0.00}$ $DS_{0.05}$	$2.87 \cdot 10^{1}$ $2.73 \cdot 10^{1}$	0.38% 5.08%	✓ ✓	$(2.84 \cdot 10^{1}, 2.90 \cdot 10^{1})$ $(2.43 \cdot 10^{1}, 3.04 \cdot 10^{1})$
k_3	$2.45 \cdot 10^{1}$	$DS_{0.00}$ $DS_{0.05}$	$4.72 \cdot 10^{2}$ $1.42 \cdot 10^{3}$	1827.91% 5693.37%		
k_{d3}	$1.80 \cdot 10^{2}$	$DS_{0.00}$ $DS_{0.05}$	$3.53 \cdot 10^{3}$ $1.22 \cdot 10^{4}$	1862.51% 6675.29%		
k_{d4}	3.60	$DS_{0.00}$ $DS_{0.05}$	3.59 3.69	0.25% 2.61%	✓ ✓	(3.55, 3.63) (3.30, 4.08)
k_5	$2.52 \cdot 10^{4}$	$DS_{0.00}$ $DS_{0.05}$	$1.17 \cdot 10^{6}$ $4.36 \cdot 10^{3}$	4562.35% 82.68%		
k_{d5}	$6.01 \cdot 10^{-2}$	$DS_{0.00}$ $DS_{0.05}$	$6.52 \cdot 10^{-1}$ 6.01	984.49% 9900.00%		
k_{d6}	$6.01 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$4.56 \cdot 10^{1}$ $6.01 \cdot 10^{1}$	7489.27% 9900.00%		

The symbol ✓ denotes the identifiable parameters according to our method. CIs for the model trained on $DS_{0.00}$ are computed assuming an uncertainty equal to 0.5% of the min-max of each time series.

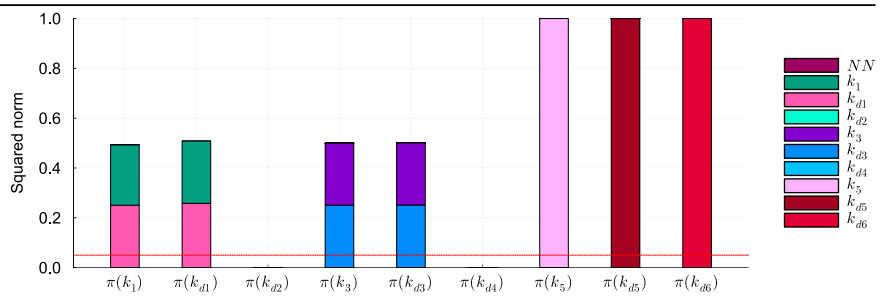

Fig. 7 | Identifiability analysis of mechanistic parameters in the cell apoptosis test case, first scenario ($DS_{0.00}$). Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the model trained on $DS_{0.00}$. The total height of the bar corresponds to the squared 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.

Table 3 | Cell apoptosis, first scenario, assuming k_{d1} and k_3 fixed to their ground truth values - estimated values of mechanistic parameters

Par.	Target	Dataset	Estimate	Rel. err.	Id.	CI
k_1	$9.61 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$9.60 \cdot 10^{-1}$ $9.75 \cdot 10^{-1}$	0.14% 1.42%	✓ ✓	$(9.47 \cdot 10^{-1}, 9.73 \cdot 10^{-1})$ $(8.45 \cdot 10^{-1}, 1.10)$
k_{d2}	$2.88 \cdot 10^{1}$	$DS_{0.00}$ $DS_{0.05}$	$2.87 \cdot 10^{1}$ $2.72 \cdot 10^{1}$	0.29% 5.53%	✓ ✓	$(2.84 \cdot 10^{1}, 2.90 \cdot 10^{1})$ $(2.45 \cdot 10^{1}, 2.99 \cdot 10^{1})$
k_{d3}	$1.80 \cdot 10^{2}$	$DS_{0.00}$ $DS_{0.05}$	$1.80 \cdot 10^{2}$ $2.03 \cdot 10^{2}$	0.04% 12.88%	✓ ✓	$(1.78 \cdot 10^{2}, 1.83 \cdot 10^{2})$ $(1.77 \cdot 10^{2}, 2.29 \cdot 10^{2})$
k_{d4}	3.60	$DS_{0.00}$ $DS_{0.05}$	3.60 3.67	0.08% 2.05%	✓ ✓	(3.56, 3.63) (3.30, 4.05)
k_5	$2.52 \cdot 10^{4}$	$DS_{0.00}$ $DS_{0.05}$	$1.46 \cdot 10^{5}$ $3.98 \cdot 10^{5}$	479.74% 1481.24%		
k_{d5}	$6.01 \cdot 10^{-2}$	$DS_{0.00}$ $DS_{0.05}$	$9.64 \cdot 10^{-1}$ $6.01 \cdot 10^{-4}$	1503.46% 99.00%		
k_{d6}	$6.01 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$2.91 \cdot 10^{1}$ $6.01 \cdot 10^{-3}$	4742.70% 99.00%		

The symbol ✓ denotes the identifiable parameters according to our method. CIs for the model trained on $DS_{0.00}$ are computed assuming an uncertainty equal to 0.5% of the min-max of each time series.

In this scenario k_{d2}, k_{d3}, and k_{d4} are still identifiable (Table 4), and can be accurately estimated (maximum relative error of 0.17% when estimated on the dataset without noise, 11.64% on the dataset with noise), whereas k_1 is not. The analysis of the components of the projections of the mechanistic parameters onto the null space of H_χ suggests that the non-identifiability of k_1 in this scenario is caused by a compensation between k_1 and the neural network (Fig. 8 and Supplementary Fig. 8).

Yeast glycolysis model

The last test case is based on a model of oscillations in yeast glycolysis[30], which has been frequently employed as a benchmark for inference in

Table 4 | Cell apoptosis, second scenario - estimated values of mechanistic parameters

Par.	Target	Dataset	Estimate	Rel. err.	Id.	CI
k_1	$9.61 \cdot 10^{-1}$	$DS_{0.00}$	1.32	37.32%		
		$DS_{0.05}$	1.18	23.06%		
k_{d2}	$2.88 \cdot 10^{1}$	$DS_{0.00}$	$2.88 \cdot 10^{1}$	0.16%	✓	$(2.45 \cdot 10^{1}, 3.32 \cdot 10^{1})$
		$DS_{0.05}$	$2.95 \cdot 10^{1}$	2.60%	✓	$(-6.86 \cdot 10^{1}, 1.28 \cdot 10^{2})$
k_{d3}	$1.80 \cdot 10^{2}$	$DS_{0.00}$	$1.80 \cdot 10^{2}$	0.17%	✓	$(1.76 \cdot 10^{2}, 1.85 \cdot 10^{2})$
		$DS_{0.05}$	$2.01 \cdot 10^{2}$	11.64%	✓	$(1.60 \cdot 10^{2}, 2.42 \cdot 10^{2})$
k_{d4}	3.60	$DS_{0.00}$	3.60	0.03%	✓	(3.56, 3.64)
		$DS_{0.05}$	3.90	8.22%	✓	(3.42, 4.37)
k_5	$2.52 \cdot 10^{4}$	$DS_{0.00}$	$1.79 \cdot 10^{6}$	7012.35%		
		$DS_{0.05}$	$2.62 \cdot 10^{3}$	89.59%		
k_{d5}	$6.01 \cdot 10^{-2}$	$DS_{0.00}$	$6.01 \cdot 10^{-4}$	99.00%		
		$DS_{0.05}$	4.93	8095.41%		
k_{d6}	$6.01 \cdot 10^{-1}$	$DS_{0.00}$	$6.01 \cdot 10^{-3}$	99.00%		
		$DS_{0.05}$	$6.01 \cdot 10^{1}$	9900.00%		

The symbol ✓ denotes the identifiable parameters according to our method. CIs for the model trained on $DS_{0.00}$ are computed assuming an uncertainty equal to 0.5% of the min-max of each time series.

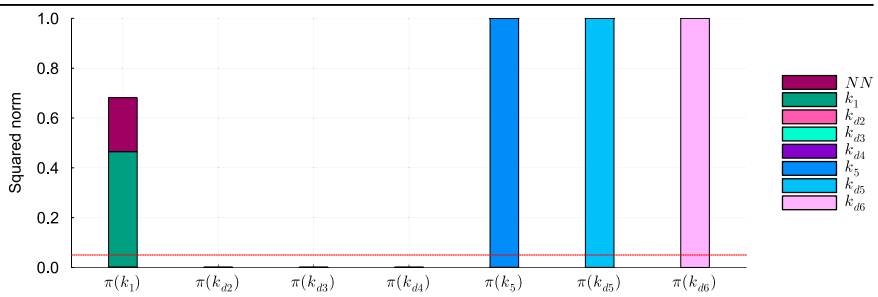

Fig. 8 | Identifiability analysis of mechanistic parameters in the cell apoptosis test case, second scenario ($DS_{0.00}$). Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the model trained on $DS_{0.00}$. The total height of the bar corresponds to the 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.

computational systems biology[6,52]. The model is described by the following system of ODE:

$$\begin{cases} \frac{dy_1(t)}{dt} = J_0 - \frac{k_1 y_1 y_6}{1+(y_6/K_1)^q} \\ \frac{dy_2(t)}{dt} = 2\frac{k_1 y_1 y_6}{1+(y_6/K_1)^q} - k_2 y_2 (N - y_5) - k_6 y_2 y_5 \\ \frac{dy_3(t)}{dt} = k_2 y_2 (N - y_5) - k_3 y_3 (A - y_6) \\ \frac{dy_4(t)}{dt} = k_3 y_3 (A - y_6) - k_4 y_4 y_5 - \kappa (y_4 - y_7) \\ \frac{dy_5(t)}{dt} = k_2 y_2 (N - y_5) - k_4 y_4 y_5 - k_6 y_2 y_5 \\ \frac{dy_6(t)}{dt} = -2\frac{k_1 y_1 y_6}{1+(y_6/K_1)^q} + 2k_3 y_3 (A - y_6) - k_5 y_6 \\ \frac{dy_7(t)}{dt} = \psi \kappa (y_4 - y_7) - k y_7. \end{cases} \quad (15)$$

In this test case, we assume a complete absence of mechanistic knowledge concerning y_1. This variable has been chosen to maximize the challenge, as its nonlinear oscillatory dynamics have proven to be particularly difficult for automatic identification of dynamical systems[53]. Our objective is to estimate all the parameters of the systems (except J_0, which appears only in the derivative of y_1, and therefore is not present in the HNODE model). As in the cell apoptosis case, we assume a knowledge of the variables directly influencing the dynamics of y_1, thus the equation of y_1 will be approximated by a neural network depending solely on y_1 and y_6. Given the greater number of system parameters compared to the first two test cases, we assume a search space ranging from 10^{-1} to 10^{1} times the ground truth value for each mechanistic parameter. These ground truth values, derived from the values employed by Ruoff et al.[30], are specified in Section Supplementary Note 18.

In the first scenario, we assume that all system variables are observable. The observations are generated by numerically integrating Eq. (15) from 0 to 6 min, using the parameters and initial conditions specified in Supplementary Note 18. The trajectory is sampled every 0.8 min (21 observation time points). We consider observation datasets with and without noise, labeled as $DS_{0.00}$ and $DS_{0.05}$ respectively. The noise dataset ($DS_{0.05}$) is generated as in the two previous test cases. We execute the pipeline, switching to stiff configurations (as described for the cell apoptosis case) after observing the first trials of the initial step. The tuned hyperparameters are listed in Supplementary Table 17. The trained HNODE models fit the observations (Fig. 9 for the dynamics of the variables y_1 and y_2, Supplementary Figs. 11 and 12 for all the model variables).

In the noiseless case, the pipeline assesses the identifiability of all the parameters, whereas, in the presence of noise, k_1 is classified as not identifiable. By comparing the norm of the projections of the mechanistic parameters onto the null space of H_χ (Supplementary Fig. 13), we can notice that also when estimated on $DS_{0.00}$, the norm of the k_1 projection falls just below our identifiability threshold. In both cases, the projections of k_1 are composed by k_1 itself and by the neural network parameters, thus embodying a potential compensation between k_1 and the neural network. This is plausible since the parameter k_1 in the HNODE model directly multiplies the variable y_1 parameterized by the neural network.

The identifiable parameters are estimated with a maximum relative error of 1.94% (Table 5) on $DS_{0.00}$, with the ground truth value of the parameter always falling within the estimated CIs. When estimated on $DS_{0.05}$, the maximum relative error of the identifiable parameter estimates is 52.60%, with the CIs containing the relative ground truth value for 11 out of 12 parameters. Given the challenge of discerning whether the relative error of parameter estimates originates from the lack of mechanistic knowledge or from the noise in the observation dataset, we attempted to estimate all parameter values using the fully mechanistic model on dataset $DS_{0.05}$. The outcomes indicate that the mean and maximum relative errors obtained with our pipeline are consistent with what is obtained with the fully mechanistic model (Supplementary Note 24).

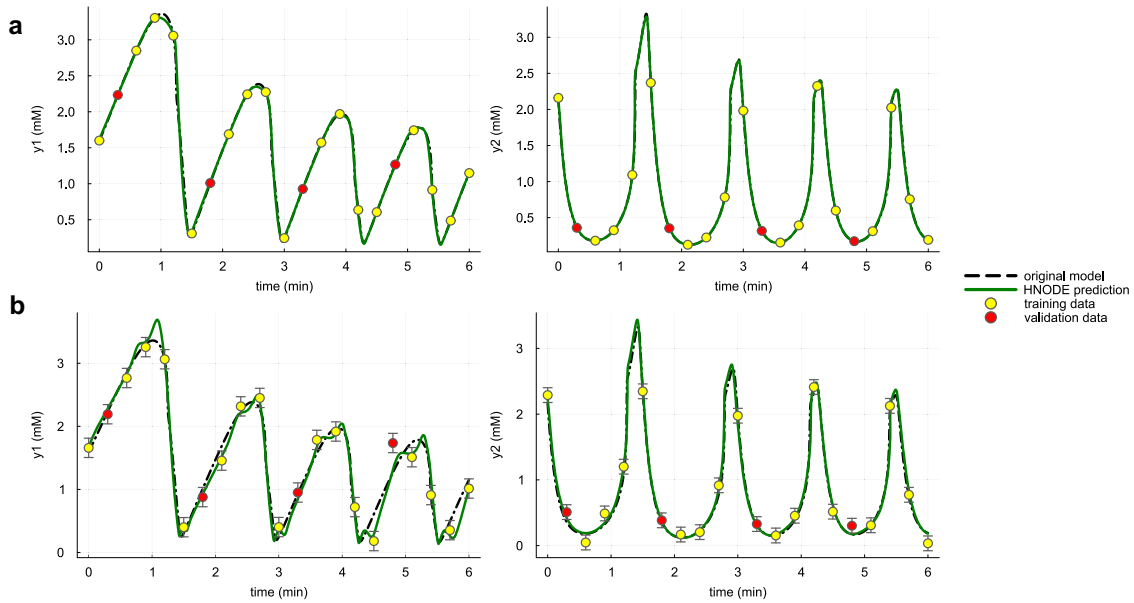

Fig. 9 | Dynamics predicted by the HNODE model in the yeast glycolysis test case, first scenario. The dynamics of y_1 and y_2 predicted by the HNODE model trained on $DS_{0.00}$ and $DS_{0.05}$ (shown in panels **a** and **b** respectively) are compared with the original model. The points represent the observations of the system, divided into training and validation sets.

Table 5 | Yeast glycolysis, first scenario: estimated values of mechanistic parameters

Par.	Target	Dataset	Estimate	Rel. err.	Id.	CI
k_1	$1.00 \cdot 10^2$	$DS_{0.00}$ $DS_{0.05}$	$1.02 \cdot 10^2$ $1.25 \cdot 10^2$	1.94% 25.43%	✓	$(8.46 \cdot 10^1, 1.19 \cdot 10^2)$
K_1	$5.20 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$5.25 \cdot 10^{-1}$ $4.65 \cdot 10^{-1}$	0.95% 10.61%	✓ ✓	$(4.86 \cdot 10^{-1}, 5.67 \cdot 10^{-1})$ $(1.58 \cdot 10^{-1}, 1.37)$
q	4.00	$DS_{0.00}$ $DS_{0.05}$	4.04 3.79	1.02% 5.33%	✓ ✓	(3.87, 4.21) (2.16, 5.42)
k_2	6.00	$DS_{0.00}$ $DS_{0.05}$	5.96 9.16	0.68% 52.60%	✓ ✓	(4.68, 7.24) $(9.16 \cdot 10^{-1}, 1.99 \cdot 10^1)$
N	1.00	$DS_{0.00}$ $DS_{0.05}$	1.00 $7.04 \cdot 10^{-1}$	0.13% 29.61%	✓ ✓	$(8.20 \cdot 10^{-1}, 1.18)$ $(1.0 \cdot 10^{-1}, 1.34)$
k_6	$1.20 \cdot 10^1$	$DS_{0.00}$ $DS_{0.05}$	$1.19 \cdot 10^1$ 9.84	0.57% 17.97%	✓ ✓	$(1.16 \cdot 10^1, 1.23 \cdot 10^1)$ $(7.11, 1.26 \cdot 10^1)$
k_3	$1.60 \cdot 10^1$	$DS_{0.00}$ $DS_{0.05}$	$1.59 \cdot 10^1$ $1.38 \cdot 10^1$	0.59% 13.76%	✓ ✓	$(1.56 \cdot 10^1, 1.62 \cdot 10^1)$ $(1.12 \cdot 10^1, 1.64 \cdot 10^1)$
A	4.00	$DS_{0.00}$ $DS_{0.05}$	4.00 4.43	0.02% 10.69%	✓ ✓	(3.98, 4.02) (4.21, 4.65)
k_4	$1.00 \cdot 10^2$	$DS_{0.00}$ $DS_{0.05}$	$9.94 \cdot 10^1$ $1.02 \cdot 10^2$	0.64% 1.75%	✓ ✓	$(9.73 \cdot 10^1, 1.01 \cdot 10^2)$ $(8.50 \cdot 10^1, 1.18 \cdot 10^2)$
κ	$1.30 \cdot 10^1$	$DS_{0.00}$ $DS_{0.05}$	$1.29 \cdot 10^1$ $1.15 \cdot 10^1$	0.49% 11.84%	✓ ✓	$(1.26 \cdot 10^1, 1.33 \cdot 10^1)$ $(8.98, 1.39 \cdot 10^1)$
k_5	1.28	$DS_{0.00}$ $DS_{0.05}$	1.28 1.31	0.03% 2.40%	✓ ✓	(1.26, 1.30) (1.17, 1.45)
ψ	$1.00 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$1.00 \cdot 10^{-1}$ $1.31 \cdot 10^{-1}$	0.09% 30.55%	✓ ✓	$(9.55 \cdot 10^{-2}, 1.05 \cdot 10^{-1})$ $(8.13 \cdot 10^{-2}, 1.80 \cdot 10^{-1})$
k	1.80	$DS_{0.00}$ $DS_{0.05}$	1.80 2.08	0.02% 15.78%	✓ ✓	(1.76, 1.84) (1.68, 2.48)

The symbol ✓ denotes the identifiable parameters according to our method. CIs for the model trained on $DS_{0.00}$ are computed assuming an uncertainty equal to 0.5% of the min-max of each time series.

In the second scenario, we assume that only the variables y_5 and y_6 are observable. We chose these variables because, in the original model, all parameters are identifiable when y_5 and y_6 are observed[6]. Following a consistent approach with the previous test case, the in silico observations are generated as in the first scenario, with the sampling frequency halved to 0.4 minutes, resulting in 42 observed time points. The pipeline is run analogously to the first scenario. The tuned hyperparameters are listed in Supplementary Table 18, and the dynamics predicted by the HNODE model (Supplementary Figs. 14 and 15) fit the experimental data.

Similar to the first scenario, the results regarding parameter identifiability vary when the parameters are estimated on $DS_{0.00}$ and

$DS_{0.05}$ (Table 6). When the parameters are estimated using the dataset without noise, the pipeline classifies six parameters as identifiable. However, when estimated using $DS_{0.05}$, the parameters K_1 and k_4, classified as identifiable on $DS_{0.00}$, are classified as non-identifiable. By analyzing the projections of the parameters onto the null subspace of H_χ (Fig. 10), we can notice that K_1 is classified as non-identifiable on $DS_{0.05}$, but its projection falls just above the threshold. Conversely, parameter k_4 is classified as identifiable on $DS_{0.00}$, yet its projection falls just below the threshold. Interestingly, on both datasets, the neural network parameters significantly contribute to the projection of nearly all non-identifiable mechanistic parameters, except for k_2 and N. This observation is consistent with our findings on the cell apoptosis test case: when the unknown variable y_1 is not observable, the neural network can compensate for changes in mechanistic parameters, also not directly linked to y_1 in the HNODE model.

The identifiable parameters are estimated with a maximum relative error of 9.26% on $DS_{0.00}$. On the dataset with error $DS_{0.05}$ the identifiable parameters are estimated with a maximum relative error of 11.05%. In both cases, the ground truth value consistently falls within the confidence interval estimated.

Table 6 | Yeast glycolysis, second scenario: estimated values of mechanistic parameters

Par.	Target	Dataset	Estimate	Rel. err.	Id.	CI
k_1	$1.00 \cdot 10^2$	$DS_{0.00}$ $DS_{0.05}$	$5.42 \cdot 10^1$ $7.02 \cdot 10^1$	45.77% 29.84%		
K_1	$5.20 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$5.20 \cdot 10^{-1}$ $7.91 \cdot 10^{-1}$	0.04% 52.12%	✓	$(2.50 \cdot 10^{-1}, 1.08)$
q	4.00	$DS_{0.00}$ $DS_{0.05}$	3.63 4.44	9.26% 11.05%	✓ ✓	(1.49, 5.76) $(9.45 \cdot 10^{-1}, 7.94)$
k_2	6.00	$DS_{0.00}$ $DS_{0.05}$	$1.04 \cdot 10^1$ 2.52	74.00% 57.96%		
N	1.00	$DS_{0.00}$ $DS_{0.05}$	$6.61 \cdot 10^{-1}$ 2.26	33.94% 126.02%		
k_6	$1.20 \cdot 10^1$	$DS_{0.00}$ $DS_{0.05}$	$1.17 \cdot 10^1$ $1.07 \cdot 10^1$	2.49% 10.74%	✓ ✓	$(9.58, 1.38 \cdot 10^1)$ $(4.71, 1.67 \cdot 10^1)$
k_3	$1.60 \cdot 10^1$	$DS_{0.00}$ $DS_{0.05}$	$1.63 \cdot 10^1$ $5.98 \cdot 10^1$	1.61% 273.76%	✓	$(1.29 \cdot 10^1, 1.96 \cdot 10^1)$
A	4.00	$DS_{0.00}$ $DS_{0.05}$	4.00 3.79	0.00% 5.27%	✓ ✓	(3.90, 4.10) (3.39, 4.19)
k_4	$1.00 \cdot 10^2$	$DS_{0.00}$ $DS_{0.05}$	$1.06 \cdot 10^2$ $1.94 \cdot 10^2$	6.26% 93.70%		
κ	$1.30 \cdot 10^1$	$DS_{0.00}$ $DS_{0.05}$	$1.19 \cdot 10^1$ $4.35 \cdot 10^1$	8.83% 234.70%		
k_5	1.28	$DS_{0.00}$ $DS_{0.05}$	1.33 1.35	3.74% 5.48%	✓ ✓	(1.25, 1.41) (1.05, 1.65)
ψ	$1.00 \cdot 10^{-1}$	$DS_{0.00}$ $DS_{0.05}$	$1.17 \cdot 10^{-1}$ $3.97 \cdot 10^{-1}$	16.89% 297.04%		
k	1.80	$DS_{0.00}$ $DS_{0.05}$	2.21 6.65	23.02% 269.66%		

The symbol ✓ denotes the identifiable parameters according to our method. CIs for the model trained on $DS_{0.00}$ are computed assuming an uncertainty equal to 0.5% of the min-max of each time series.

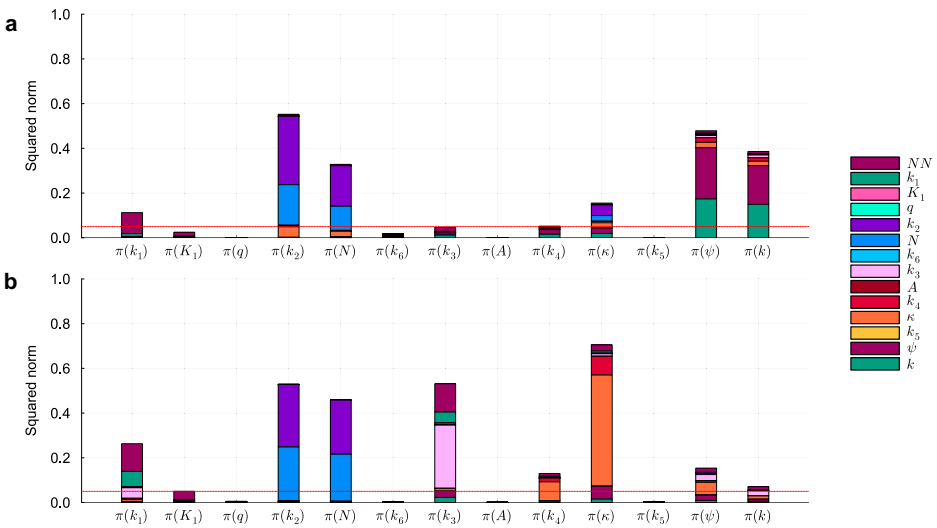

Fig. 10 | Identifiability analysis of mechanistic parameters in the yeast glycolysis test case, second scenario. Squared norm of the projections of the mechanistic parameters onto the null subspace of H_χ for the model trained on $DS_{0.00}$ and $DS_{0.05}$ (panels **a** and **b** respectively). The 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.

Discussion

We have introduced a pipeline for estimating mechanistic parameters and discussing their local identifiability within an incompletely specified mechanistic model. This workflow leverages the HNODE framework to embed the original model into a larger model, in which the unknown portions of the system are described by neural networks. The primary novelties of our work lie in our approach to conducting a global exploration of the mechanistic parameter search space and our method to evaluate the parameter identifiability. First, we include the initial values of mechanistic parameters into the hyperparameters to be tuned. This involves combining Bayesian Optimization with gradient-based search, with the goal of globally exploring the parameter search space. Second, to assess the local identifiability at-a-point of the parameters, we extend a classical approach for identifiability analysis in mechanistic models. Notably, identifiability analysis for HNODE models has not been previously investigated in the literature, which has predominantly focused on ensuring the identifiability of the mechanistic component through regularization of the training cost function[26,28].

The primary limitations of our work include: (1) the assumption of having access to initial conditions for all state variables within systems, even under partial observability; (2) the computational cost associated with the hyperparameter tuning phase; (3) the intrinsic local nature of identifiability analysis; (4) the arbitrary selection of hyperparameters, ϵ and δ, for identifiability analysis; and (5) the methodology employed for estimating confidence intervals. All these limitations are discussed in the following paragraphs.

We assume knowledge of the initial states of all system variables, possibly with noise, even when only a subset of them is observable. In real-world scenarios, this assumption may not always be valid. While in certain cases, scientists have concrete physical insights into the system initial state, in others, initial concentrations need to be estimated alongside the parameters. Notably, this limitation is shared by other approaches to system identification[6].

The initial phase of hyperparameter tuning and exploration of the mechanistic search space is the most computationally expensive step within the pipeline (for a discussion of the computational times required to run the pipeline across the different test cases, see Supplementary Note 26). We employed the TPE algorithm, and this algorithm has proven effective even in high-dimensional parameter search spaces, as evidenced by the Yeast Glycolysis test case, featuring 13 mechanistic parameters and 6 hyperparameters. However, employing the TPE algorithm entails sequentially repeating the training of the candidate HNODE model, each time with a limited number of epochs. It is important to note that the problem of hyperparameter tuning is a long-standing problem and different approaches have been proposed for this goal[54]. These approaches range from random and grid search to Bayesian optimization and genetic algorithms: in this work, we did not conduct a comparative analysis of TPE performance against other hyperparameter tuning methods.

The local nature of our identifiability analysis implies that it might overlook parameter non-identifiability if distinct parameterizations leading to similar model simulations are isolated in the search space, as separate local minima of the cost function.

The hyperparameters ϵ and δ used for identifiability analysis, although related, have different meanings. The threshold ϵ intuitively distinguishes what we assume to be negligible changes in the HNODE model behavior from non-negligible changes. This hyperparameter is also present in existing Hessian-based methods for identifiability analysis of mechanistic models[42]. The threshold δ, introduced in our method, intuitively quantifies how much it is possible to perturb the mechanistic parameter with a negligible effect on the model simulations. δ has been introduced to overcome the limitation of the dominant parameter approach in the context of HNODE. Existing Hessian-based methods for identifiability analysis of mechanistic models[42] primarily categorize parameters as non-identifiable if they have the highest projection (in absolute value) onto eigenvectors associated with null eigenvalues. We decided against using this approach because we observed empirically that in HNODE models when compensation occurs between the neural network and mechanistic parameters, the dominant parameter in the null direction is typically a neural network parameter. Thus, considering only the dominant parameter, we risk overlooking the role of the mechanistic parameter. In our test cases, we employed $\epsilon = 10^{-5}$ and $\delta = 0.05$. The results of the first two test cases are quite robust to the choice of ϵ and δ (Supplementary Notes 4 and 17). However, the third test case is more complex: here changes in ϵ and δ would lead to different results for certain parameters (Supplementary Note 25).

The fifth limitation of our work is the CI estimation. The FIM-based approach we employed has several limitations[55]. By using it, we implicitly assume the unbiasedness and normality of our estimators[47]. Additionally, since our estimators are not linear, the CIs estimated are only lower bounds of the real CIs[47]. We opted for this method because of its computational advantages over other methods, such as likelihood profiling or bootstrapping-based methods[13,55], which would necessitate multiple trainings of the HNODE model.

We tested our pipeline in various in silico scenarios of increasing complexity, assuming a partial lack of mechanistic knowledge in three models acknowledged as benchmarks in computational biology. In each test case, we replace either a portion or an entire equation of the system with a neural network, creating a hybrid model in which the neural network influences the dynamics of all system variables (as detailed in Supplementary Note 27). These tests encompass various conditions, including different levels of noise in the training data and different assumptions regarding the observability of the system variables. Across all the examined scenarios, our pipeline consistently enabled the analysis of parameter identifiability and the accurate estimation of identifiable parameters.

Despite its limitations, the proposed workflow represents an initial step toward adapting traditional methods utilized in entirely mechanistic models to the HNODE modeling scenario. In future works, we aim to address the limitations of the pipeline. Firstly, we will consider different state-of-the-art approaches for hyperparameter tuning, such as Gaussian process[56], evolutionary[54], and genetic approaches[57], comparing their performances with TPE. Secondly, we plan to expand the number of test cases to derive a less arbitrary choice of the thresholds ϵ and δ, which could vary based on the number of parameters in the HNODE model. Thirdly, we intend to compare the FIM-based method for estimating CIs with other approaches that have demonstrated greater reliability in completely mechanistic models, as described by Kreutz et al.[13] and Joshi et al.[55].

Methods

Algorithmic setup

All the results presented for the test cases described in the Results Section have been obtained employing the following hyperparameters. The observation data point datasets are partitioned into training and validation sets with an 8:2 ratio. The first stage of the hyperparameter tuning consists in 500 iterations of TPE: in each trial, the HNODE model is trained for 500 epochs using Adam optimizer. In the second stage of the hyperparameter tuning (performed only on datasets with noise) the model is trained for 2000 epochs in each trial. The comprehensive training of the HNODE model encompasses 10000 epochs utilizing the Adam optimizer. Subsequent refinement involves the application of the L-BFGS algorithm until convergence, with a maximum of 5000 epochs (performed only on noiseless datasets). The optimized mechanistic parameters are clipped to the lower or upper bounds of the parameter search spaces when they exceed those limits. Each model is trained with 10 initializations of the neural network, with weights initialized using Glorot Uniform[58]. The training resulting in the lowest validation cost is then selected. The identifiability analysis is performed using the thresholds $\epsilon = 10^{-5}$ and $\delta = 0.05$; a discussion regarding these thresholds is provided in the Discussion Section.

Implementation

All the computations have been run on a Debian-based Linux cluster node, with two 24-core Xeon(R) CPU X5650 CPUs and 250 GB of RAM. The code

is implemented in Julia v1.9.1[59], using the SCiML environment[23], a software suite for modeling and simulation that also incorporates machine learning algorithms. The hyperparameter tuning has been performed with Optuna[60].

Data availability

No experimental data was used as part of this study. The in silico generated datasets are available at https://github.com/cosbi-research/HNODECB/tree/main/datasets.

Code availability

The scripts to reproduce the results are available at https://github.com/cosbi-research/HNODECB.

Received: 9 July 2024; Accepted: 21 October 2024;
Published online: 29 November 2024

References

1. Motta, S. & Pappalardo, F. Mathematical modeling of biological systems. *Brief. Bioinforma.* **14**, 411–422 (2013).
2. Mogilner, A., Wollman, R. & Marshall, W. F. Quantitative modeling in cell biology: what is it good for? *Dev. Cell* **11**, 279–287 (2006).
3. Gábor, A. & Banga, J. R. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Syst. Biol.* **9**, 1–25 (2015).
4. Baker, R. E., Pena, J.-M., Jayamohan, J. & Jérusalem, A. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol. Lett.* **14**, 20170660 (2018).
5. Cornish-Bowden, A.Fundamentals of enzyme kinetics (John Wiley & Sons, 2013).
6. Yazdani, A., Lu, L., Raissi, M. & Karniadakis, G. E. Systems biology informed deep learning for inferring parameters and hidden dynamics. *PLoS Comput. Biol.* **16**, e1007575 (2020).
7. Sun, J., Garibaldi, J. M. & Hodgman, C. Parameter estimation using metaheuristics in systems biology: a comprehensive review. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **9**, 185–202 (2011).
8. Mendes, P. & Kell, D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinforma. (Oxf., Engl.)* **14**, 869–883 (1998).
9. Reali, F., Priami, C. & Marchetti, L. Optimization algorithms for computational systems biology. *Front. Appl. Math. Stat.* **3**, 6 (2017).
10. Liepe, J. et al. A framework for parameter estimation and model selection from experimental data in systems biology using approximate bayesian computation. *Nat. Protoc.* **9**, 439–456 (2014).
11. Linden, N. J., Kramer, B. & Rangamani, P. Bayesian parameter estimation for dynamical models in systems biology. *PLoS Comput. Biol.* **18**, e1010651 (2022).
12. Meskin, N., Nounou, H., Nounou, M., Datta, A. & Dougherty, E. R. Parameter estimation of biological phenomena modeled by s-systems: an extended kalman filter approach. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, 4424–4429 (IEEE, 2011).
13. Kreutz, C., Raue, A., Kaschek, D. & Timmer, J. Profile likelihood in systems biology. *FEBS J.* **280**, 2564–2571 (2013).
14. Anstett-Collin, F., Denis-Vidal, L. & Millérioux, G. A priori identifiability: an overview on definitions and approaches. *Annu. Rev. Control* **50**, 139–149 (2020).
15. Lam, N. N., Docherty, P. D. & Murray, R. Practical identifiability of parametrised models: a review of benefits and limitations of various approaches. *Math. Comput. Simul.* **199**, 202–216 (2022).
16. Yeo, H. C. & Selvarajoo, K. Machine learning alternative to systems biology should not solely depend on data. *Brief. Bioinf.* **23**, bbac436 (2022).
17. Engelhardt, B., Frőhlich, H. & Kschischo, M. Learning (from) the errors of a systems biology model. *Sci. Rep.* **6**, 20772 (2016).
18. Zou, B. J., Levine, M. E., Zaharieva, D. P., Johari, R. & Fox, E. B. Hybrid square neural ode causal modeling. arXiv preprint arXiv:2402.17233 (2024).
19. Lanzieri, D., Lanusse, F. & Starck, J.-L. Hybrid physical-neural odes for fast n-body simulations. arXiv preprint arXiv:2207.05509 (2022).
20. Grigorian, G., George, S. V., Lishak, S., Shipley, R. J. & Arridge, S. A hybrid neural ordinary differential equation model of the cardiovascular system. *J. R. Soc. Interface* **21**, 20230710 (2024).
21. Alber, M. et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digital Med.* **2**, 115 (2019).
22. Zhang, T. et al. Two heads are better than one: current landscape of integrating qsp and machine learning: an isop qsp sig white paper by the working group on the integration of quantitative systems pharmacology and machine learning. *J. Pharmacokinetics Pharmacodynamics* **49**, 5–18 (2022).
23. Rackauckas, C. et al. Universal differential equations for scientific machine learning. arXiv preprint arXiv:2001.04385 (2020).
24. Bräm, D. S., Nahum, U., Schropp, J., Pfister, M. & Koch, G. Low-dimensional neural odes and their application in pharmacokinetics. Journal of Pharmacokinetics and Pharmacodynamics 1–18 (2023).
25. Valderrama, D., Ponce-Bobadilla, A. V., Mensing, S., Fröhlich, H. & Stodtmann, S. Integrating machine learning with pharmacokinetic models: Benefits of scientific machine learning in adding neural networks components to existing pk models. CPT: Pharmacometrics & Systems Pharmacology (2023).
26. Takeishi, N. & Kalousis, A. Deep grey-box modeling with adaptive data-driven models toward trustworthy estimation of theory-driven models. In *International Conference on Artificial Intelligence and Statistics*, 4089–4100 (PMLR, 2023).
27. Kidger, P. On neural differential equations. arXiv preprint arXiv:2202.02435 (2022).
28. Yin, Y. et al. Augmenting physical models with deep networks for complex dynamics forecasting. *J. Stat. Mech.: Theory Exp.* **2021**, 124012 (2021).
29. 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 Proc. -Syst. Biol.* **153**, 425–432 (2006).
30. Ruoff, P., Christensen, M. K., Wolf, J. & Heinrich, R. Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations. *Biophys. Chem.* **106**, 179–192 (2003).
31. Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations. Advances in neural information processing systems **31** (2018).
32. Errico, R. M. What is an adjoint model? *Bull. Am. Meteorological Soc.* **78**, 2577–2592 (1997).
33. Allaire, G. A review of adjoint methods for sensitivity analysis, uncertainty quantification and optimization in numerical codes. *Ing. énieurs de. l'Automob.* **836**, 33–36 (2015).
34. Ghosh, A., Behl, H., Dupont, E., Torr, P. & Namboodiri, V. Steer: Simple temporal regularization for neural ode. *Adv. Neural Inf. Process. Syst.* **33**, 14831–14843 (2020).
35. Kim, S., Ji, W., Deng, S., Ma, Y. & Rackauckas, C. Stiff neural ordinary differential equations. *Chaos: Interdiscip. J. Nonlinear Sci.* **31**, 093122 (2021).
36. Turan, E. M. & Jäschke, J. Multiple shooting for training neural differential equations on time series. *IEEE Control Syst. Lett.* **6**, 1897–1902 (2021).
37. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
38. Liu, D. C. & Nocedal, J. On the limited memory bfgs method for large scale optimization. *Math. Program.* **45**, 503–528 (1989).
39. Gao, Y., Yu, T. & Li, J. Bayesian optimization with local search. In *Machine Learning, Optimization, and Data Science: 6th International Conference, LOD 2020, Siena, Italy, July 19–23, 2020, Revised Selected Papers, Part II 6*, 350–361 (Springer, 2020).
40. Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B. Algorithms for hyper-parameter optimization. Advances in neural information processing systems **24** (2011).

41. Falkner, S., Klein, A. & Hutter, F. Bohb: Robust and efficient hyperparameter optimization at scale. In *International conference on machine learning*, 1437–1446 (PMLR, 2018).
42. 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 Syst. Biol.* **3**, 1–21 (2009).
43. Gutenkunst, R. N. et al. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189 (2007).
44. Transtrum, M. K. et al. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J. Chem. Phys*. **143** (2015).
45. Jagadeesan, P., Raman, K. & Tangirala, A. K. Sloppiness: Fundamental study, new formalism and its application in model assessment. *PLoS ONE* **18**, e0282609 (2023).
46. Rodriguez-Fernandez, M., Banga, J. R. & Doyle III, F. J. Novel global sensitivity analysis methodology accounting for the crucial role of the distribution of input parameters: application to systems biology models. *Int. J. Robust. Nonlinear Control* **22**, 1082–1102 (2012).
47. Tangirala, A. K. Principles of system identification: theory and practice (Crc Press, 2018).
48. Stoica, P. & Marzetta, T. L. Parameter estimation problems with singular information matrices. *IEEE Trans. Signal Process.* **49**, 87–90 (2001).
49. Wangersky, P. J. Lotka-volterra population models. *Annu. Rev. Ecol. Syst.* **9**, 189–218 (1978).
50. Verner, J. H. Numerically optimal runge–kutta pairs with interpolants. *Numer. Algorithms* **53**, 383–396 (2010).
51. Hosea, M. & Shampine, L. Analysis and implementation of tr-bdf2. *Appl. Numer. Math.* **20**, 21–37 (1996).
52. Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).
53. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937 (2016).
54. Vincent, A. M. & Jidesh, P. An improved hyperparameter optimization framework for automl systems using evolutionary algorithms. *Sci. Rep.* **13**, 4737 (2023).
55. Joshi, M., Seidel-Morgenstern, A. & Kremling, A. Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems. *Metab. Eng.* **8**, 447–455 (2006).
56. Wu, J. et al. Hyperparameter optimization for machine learning models based on bayesian optimization. *J. Electron. Sci. Technol.* **17**, 26–40 (2019).
57. Aszemi, N. M. & Dominic, P. Hyperparameter optimization in convolutional neural network using genetic algorithms. *Int. J. Adv. Comput. Sci. Appl.* **10** (2019).
58. Glorot, X. & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 249–256 (JMLR Workshop and Conference Proceedings, 2010).
59. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM Rev.* **59**, 65–98 (2017).
60. 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).

Acknowledgements

This study received no funding.

Author contributions

A.F., F.R., and S.G. ideated the workflow and selected the test cases under the supervision of L.M. and G.I. S.G. developed the computational workflow. L.M. and G.I. provided overall guidance for the project. All authors contributed to manuscript preparation and approved the final version.

Competing interests

A.F. was not affiliated with Roche Pharma Research and Early Development during the conduct of this research and therefore declares no competing interests. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41540-024-00460-3.

Correspondence and requests for materials should be addressed to Luca Marchetti.