S1 UTILISED ALGORITHMS

Since analytical solutions of non-linear ODE systems are in general not available, a numerical integration has to be performed. In this work, the dynamical system and its sensitivities were integrated by the *CVODES* integrator of the *SUNDIALS* suite (Hindmarsh *et al.*, 2005; Serban and Hindmarsh, 2005). Therein, an implicit *BDF* integration method (Gear, 1971) with attached *KLU* sparse solver was chosen (Davis and Palamadai Natarajan, 2010). Numerical optimisation was conducted using a trust-region based, large scale nonlinear optimisation algorithm implemented in the MATLAB function *LSQNONLIN* (Coleman and Li, 1996). The inner derivatives of the likelihood, which are used in the integration mechanism of the prediction bands (PBs) and in gradient-based parameter estimation were computed via supplied forward sensitivities (Leis and Kramer, 1988). The advantage of this procedure is the better accuracy of the sensitivities. For the mathematical modelling and visualisation, the open-source and freely available d2d framework (Raue *et al.*, 2015), based on MATLAB, was used.

S2 DERIVATION OF INTEGRATION FORMULA FOR PREDICTION BANDS

The derivation of the integration mechanism for prediction bands is conducted based on Equation (10) of the main document:

$$\chi^2(\tilde{\theta}(\tau), z(\tau)) = \underbrace{\min_z VPL(z(\tau)) + icdf(\chi^2_{1,\alpha})}_{=const.}, \tag{S-1}$$

$$\nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau)) = 0\,, \tag{S-2}$$

wherein both $z(\tau)$ and $\tilde{\theta}(\tau)$ are explicitly time dependent. The auxiliary data point z_τ at the threshold of a specified confidence level α, computed via the validation profile likelihood approach, serves as starting point for the integration algorithm. Further on, a possibly time dependent control $u(\tau)$ is omitted and $\theta(\tau)$ is substituted by θ for better readability. Inserting the negative log-likelihood of the validation profile likelihood and its derivative with respect to the parameters θ (see Equation (7)), Equations (S-1) and (S-2) read

$$\begin{aligned}\chi^2(\tilde{\theta}, z(\tau)) &= \underbrace{\sum_i \left(\frac{y_i - g(x(t_i,\tilde{\theta}),\tilde{\theta}))}{\sigma_i}\right)^2}_{=:\Phi(\tilde{\theta})} + \\ &+ \underbrace{\left(\frac{z(\tau) - g_z(x(\tau,\tilde{\theta}),\tilde{\theta}))}{\tilde{\sigma}}\right)^2}_{\text{Contribution of auxiliary data point } =\tilde{\chi}^2(\tilde{\theta}, z(\tau))} = const.\,,\end{aligned} \tag{S-3}$$

and

$$\begin{aligned}&\nabla_\theta \chi^2(\tilde{\theta}, z(\tau)) = \\ &- \sum_i 2\left(\frac{y_i - g(x(t_i,\tilde{\theta}),\tilde{\theta}))}{\sigma_i}\right)\left(\frac{\partial g_z}{\partial x}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} + \left.\frac{\partial g_z}{\partial \theta}\right|_{\tilde{\theta}}\right) - \\ &- 2\left(\frac{z(\tau) - g_z(x(\tau,\tilde{\theta}),\tilde{\theta}))}{\tilde{\sigma}}\right)\left(\frac{\partial g_z}{\partial x}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} + \left.\frac{\partial g_z}{\partial \theta}\right|_{\tilde{\theta}}\right) = 0\,.\end{aligned} \tag{S-4}$$

To obtain the time course of the prediction interval, we first calculate the time derivative of Equation (S-3):

$$\begin{aligned}&\frac{d}{d\tau}\left(\chi^2(\tilde{\theta}(\tau), z(\tau)) - const.\right) \\ &= \nabla_\theta \Phi|_{\tilde{\theta}} \frac{\partial \tilde{\theta}}{\partial \tau} + \frac{2\left[g_z(x(\tau,\tilde{\theta}),\tilde{\theta}) - z(\tau)\right]}{\tilde{\sigma}^2} \times \\ &\times \left(\frac{\partial g_z}{\partial x}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}}\frac{\partial \tilde{\theta}}{\partial \tau} + \left.\frac{\partial g_z}{\partial \theta}\right|_{\tilde{\theta}}\frac{\partial \tilde{\theta}}{\partial \tau} + \frac{\partial g_z}{\partial x}\frac{\partial x}{\partial \tau} - \frac{dz}{d\tau}\right) \\ &= \underbrace{\nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau))|_{\tilde{\theta}}\frac{\partial \tilde{\theta}}{\partial \tau}}_{=0} + \underbrace{2\left[g_z(x(\tau,\tilde{\theta}),\tilde{\theta}) - z(\tau)\right]/\tilde{\sigma}^2}_{\neq 0} \times \\ &\times \left(\frac{\partial g_z}{\partial x}\frac{\partial x}{\partial \tau} - \frac{dz}{d\tau}\right) \overset{!}{=} 0,\end{aligned} \tag{S-5}$$

whereby the first term equals zero because $VPL(z(\tau))$ minimises the likelihood function. Thus, the ODE for the time evolution of $g_z(x(\tau,\tilde{\theta}),\tilde{\theta})$ can be stated:

$$\frac{\partial g_z(x(\tau,\tilde{\theta}),\tilde{\theta})}{\partial \tau} = \frac{dz}{d\tau}. \tag{S-6}$$

To obtain an ODE for the parameters $\tilde{\theta}$, Equation (S-2) is differentiated with respect to τ,

$$\begin{aligned}&\frac{d}{d\tau}\left(\nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau))\right) = H\Phi|_{\tilde{\theta}}\frac{d\tilde{\theta}}{d\tau} + \\ &+ \frac{2\left[\frac{\partial g_z}{\partial x}\frac{\partial x}{\partial \tau} + \frac{\partial g_z}{\partial x}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}}\frac{d\tilde{\theta}}{d\tau} + \left.\frac{\partial g_z}{\partial \theta}\right|_{\tilde{\theta}}\frac{d\tilde{\theta}}{d\tau} - \frac{dz}{d\tau}\right]}{\tilde{\sigma}^2} \times \\ &\times \left(\frac{\partial g_z}{\partial x}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} + \left.\frac{\partial g_z}{\partial \theta}\right|_{\tilde{\theta}}\right) + \\ &+ \frac{2[g_z(x(\tau,\tilde{\theta}),\tilde{\theta}) - z(\tau)]}{\tilde{\sigma}^2}\Bigg(\frac{\partial g_z}{\partial x}\frac{\partial}{\partial \tau}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} + \\ &+ \left.\frac{\partial^2 g_z}{\partial x \partial \theta}\right|_{\tilde{\theta}}\frac{\partial x}{\partial \tau} + \frac{\partial^2 g_z}{\partial x^2}\frac{\partial x}{\partial \tau}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} + \\ &+ \left(\frac{\partial \tilde{\theta}}{\partial \tau}\right)^2\left(2\left.\frac{\partial^2 g_z}{\partial x \partial \theta}\right|_{\tilde{\theta}}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} + \frac{\partial g_z}{\partial x}\left.\frac{\partial^2 x}{\partial \theta^2}\right|_{\tilde{\theta}} + \left.\frac{\partial^2 g_z}{\partial \theta^2}\right|_{\tilde{\theta}}\right)\Bigg),\end{aligned}$$

which leads with Equation (S-5) and omitted second order derivatives to

$$\begin{aligned}&\nabla_\theta^\top \nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau))\frac{d\tilde{\theta}}{d\tau} + \frac{2[g_z(x(\tau,\tilde{\theta}),\tilde{\theta}) - z(\tau)]}{\tilde{\sigma}^2} \times \\ &\times \frac{\partial g_z}{\partial x}\frac{\partial}{\partial \tau}\left.\frac{\partial x}{\partial \theta}\right|_{\tilde{\theta}} \overset{!}{=} -\gamma \nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau)).\end{aligned} \tag{S-7}$$

As described in the main text, the self-correction term on the right hand side of equation (S-7) compensates for the omitted second order derivatives. The second ODE expressing the time evolution of $\tilde{\theta}$ is then obtained as

$$\frac{d}{d\tau}\tilde{\theta} = (\nabla_\theta^\top \nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau)))^{-1} \times$$
$$\times \left(\left(\frac{2[z(\tau) - g_z(x(\tau, \tilde{\theta}), \tilde{\theta})]}{\tilde{\sigma}^2} \right) \frac{\partial g_z}{\partial x} \frac{\partial}{\partial \tau} \frac{\partial x}{\partial \theta} \bigg|_{\tilde{\theta}} - \right. \tag{S-8}$$
$$\left. - \gamma \nabla_\theta \chi^2(\tilde{\theta}(\tau), z(\tau)) \right) .$$

The integration of prediction bands is performed via the Runge-Kutta scheme of fourth order (Butcher, 1963). Hence, both $\dot{z}(\tau, \tilde{\theta})$ and $\dot{\tilde{\theta}}(\tau, \tilde{\theta})$ of Equations S-6 and S-8, respectively, are evaluated to obtain

$$k_1 = \begin{pmatrix} \dot{z}(\tau, \tilde{\theta}) \\ \dot{\tilde{\theta}}(\tau, \tilde{\theta}) \end{pmatrix} . \tag{S-9}$$

Then, k_2, k_3, k_4 are consecutively computed through

$$\begin{aligned} k_2 &= \begin{pmatrix} \dot{z}(\tau + \frac{\Delta\tau}{2}, \tilde{\theta} + \frac{\Delta\tau}{2} k_1) \\ \dot{\tilde{\theta}}(\tau + \frac{\Delta\tau}{2}, \tilde{\theta} + \frac{\Delta\tau}{2} k_1) \end{pmatrix} , \\ k_3 &= \begin{pmatrix} \dot{z}(\tau + \frac{\Delta\tau}{2}, \tilde{\theta} + \frac{\Delta\tau}{2} k_2) \\ \dot{\tilde{\theta}}(\tau + \frac{\Delta\tau}{2}, \tilde{\theta} + \frac{\Delta\tau}{2} k_2) \end{pmatrix} , \\ k_4 &= \begin{pmatrix} \dot{z}(\tau + \Delta\tau, \tilde{\theta} + \Delta\tau k_3) \\ \dot{\tilde{\theta}}(\tau + \Delta\tau, \tilde{\theta} + \Delta\tau k_3) \end{pmatrix} . \end{aligned} \tag{S-10}$$

Thereby, $\Delta\tau$ denotes the step size of the integration. Finally, z and $\tilde{\theta}$ are updated via

$$\begin{pmatrix} z(\tau_i + \Delta\tau) \\ \tilde{\theta}(\tau_i + \Delta\tau) \end{pmatrix} = \begin{pmatrix} z(\tau_i) \\ \tilde{\theta}(\tau_i) \end{pmatrix} + \frac{\Delta\tau}{6}(k_1 + 2k_2 + 2k_3 + k_4) . \tag{S-11}$$

S3 INTEGRATION WITHIN THE DATA2DYNAMICS FRAMEWORK

The d2d framework can be found on 'https://bitbucket.org/d2d-development/d2d-software/wiki/Home' and requires MATLAB with symbolic and numerical toolbox as well as a working C compiler set up as MEX compiler in MATLAB. After installation of the d2d framework, manuals about model setup and analysis can be found on the homepage as well. Further, an introduction on prediction bands and its computation in the framework can be found at 'Instructions on prediction bands'. In addition, the examples of this paper are included and can be found in the folder 'Examples'. Exemplary, the setup file of the ABC_toyModel folder is stated here, with description in italic. Execution of this file in MATLAB will automatically perform all these steps and calculate the PBs.

Initialise framework:
arInit;

Load ABC toy model:
arLoadModel('ABC_model');
Load data with equidistant observation of state B and C for t=0,10,..100:
arLoadData('ABC_data_BCobs');

Write and compile C files for ODE integration with forward sensitivities:
arCompileAll();

Take measurement errors as provided by the data file:
ar.config.fiterrors = -1;
arSetPars('sd_B_au',[],2);
arSetPars('sd_C_au',[],2);

Optimise likelihood to obtain best fit:
arFit();

Get information about prediction profile function:
help doPPL

Calculate prediction bands for the three states:
doPPL(1,1,1:3,0,0,1,0.25);

plot prediction bands:
ar.config.ploterrors = -1;
arPlot2

S4 ABC TOY EXAMPLE AND ANALYTICAL SOLUTION

In this section, detailed information about the ABC toy model of Section 3.1 of the main text, available in the folder 'ABC_toyModel' in the d2d framework, will be given. The model contains three states, A_state, B_state and C_state, with the following ODE system determining the time evolution of the dynamical variables:

$$\begin{aligned} \mathrm{d[A\_state]/dt} &= & -p_1\mathrm{A\_state} \\ \mathrm{d[B\_state]/dt} &= & +p_1\mathrm{A\_state} - p_2\mathrm{B\_state} \\ \mathrm{d[C\_state]/dt} &= & +p_2\mathrm{B\_state} \end{aligned} \tag{S-12}$$

Further, the model contains 2 observables:

- **Observable 1:** C

  $$\mathrm{C(t)} = \quad \mathrm{[C\_state]}$$

  With error model:

  $$\sigma\{\mathrm{C}\}(\mathrm{t}) = \quad 0.1$$

- **Observable 2:** B

  $$\mathrm{B(t)} = \quad \mathrm{[B\_state]}$$

  With error model:

  $$\sigma\{\mathrm{B}\}(\mathrm{t}) = \quad 0.1$$

In addition, the initial values of $B(t)$ and $C(t)$ are fixed to zero:

$$\begin{aligned} \mathrm{init\_B\_state} &\rightarrow 0 \\ \mathrm{init\_C\_state} &\rightarrow 0 \end{aligned}$$

S4.1 Analytical solution

In the paper of Gellene (1995), the analytical solution of the ABC system is discussed in detail. It is given by

$$
\begin{aligned}
\mathrm{d[A\_state]/dt} &= \mathrm{init\_A\_state}\ \exp^{-p_1 t}\ , \\
\mathrm{d[B\_state]/dt} &= \frac{p_1 \mathrm{init\_A\_state}}{p_2 - p_1}\left(\exp^{-p_1 t} - \exp^{-p_2 t}\right)\ , \\
\mathrm{d[C\_state]/dt} &= \mathrm{init\_A\_state}\left(1 + \frac{p_1 \exp^{-p_2 t} - p_2 \exp^{-p_1 t}}{p_2 - p_1}\right)\ .
\end{aligned}
\tag{S-13}
$$

Based on this, the first and second-order derivatives with respect to the parameters p_1, p_2 and $\mathrm{init\_A\_state}$ can be calculated. Equation (S-13) and its sensitivities can be used instead of an integration with forward sensitivities in order to calculate the validation profile likelihood for the ABC toy model for distinct time points. In Figure 1 of the main text, the 95% thresholds of the analytically derived validation profiles are taken for comparison with the PBs. In addition, a stand-alone MATLAB code of the ABC toy model can be found in a zip file at http://www.fdmold.uni-freiburg.de/~hhass. It features the exact data points used here and analytical solutions for the calculation of the validation profiles in order to reproduce our results.

S4.2 Model fit and plots

The agreement of the model observables and the simulated data, given in Table ST1, yields a value of the objective function $\chi^2 = 11.51$ for 21 data points in this data set. The model observables and the simulated data are shown in Figure SF1.

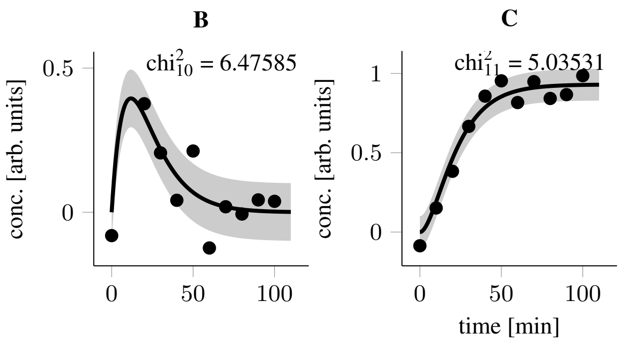

Fig. SF1: **ABC toy model observables and simulated data.** The observables are displayed as solid lines. The error describing the measurement noise is indicated by shades.

S4.3 Estimated model parameters

In Table ST2 the estimated parameter values are given. The parameter name prefix init_ indicates the initial value of a dynamic variable.

name	time [min]	value [au]	σ [au]
C_au	0	-0.086488	0.1
C_au	10	0.151805	0.1
C_au	20	0.383077	0.1
C_au	30	0.666288	0.1
C_au	40	0.856965	0.1
C_au	50	0.953492	0.1
C_au	60	0.81654	0.1
C_au	70	0.948258	0.1
C_au	80	0.842301	0.1
C_au	90	0.866566	0.1
C_au	100	0.985896	0.1
B_au	0	-0.0809499	0.1
B_au	20	0.376395	0.1
B_au	30	0.205873	0.1
B_au	40	0.0415357	0.1
B_au	50	0.212383	0.1
B_au	60	-0.123839	0.1
B_au	70	0.0190627	0.1
B_au	80	-0.00616601	0.1
B_au	90	0.0429027	0.1
B_au	100	0.037974	0.1

Table ST1. Simulated data for the ABC toy model

name	$\log_{10}(\theta_{min})$	$\log_{10}(\tilde{\theta})$	$\log_{10}(\theta_{max})$	$\tilde{\theta}$
init_A_state	-5	-0.0315	+3	$9.30 \cdot 10^{-01}$
p1	-5	-1.0097	+3	$9.78 \cdot 10^{-02}$
p2	-5	-1.1405	+3	$7.24 \cdot 10^{-02}$

Table ST2. Estimated parameter values
$\log_{10}(\tilde{\theta})$ indicates the estimated value of the parameters. $\log_{10}(\theta_{min})$ and $\log_{10}(\theta_{max})$ indicate the upper and lower bounds for the parameters. The $\tilde{\theta}$-column indicates the non-logarithmic value of the parameter estimate.

S5 ESTABLISHED MODELS

To verify the usability and efficiency of the integration method, PBs were calculated for three established models for cellular signalling, of Bachmann *et al.* (2011), Raia *et al.* (2011) and Swameye *et al.* (2003). Detailed information and supplementary data of these models can be found at 'Bachmann et. al paper', 'Raia et. al paper' and 'Swameye et. al paper'. In addition, the data in xls format and the ODE systems for these models can be found within the d2d framework, or on its homepage in the following folders:

- 'Bachmann et. al data'
- 'Bachmann et. al ODEs'
- 'Raia et. al data'
- 'Raia et. al ODEs'
- 'Swameye et. al data'
- 'Swameye et. al ODEs'

REFERENCES

Bachmann, J., Raue, A., *et al.* (2011). Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Molecular Systems Biology*, **7**(1), 516.

Butcher, J. C. (1963). Coefficients for the study of runge-kutta integration processes. *Journal of the Australian Mathematical Society*, **3**(02), 185–201.

Coleman, T. and Li, Y. (1996). An interior, trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization*, **6**, 418–445.

Davis, T. A. and Palamadai Natarajan, E. (2010). Algorithm 907: KLU, a direct sparse solver for circuit simulation problems. *ACM Transactions on Mathematical Software (TOMS)*, **37**(3), 36.

Gear, C. W. (1971). The automatic integration of ordinary differential equations. *Communications of the ACM*, **14**(3), 176–179.

Gellene, G. I. (1995). Application of kinetic approximations to the $A \rightleftarrows B \to C$ reaction system. *Journal of Chemical Education*, **72**(3), 196.

Hindmarsh, A. C., Brown, P. N., *et al.* (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software (TOMS)*, **31**(3), 363–396.

Leis, J. R. and Kramer, M. A. (1988). The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Transactions on Mathematical Software (TOMS)*, **14**(1), 45–60.

Raia, V., Schilling, M., *et al.* (2011). Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets. *Cancer Research*, **71**(3), 693–704.

Raue, A., Steiert, B., *et al.* (2015). Data2dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*, **31**(21), 3558–3560.

Serban, R. and Hindmarsh, A. C. (2005). CVODES: the sensitivity-enabled ODE solver in SUNDIALS. In *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pages 257–269. American Society of Mechanical Engineers.

Swameye, I., Müller, T., *et al.* (2003). Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences*, **100**(3), 1028–1033.