BioPSy: An SMT-based Tool for Guaranteed Parameter Set Synthesis of Biological Models

Curtis Madsen, Fedor Shmarov, and Paolo Zuliani

School of Computing Science, Newcastle University, Newcastle upon Tyne, UK

17 September 2015
CMSB

Outline

Introduction

- A key component to modelling biological systems is selecting the correct model parameters (e.g., reaction rate constants).
- Small parameter variations can lead to vastly different results when simulating biological models.
- Methods for performing parameter synthesis and estimation are necessary to determine acceptable values for the parameters of such models.

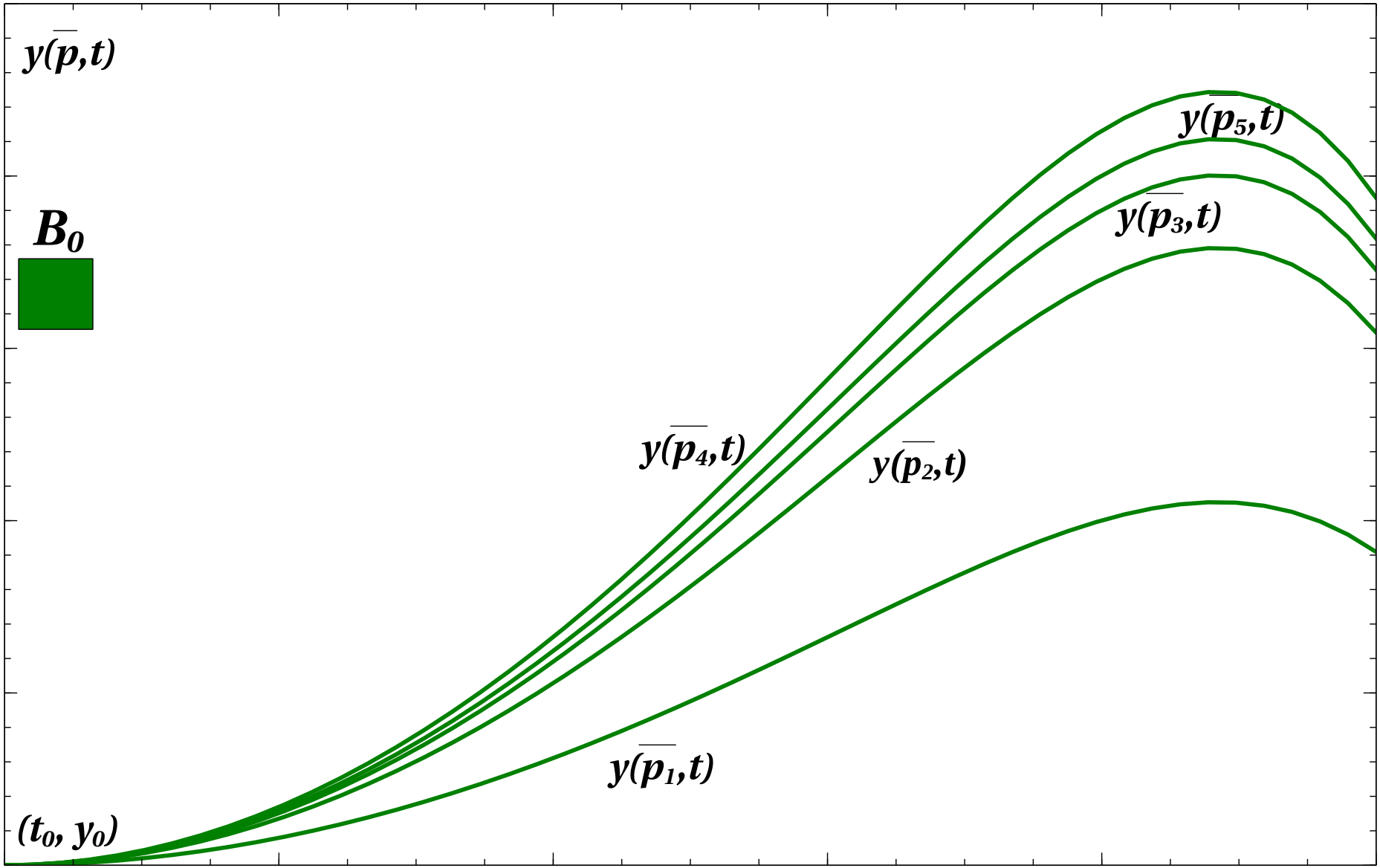
Parameter Synthesis

- The *parameter set synthesis* problem consists of identifying sets of parameter values for which a given system model satisfies a desired behaviour.
- This problem is categorised as a reachability problem where the solution to a set of ODEs can be computed but some parameter values that lead to the solution are missing.
- For example:
  - The parameter, k, can be synthesised in the ODE model given by $x'(t) = kt$.
  - Given the time-series data in which $x = \{0, 1, 4, 9\}$ for $t = \{0, 1, 2, 3\}$, it is easy to see that k should be 2.
  - However, the values of x might vary by 0.1 due to noise.
  - The solution to the parameter synthesis problem might yield the interval $[1.978, 2.022]$ for k.

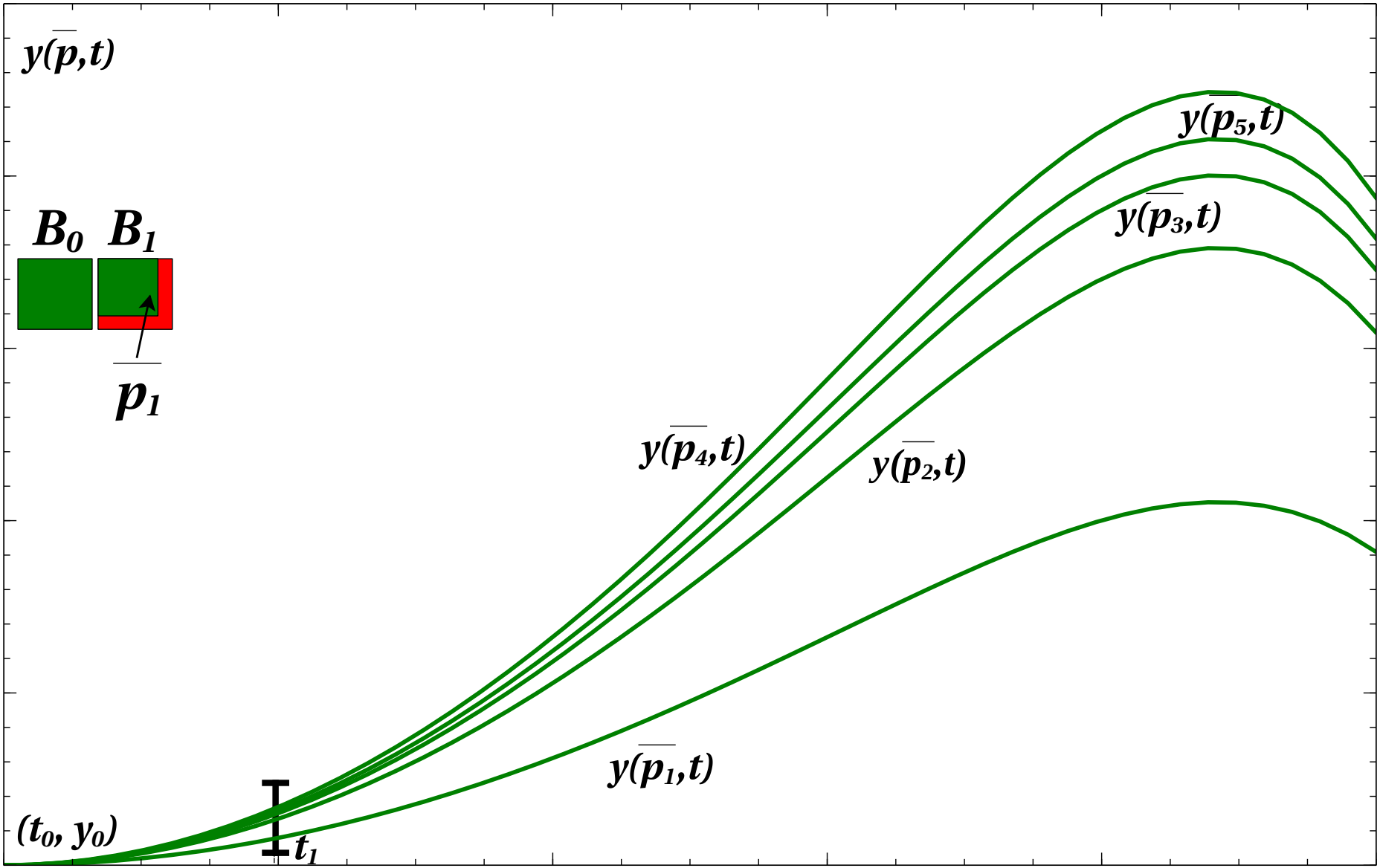
BioPSy

- We have developed BioPSy, a tool for guaranteed parameter set synthesis in biological systems.
- BioPSy is capable of determining *ranges* (intervals) of parameters for which a model's temporal behaviour remains in satisfactory states.
- It takes an SBML model and experimental data as input.
- Additionally, users can specify a list of parameters to synthesise, a tolerable noise value for the experimental data, and the parameter synthesis precision.

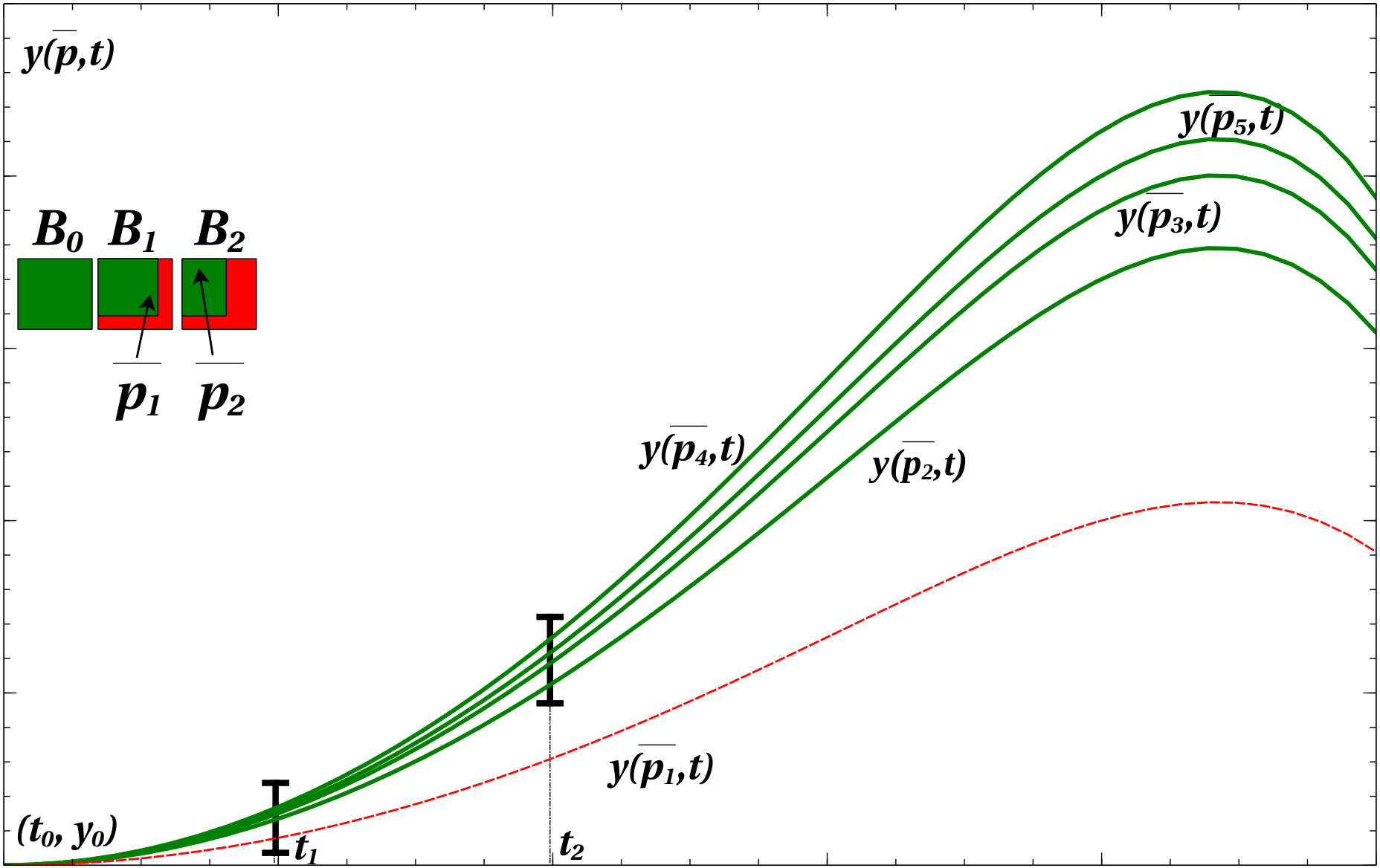
Intuition



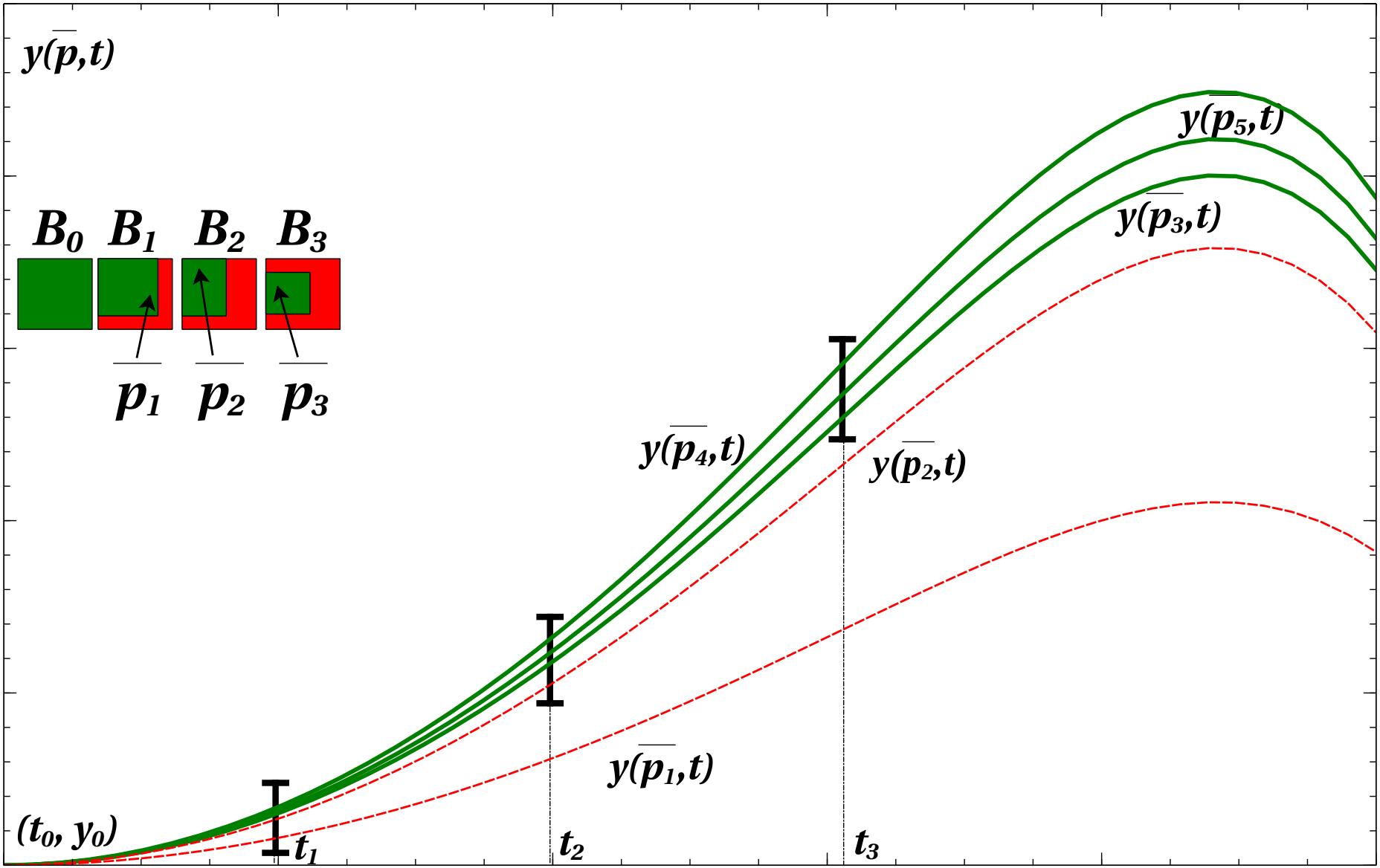
Intuition


$y(\bar{p},t)$
B_0 B_1
$\bar{p}_1$
$y(\bar{p}_5,t)$
$y(\bar{p}_3,t)$
$y(\bar{p}_4,t)$
$y(\bar{p}_2,t)$
$y(\bar{p}_1,t)$
(t_0, y_0)
t_1

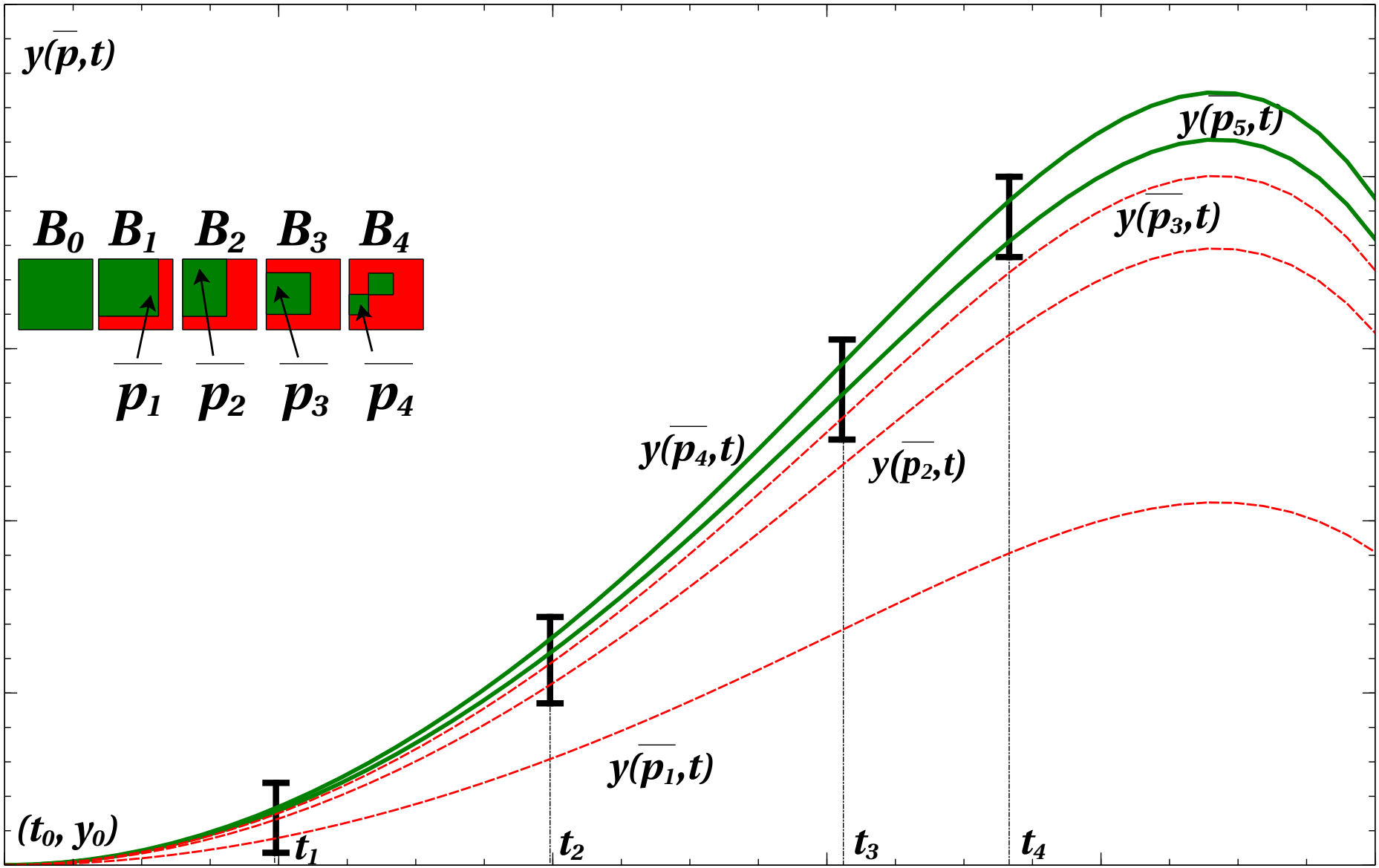
Intuition


$y(\bar{p},t)$
B_0 B_1 B_2
$\bar{p}_1$ $\bar{p}_2$
$y(\bar{p}_5,t)$
$y(\bar{p}_3,t)$
$y(\bar{p}_4,t)$
$y(\bar{p}_2,t)$
$y(\bar{p}_1,t)$
(t_0, y_0)
t_1
t_2

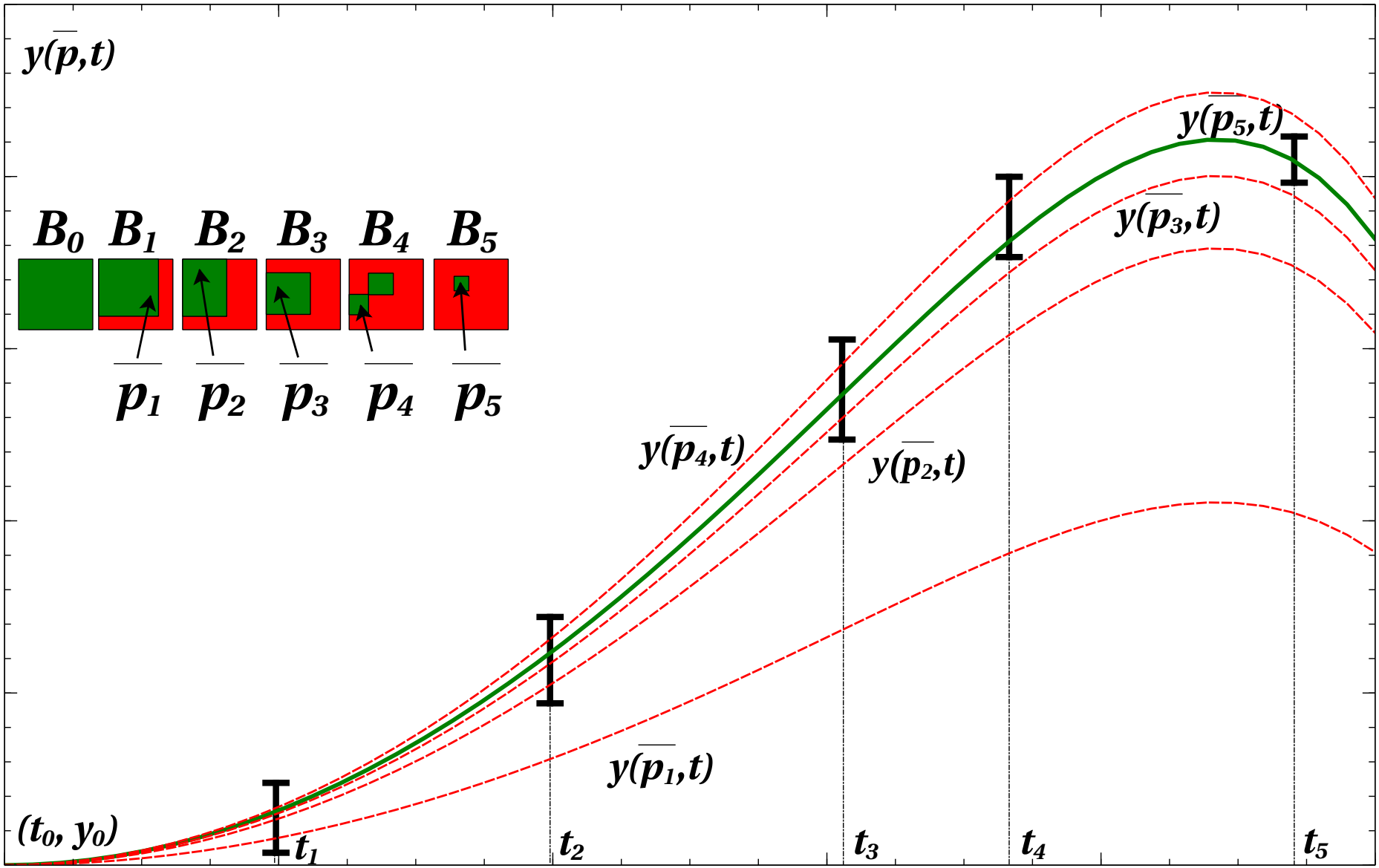
Intuition


$y(\bar{p},t)$
B_0 B_1 B_2 B_3
$\bar{p}_1$ $\bar{p}_2$ $\bar{p}_3$
$y(\bar{p}_5,t)$
$y(\bar{p}_3,t)$
$y(\bar{p}_4,t)$
$y(\bar{p}_2,t)$
$y(\bar{p}_1,t)$
(t_0, y_0)
t_1
t_2
t_3

Intuition


$y(\bar{p},t)$
B_0 B_1 B_2 B_3 B_4
$\bar{p}_1$ $\bar{p}_2$ $\bar{p}_3$ $\bar{p}_4$
$y(\bar{p}_5,t)$
$y(\bar{p}_3,t)$
$y(\bar{p}_4,t)$
$y(\bar{p}_2,t)$
$y(\bar{p}_1,t)$
(t_0, y_0)
t_1 t_2 t_3 t_4

Intuition


$y(\bar{p},t)$
B_0 B_1 B_2 B_3 B_4 B_5
$\bar{p}_1$ $\bar{p}_2$ $\bar{p}_3$ $\bar{p}_4$ $\bar{p}_5$
$y(\bar{p}_5,t)$
$y(\bar{p}_3,t)$
$y(\bar{p}_4,t)$
$y(\bar{p}_2,t)$
$y(\bar{p}_1,t)$
(t_0, y_0)
t_1 t_2 t_3 t_4 t_5

Boxes

- `BioPSy` returns three sets of boxes:
  - A set of **sat** boxes: for all the parameter values in this set, the model is *formally* and *numerically* guaranteed to satisfy the noisy time-series data;
  - A set of **unsat** boxes: for no point in this set, the model satisfies the noisy time-series data; and
  - A set of **undet** boxes: because of the given precision (or due to the undecidability of the problem), `BioPSy` is unable to determine if they satisfy the noisy time-series data.

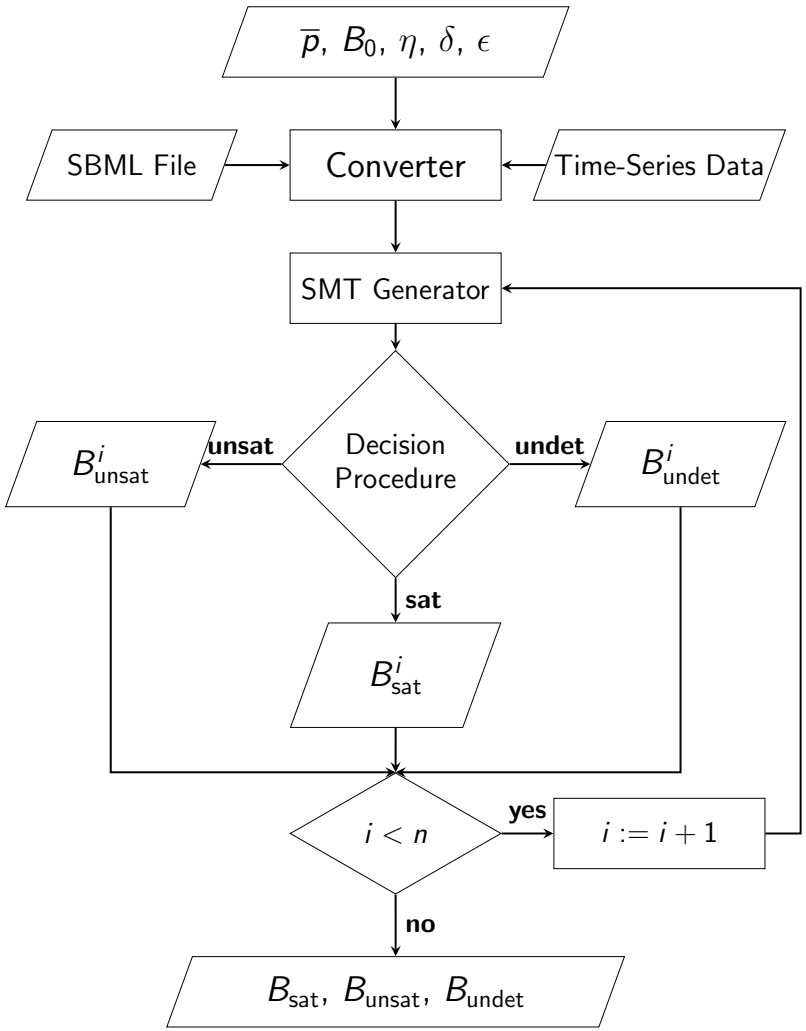
SMT Formulation

- The ODEs (derived from the SBML), parameters, and the noisy time-series data are converted into a collection of SMT problems.
- Each problem represents an *initial value problem* (IVP) constrained by the initial time point and one of the subsequent time points.
- Each individual SMT problem contains assertions that constrain:
  - The solution of each ODE to be contained within the interval found in the noisy time-series data; and
  - The parameters being synthesised to be within the synthesised ranges from the previous time point.

Splitting Boxes

- Initial boxes for the parameter set are passed one-by-one as SMT problems to a decision procedure utilising the SMT solver, `dReal`.
- The synthesis algorithm iteratively splits each box until a minimum size is reached or the current box is either **unsat** or **sat**.
- This process continues incrementally until all the points in the time series are processed.

Algorithm



- δ can be arbitrarily small but must be positive since solving first-order real formulae with general nonlinear functions is an undecidable problem.
- ϵ can be arbitrarily small as well, must also be positive, and determines the level of granularity used to search the parameter space.
- A small η makes it more difficult to identify acceptable ranges but leads to better compliance with the time-series data.

Figure: $\overline{p} = \{p_1, \ldots p_m\}$: model parameters to synthesise, B_0 = initial set of parameter ranges, η = acceptable noise, δ = SMT solver precision, ϵ = precision of parameter synthesis.

Results

- `BioPSy` has been successfully used to synthesise parameter ranges for three models:
  - A personalized prostate cancer therapy model;
  - A cell cycle model; and
  - A human starvation model.

Personalized Prostate Cancer Therapy Model

$$\frac{dx}{dt} = \left(\frac{\alpha_x}{1 + e^{(k_1 - z)k_2}} - \frac{\beta_x}{1 + e^{(z - k_3)k_4}} - m_1 \left(1 - \frac{z}{z_0}\right) - c_1 \right) x + c_2$$

$$\frac{dy}{dt} = m_1 \left(1 - \frac{z}{z_0}\right) x + \left(\alpha_y \left(1 - d_0 \frac{z}{z_0}\right) - \beta_y \right) y$$

$$\frac{dz}{dt} = -z\gamma - c_3$$

$$v = x + y$$

- v - *prostate specific antigen* (PSA)
- x - *hormone sensitive cells* (HSCs)
- y - *castration resistant cells* (CRCs)
- z - androgen

Ideta, A.M., Tanaka, G., Takeuchi, T., Aihara, K.: A mathematical model of intermittent androgen suppression for prostate cancer. Journal of Nonlinear Science 18(6), 593–614 (2008)

Personalized Prostate Cancer Therapy

- A prostate cancer patient was on treatment for 5 nonconsecutive times throughout 6 years and monitored every month (such as PSA and androgen levels were documented).
- Every period of time-series data contains around 4-5 time points.
- Parameter synthesis was performed using real clinical data[1].
- For each time-series, we synthesise $\alpha_y \times \beta_x \in [0.0, 0.05] \times [0.0, 0.05]$ with:
  - Tolerable amount of noise, $\eta = 1.4$;
  - Parameter synthesis precision, $\epsilon = 10^{-3}$; and
  - SMT solver precision, $\delta = 10^{-3}$.

[1] `http://www.nicholasbruchovsky.com/clinicalResearch.html`

Parameter Synthesis

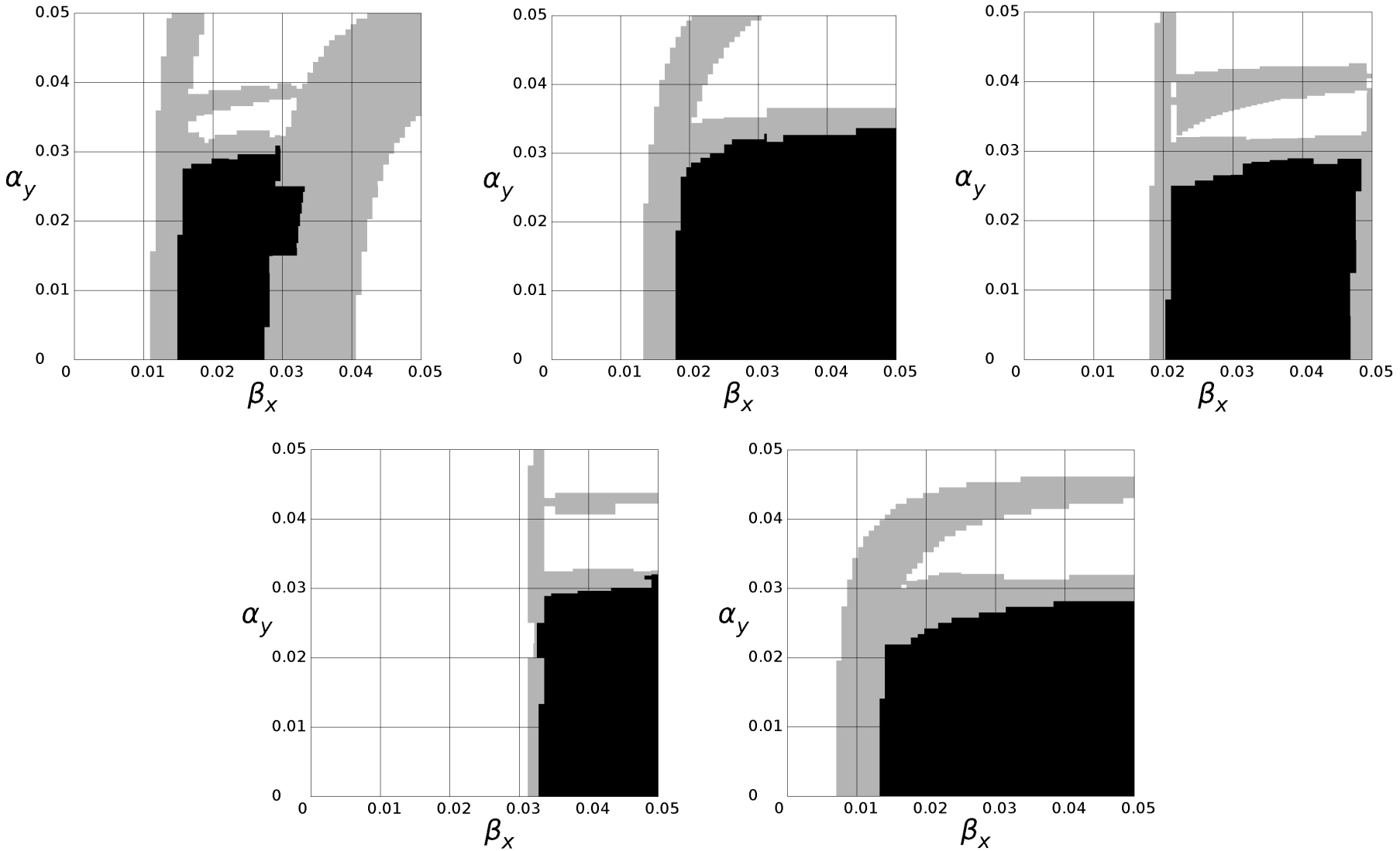


Figure: **white** - infeasible boxes; **black** - feasible boxes; and **gray** - undetermined boxes. **Runtime**: 12 hours for set of time-series data.

Parameter Synthesis Combined Result

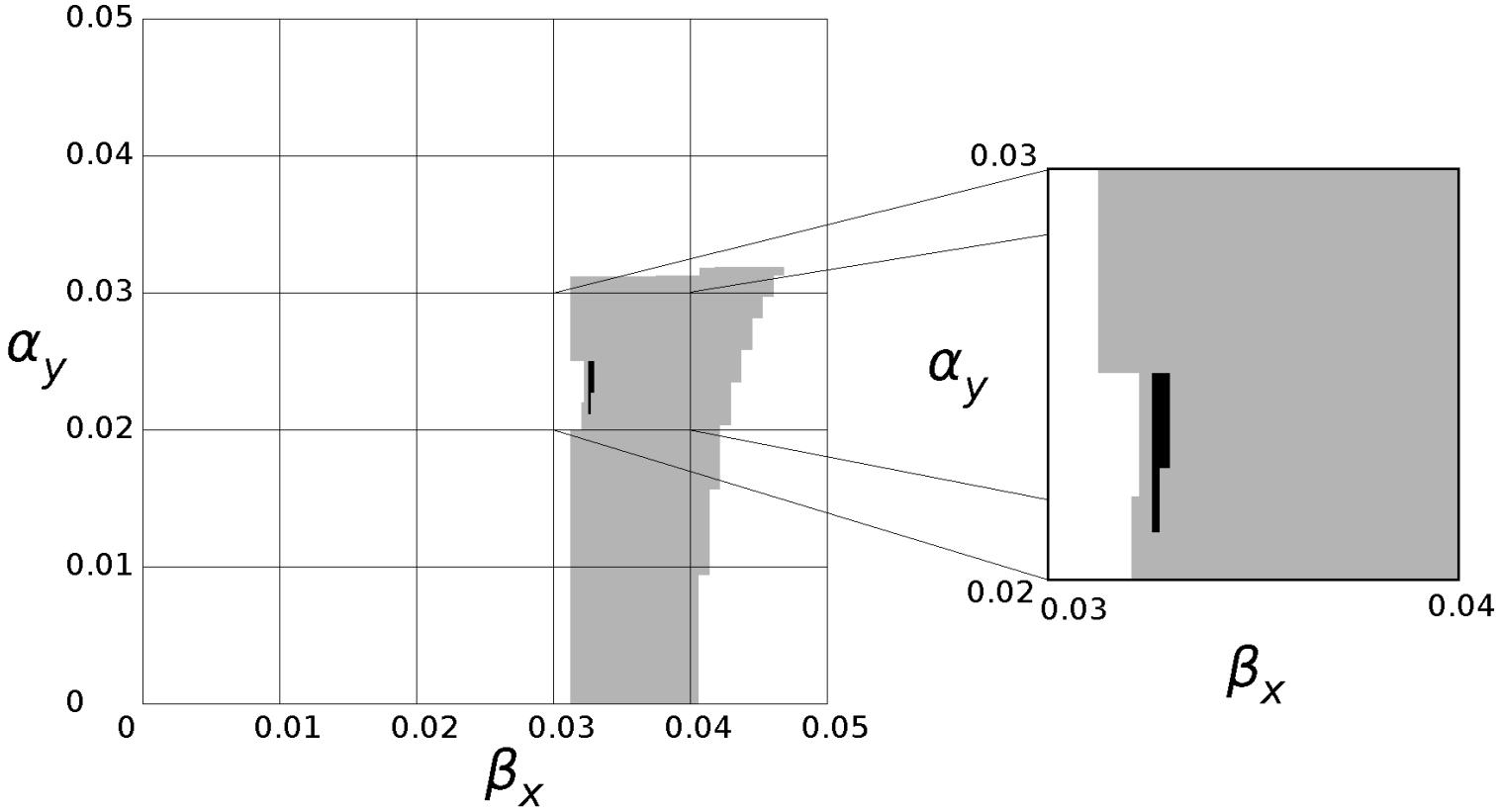


Figure: **white** - infeasible boxes; **black** - feasible boxes; and **gray** - undetermined boxes.

- The feasible set: $\alpha_y \times \beta_x \in [0.0225, 0.025] \times [0.0325, 0.0332031] \bigcup [0.0210938, 0.0225] \times [0.0325, 0.0327344]$.

Parameter Checking

- Checking parameter values is easier than synthesising parameter sets.
- Parameter values in this study were obtained using `COPASI` and verified with `BioPSy`.

Method	α_x	α_y	β_x	β_y	`BioPSy`				
					S_1	S_2	S_3	S_4	S_5
Evolut. Prog.	-0.216	-2.68×10^{-6}	0.0272	0.000135	n	y	y	n	y
Hooke & Jeeves	-0.309	-0.279	0.029	-0.24	y	y	y	y	y
Levenberg-Marquardt	-0.17	-32.0	0.00661	-10.5	n	n	n	n	n
Praxis	-0.233	-0.00698	0.0240	0.187	y	y	y	n	y
Scatter Search	-0.17	-31.9	0.00661	-10.5	n	n	n	n	n
Simulated Annealing	-0.249	6.4×10^{149}	0.0227	-2.27×10^{148}	n	n	n	n	n
Truncated Newton	-0.236	-0.00792	0.0244	0.0116	y	y	y	n	y

Cell Cycle

- In this model, two proteins, CDC2 and Cyclin, combine to form a heterodimer.
- The heterodimer controls major events in a cell causing it to:
  - Reach a steady state;
  - Act as a spontaneous oscillator; or
  - Act as an excitable switch.
- The cell cycle model has reference BIOMD0000000006 in the BioModels Database[2].
- The simulated time-series data contains 10 time points.

Tyson, J.J.: Modeling the cell division cycle: cdc2 and cyclin interactions. Proceedings of the National Academy of Sciences 88(16), 7328–7332 (1991)

[2] `https://www.ebi.ac.uk/biomodels-main/`

Cell Cycle Model

- For this experiement, we synthesise $k_4' \times k_4 \in [0.01, 0.02] \times [175, 185]$ with:
  - Tolerable amount of noise, $\eta = 10^{-3}$;
  - Parameter synthesis precision, $\epsilon = 0.1$; and
  - SMT solver precision, $\delta = 10^{-3}$.

$$\frac{du}{dt} = k_4 (v - u) \left(\frac{k_4'}{k_4} + u^2 \right) - k_6 u$$
$$\frac{dv}{dt} = \kappa - k_6 u$$

- u - CDC2
- v - Cyclin

Parameter Synthesis

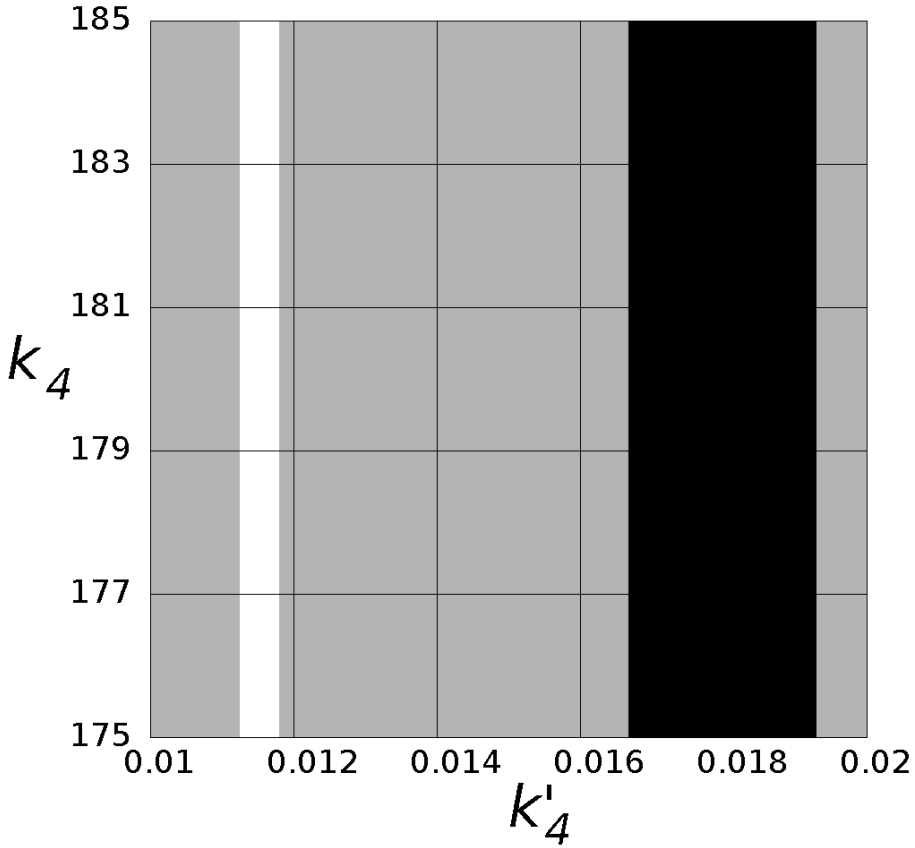


Figure: **white** - infeasible boxes; **black** - feasible boxes; and **gray** - undetermined boxes. **Runtime**: 10 minutes.

- The feasible set: $k'_4 \times k_4 \in [0.0166691, 0.0192934] \times [175, 185]$.

Human Starvation

- The human starvation model tracks the amount of fat, protein in muscle mass, and ketone bodies in the human body after glucose reserves have been depleted.
- For this experiement, we synthesise $\kappa \times b \in [9, 11] \times [0.05, 0.08]$ with:
  - Tolerable amount of noise, $\eta = 0.1$;
  - Parameter synthesis precision, $\epsilon = 0.1$; and
  - SMT solver precision, $\delta = 10^{-3}$.
- The simulated time-series data contains 25 time points.

$$(\text{Fat})\frac{dF}{dt} = F\left(\frac{-a}{1+K} - \frac{1}{\lambda_F}\left(\frac{C + gL_0}{F+M} + \kappa\right)\right)$$

$$(\text{Protein})\frac{dM}{dt} = -\frac{M}{\lambda_M}\left(\frac{C + \kappa L_0}{F+M} + \kappa\right)$$

$$(\text{Ketone Bodies})\frac{dK}{dt} = \frac{VaF}{1+K} - b$$

Song, B., Thomas, D.: Dynamics of starvation in humans. Journal of Mathematical Biology 54(1), 27–43 (2007)

Parameter Synthesis

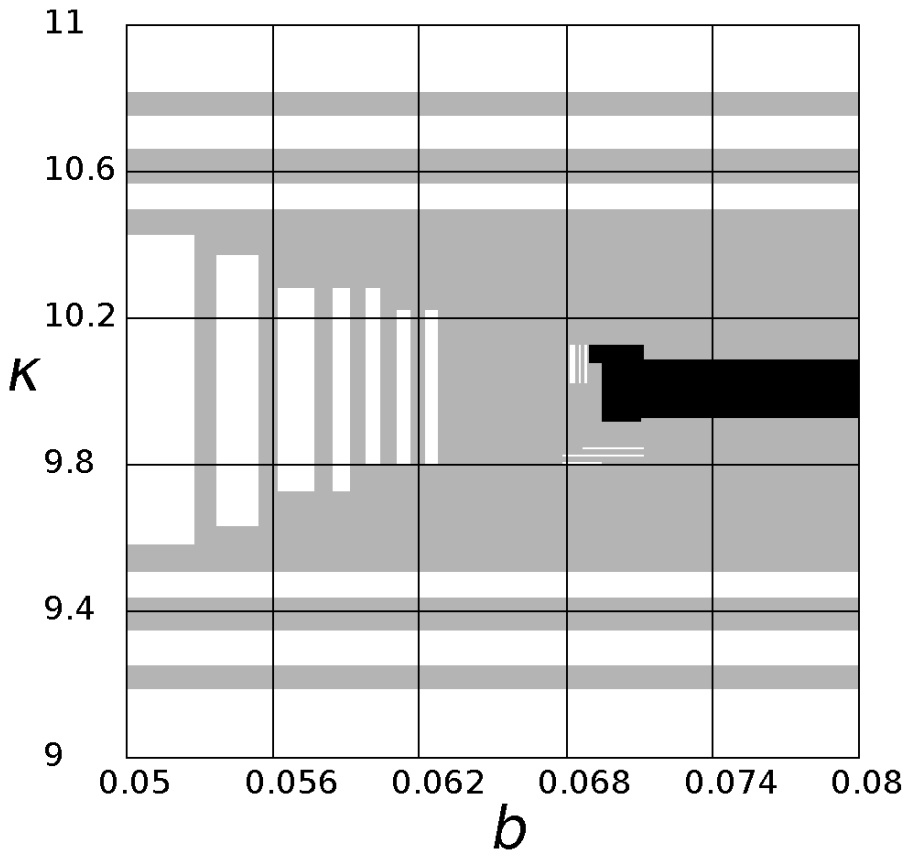


Figure: **white** - infeasible boxes; **black** - feasible boxes; and **gray** - undetermined boxes. **Runtime**: 5 minutes.

- The feasible set: $\kappa \times b \in [9.88077, 9.8832] \times [0.0764844, 0.0771875] \bigcup [9.92213, 10] \times [0.0785938, 0.08] \bigcup \ldots$.

BioPSy Demo Video

Conclusions

- `BioPSy` accepts SBML models, so it can be applied to a large number of existing biological models.
- The tool is not only limited to biological models with mass action kinetics but can handle models involving general ODEs.
- Models using parameters synthesised with `BioPSy` are formally guaranteed to behave as desired.
- Also, parameter estimates generated by other methods can be formally validated with `BioPSy`.
- `BioPSy` is freely available for download from: `https://github.com/dreal/biology`.

Future Work

- We hope to improve `BioPSy` to suggest how to proceed with undecidable ranges:
  - The tool could automatically synthesise the parameters again with a finer level of precision; and/or
  - Undecidable ranges that lie right next to each other could be automatically combined into a larger interval and re-analysed.
- Additionally, we plan to extend `BioPSy` to handle biological models that contain both continuous and discrete dynamics.

Thank you

Questions?