Supplemental Material for "Non-Negative Universal Differential Equations With Applications in Systems Biology"

1 Lotka-Volterra: Small initialisation

To investigate how the initialisation of mechanistic parameters affects optimisation performance, we initialised mechanistic parameters θ_M (1) "standard": randomly (uniform distribution) in the log-scaled $[-3, 3]$ domain, and (2) "small": near the lower bound, i.e. uniformly in the log-scaled $[-3, -2.5]$ domain. The results in this section compare this difference from initialisation in multi-start gradient-based local optimisation. To ensure results can be attributed to the change in initialisation, we sampled the neural network parameters 1000 times then reused these 1000 θ_U vectors across all experiments. UDEs and nUDEs use the same samples of initial θ_M parameters. Hence, we have four cases for initialisation of θ_M: UDEs with standard startpoints; UDEs with small startpoints; nUDEs with standard startpoints; and nUDEs with small startpoints.

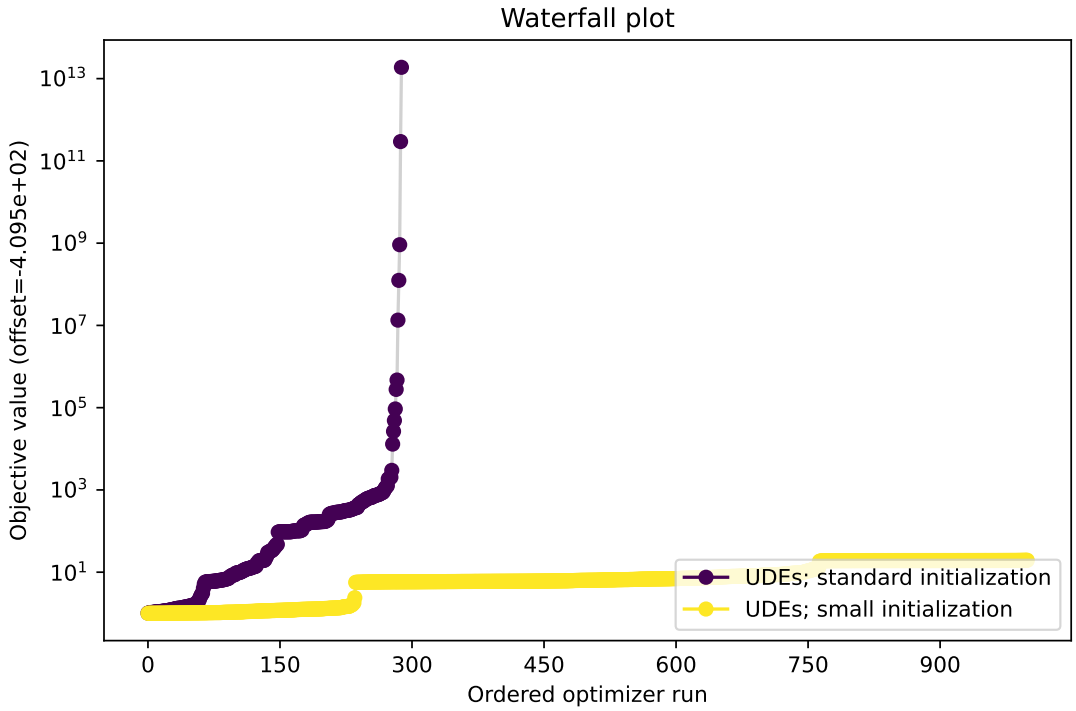

Figure 1: Unpaired waterfall plot of UDEs with small or standard startpoint initialisation.

Using small startpoints for mechanistic parameters improved optimisation substantially. The number of evaluable startpoints with UDEs increased substantially: 289/1000 standard startpoints were evaluable, whereas 1000/1000 small startpoints were evaluable (Supp. Fig. 1). As the mechanistic parameters α and γ in the UDEs (see main paper, Eq. 6) correspond to exponential growth or death terms, the small startpoints may have improved the number of evaluable startpoints by reducing the number of solutions with blow-up. Furthermore, comparing evaluable standard startpoints against evaluable small startpoints, the small startpoints generally resulted in better

optimisation endpoints. These results in UDEs were similar for nUDEs (Supp. Fig. 2), where 225/1000 standard startpoints were evaluable, and 1000/1000 small startpoints were evaluable.

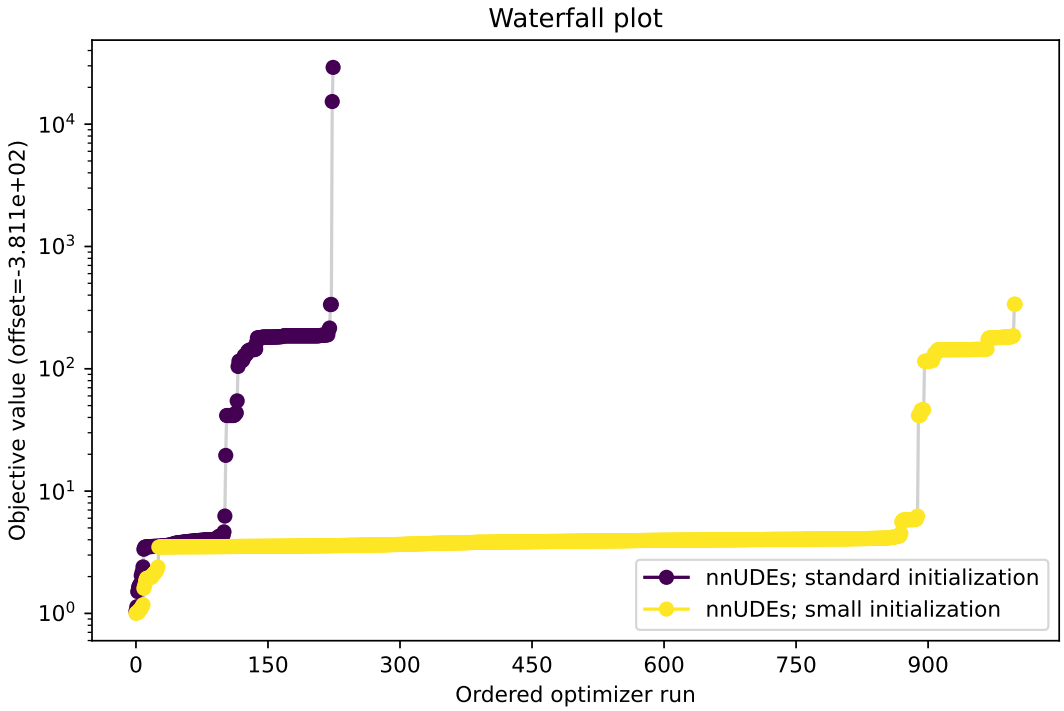

Figure 2: Unpaired waterfall plot of nUDEs with small or standard startpoint initialisation.

Although these results suggest that small startpoints for mechanistic parameters improve optimisation performance in UDEs and nUDEs, small startpoints were not used in the experiments in the main paper, because it can bias the parameter estimation. In practice, however, using small startpoints for the mechanistic parameters when high rates of non-evaluable standard startpoints are encountered, may help improve the optimisation problem.

2 Boehm model

2.1 Model specification

The system of ODEs below specifies the fully-mechanistic Boehm model. Red colour indicates the components that are dropped for the UDE.

$$
\begin{aligned}
\frac{d\ \text{STAT5A}}{dt} = \quad & -\text{BaF3\_Epo} \cdot [\text{STAT5A}]^2 \cdot \text{k\_phos} \\
& - (\text{BaF3\_Epo} \cdot [\text{STAT5A}] \cdot [\text{STAT5B}] \cdot \text{k\_phos}) \\
& + 2\ \text{k\_exp\_homo} \cdot [\text{nucpApA}] \\
& + \text{k\_exp\_hetero} \cdot [\text{nucpApB}] \\
\frac{d\ \text{STAT5B}}{dt} = \quad & -\text{BaF3\_Epo} \cdot [\text{STAT5A}] \cdot [\text{STAT5B}] \cdot \text{k\_phos} \\
& - 2\ (\text{BaF3\_Epo} \cdot [\text{STAT5A}]^2 \cdot \text{k\_phos}) \\
& + \text{k\_exp\_hetero} \cdot [\text{nucpApB}] \\
& + 2\ \text{k\_exp\_homo} \cdot [\text{nucpBpB}] \\
\frac{d\ \text{pApA}}{dt} = \quad & \text{BaF3\_Epo} \cdot [\text{STAT5A}]^2 \cdot \text{k\_phos} - \ \text{k\_imp\_homo} \cdot [\text{pApA}] \\
\frac{d\ \text{pBpB}}{dt} = \quad & \text{BaF3\_Epo} \cdot [\text{STAT5B}]^2 \cdot \text{k\_phos} - \ \text{k\_imp\_homo} \cdot [\text{pBpB}] \\
\frac{d\ \text{pApB}}{dt} = \quad & \text{BaF3\_Epo} \cdot [\text{STAT5A}] \cdot [\text{STAT5B}] \cdot \text{k\_phos} \\
& - \ \text{k\_imp\_hetero} \cdot [\text{pApB}] \\
\frac{d\ \text{nucpApA}}{dt} = \quad & \text{k\_imp\_homo} \cdot [\text{pApA}] \ - \ \text{k\_exp\_homo} \cdot [\text{nucpApA}] \\
\frac{d\ \text{nucpApB}}{dt} = \quad & \text{k\_imp\_hetero} \cdot [\text{pApB}] - \text{k\_exp\_hetero} \cdot [\text{nucpApB}] \\
\frac{d\ \text{nucpBpB}}{dt} = \quad & \text{k\_imp\_homo} \cdot [\text{pBpB}] - \text{k\_exp\_homo} \cdot [\text{nucpBpB}] \\
\text{BaF3\_Epo} = \quad & 1.25e-07^{-\text{Epo\_degradation\_BaF3} \cdot t}
\end{aligned}
\tag{1}
$$

The state variables are linked to the measurements through an observable mapping given by:

$$
\text{pSTAT5A\_rel} = \frac{100 \cdot \text{pApB} + 200 \cdot \text{pApA} \cdot \text{specC17}}{\text{pApB} + \text{STAT5A} \cdot \text{specC17} + 2 \cdot \text{pApA} \cdot \text{specC17}}
$$

$$
\text{pSTAT5B\_rel} = -\frac{100 \cdot \text{pApB} - 200 \cdot \text{pBpB} \cdot (\text{specC17} - 1)}{(\text{STAT5B} \cdot (\text{specC17} - 1) - \text{pApB}) + 2 \cdot \text{pBpB} \cdot (\text{specC17} - 1)}
$$

$$
\text{rSTAT5A\_rel} = \frac{100 \cdot \text{pApB} + 100 \cdot \text{STAT5A} \cdot \text{specC17} + 200 \cdot \text{pApA} \cdot \text{specC17}}{2 \cdot \text{pApB} + \text{STAT5A} \cdot \text{specC17} + 2 \cdot \text{pApA} \cdot \text{specC17} - \text{STAT5B} \cdot (\text{specC17} - 1) - 2 \cdot \text{pBpB} \cdot (\text{specC17} - 1)}
$$

Here we specify how the ANN used in the UDE from *Scenario 1* modifies the dynamics given in Supp. Eq. 1. The blue components are modelled by an ANN, $\mathbf{U}$, with the indices corresponding to the two outputs of $\mathbf{U}$.

$$
\begin{aligned}
\frac{d\ \text{STAT5A}}{dt} = \quad & -\text{BaF3\_Epo} \cdot [\text{STAT5A}]^2 \cdot \text{k\_phos} \\
& - (\text{BaF3\_Epo} \cdot [\text{STAT5A}] \cdot [\text{STAT5B}] \cdot \text{k\_phos}) \\
& + \mathbf{U}_1(\text{nucpApA}) \\
& + \text{k\_exp\_hetero} \cdot [\text{nucpApB}] \\
\frac{d\ \text{nucpApA}}{dt} = \quad & \text{k\_imp\_homo} \cdot [\text{pApA}] + \ \mathbf{U}_2(\text{nucpApA})
\end{aligned}
\tag{2}
$$

Below we specify how the ANN used for the UDE in *Scenario 2* modifies the dynamics given in Supp. Eq. 1. The blue components are modelled by an ANN, $\mathbf{U}$, with the indices corresponding to the five outputs of $\mathbf{U}$. The species used as input to $\mathbf{U}$ are $\boldsymbol{\nu} = \texttt{nucpApA, nucpApB, nucpBpB}$.

Table 1: Boehm UDE setup.

	Scenario 1	Scenario 2
ANN architecture	5/5/5/2	5/5/5/5
Inputs	nucpApA	nucpApA, nucpApB, nucpBpB
Outputs	nucpApA, STAT5A	nucpApA, nucpApB, nucpBpB, STAT5A, STAT5B
Activation function	tanh	tanh
Size of θ_U	82	110

$$
\begin{aligned}
\frac{d\ \text{STAT5A}}{dt} = \quad & -\text{BaF3\_Epo} \cdot [\text{STAT5A}]^2 \cdot \text{k\_phos} \\
& - \ (\text{BaF3\_Epo} \cdot [\text{STAT5A}] \cdot [\text{STAT5B}] \cdot \text{k\_phos}) \\
& + \ \mathbf{U}_1(\boldsymbol{\nu}) \\
& + \text{k\_exp\_hetero} \cdot [\text{nucpApB}] \\
\frac{d\ \text{STAT5B}}{dt} = \quad & -\text{BaF3\_Epo} \cdot [\text{STAT5A}] \cdot [\text{STAT5B}] \cdot \text{k\_phos} \\
& - \ 2\ (\text{BaF3\_Epo} \cdot [\text{STAT5A}]^2 \cdot \text{k\_phos}) \\
& + \text{k\_exp\_hetero} \cdot [\text{nucpApB}] \\
& + 2\ \text{k\_exp\_homo} \cdot [\text{nucpBpB}] \\
& + \ \mathbf{U}_2(\boldsymbol{\nu}) \\
\frac{d\ \text{nucpApA}}{dt} = \quad & \text{k\_imp\_homo} \cdot [\text{pApA}] + \ \mathbf{U}_3(\boldsymbol{\nu}) \\
\frac{d\ \text{nucpApB}}{dt} = \quad & \text{k\_imp\_hetero} \cdot [\text{pApB}] - \text{k\_exp\_hetero} \cdot [\text{nucpApB}] + \ \mathbf{U}_4(\boldsymbol{\nu}) \\
\frac{d\ \text{nucpBpB}}{dt} = \quad & \text{k\_imp\_homo} \cdot [\text{pBpB}] - \text{k\_exp\_homo} \cdot [\text{nucpBpB}] + \ \mathbf{U}_5(\boldsymbol{\nu})
\end{aligned}
\tag{3}
$$

Table 1 compares the UDE setups used in *Scenario 1* and *Scenario 2*.

2.2 Training details

All entries of $\boldsymbol{\theta}_{\mathrm{M}}$ were initialised randomly within $[10^{-5}, 10^{5}]$ (uniform distribution), and were constrained by the same bounds during estimation. The measurements were taken from the original publication [2] and the parameter estimation problem was adapted from implementation available in the PEtab benchmark collection [1].

Multi-start optimisation was performed with 1000 start points that were shared among comparable experimental setups that had the same ANN architecture and the same regularisation type. We combined the models specified for *Scenario 1* and *2* with all combinations of nUDE constraint (UDE, a $\mathbf{N}(x) = x$-nUDE and a $\mathbf{N}(x) = \tanh(x)$-nUDE) and regularisation (*parameter* and *output* regularisation). To find the optimal regularisation strengths λ_p, λ_o, we repeated the calibration procedure for five settings of the regularisation strengths {1e-2, 1e-1, 1e0, 1e1, 1e2}.

2.3 Results

Here we present additional results obtained from the Boehm model calibration.

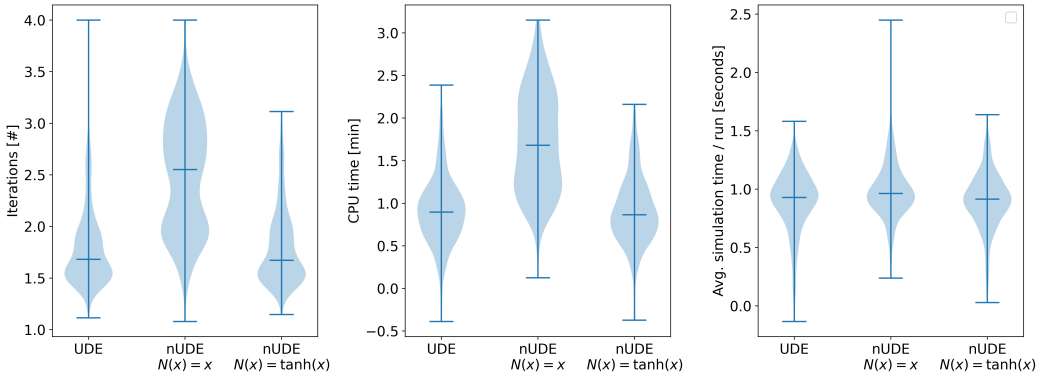

Figure 3: Computational Performance of the *Scenario 1* UDE, $N(x) = x$-nUDE and $N(x) = \tanh(x)$-nUDE. Horizontal lines in violin plots indicate minimum, median and maximum.

Table 2: Boehm *Scenario 2*: NMSE between the reference ODE solution and the best UDE models with and without regularisation. Best NMSE per regularisation type in bold numbers.

Regularisation type	λ					
	0	0.01	0.1	1	10	100
Parameter	145.993	127.900	61.223	49.504	**12.840**	13.513
Output	145.993	21.215	**12.691**	13.995	1.78e+7	1.78e+7

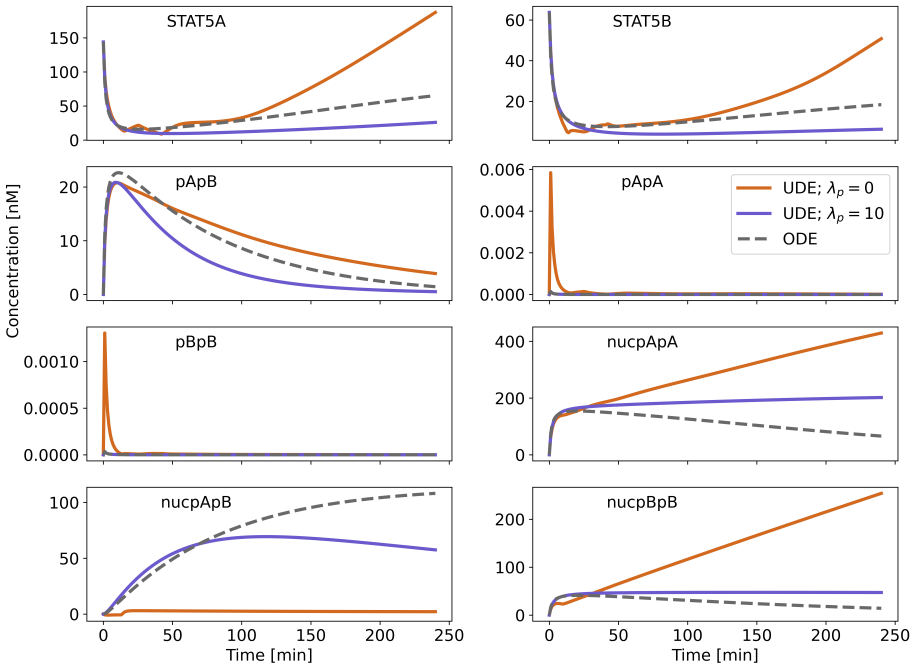

Figure 4: Simulation of the best UDEs in *Scenario 2*, without (orange) and with (purple, $\lambda_p = 10$) parameter regularisation. Shown are the simulations for all 8 non-observed state variables in comparison to the reference from optimising the full mechanistic model (grey, dashed line).

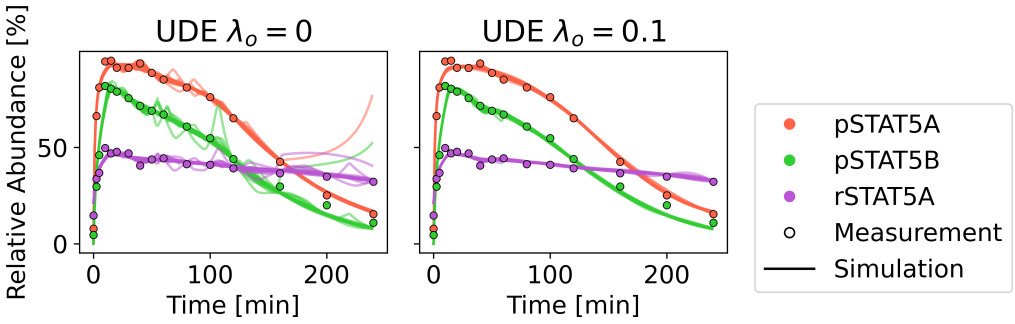

Figure 5: Ensembles of 20 best UDE model variants in *Scenario 2*. Best fits are shown for the unregularised UDE and an output-regularised UDE with $\lambda_o = 0.1$.

References

[1] "Benchmarking-Initiative/Benchmark-Models-PEtab: Benchmark Collection as at 2023-07-17". In: (2023). DOI: 10.5281/zenodo.8155058.

[2] Martin E Boehm, Lorenz Adlung, Marcel Schilling, et al. "Identification of isoform-specific dynamics in phosphorylation-dependent STAT5 dimerization by quantitative mass spectrometry and mathematical modeling". In: *J. Proteome Res.* 13.12 (2014), pp. 5685–5694.