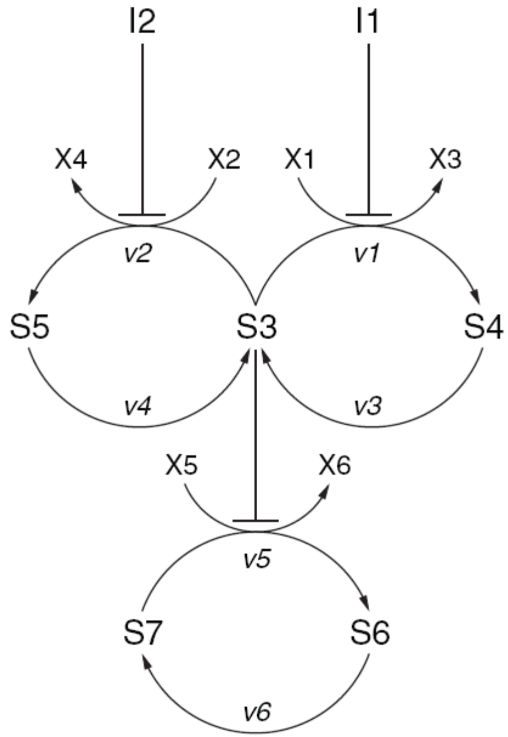
Improved parameter estimation for completely observed ordinary differential equations with application to biological systems

Peter Gennemark
University of Göteborg
Uppsala University

Dag Wedelin
Chalmers University of Technology

An Ordinary Differential Equation (ODE) model



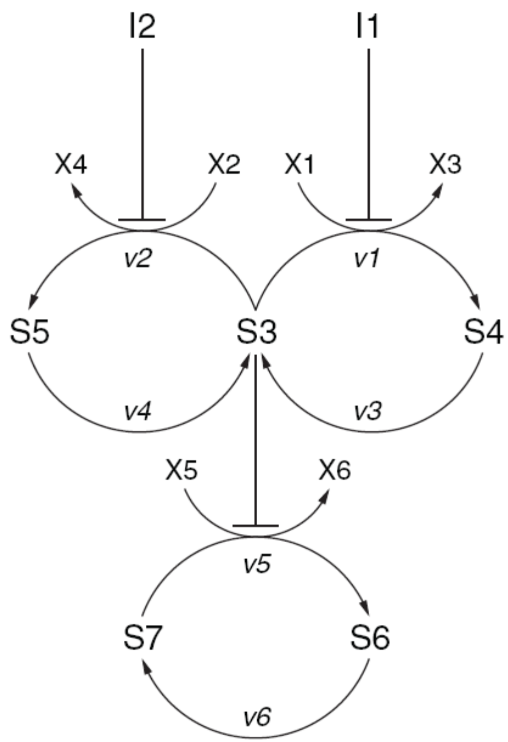
$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$

$$S_4'(t) = v_1 - v_3$$

$$S_5'(t) = v_2 - v_4$$

$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

An Ordinary Differential Equation (ODE) model



$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$
$$S_4'(t) = v_1 - v_3$$
$$S_5'(t) = v_2 - v_4$$
$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

$$v_1 = \frac{S_3(t)V_{\max1}}{(S_3(t) + K_{D1})(1 + I_1(t)/K_{I1})}$$

$$v_2 = \frac{S_3(t)V_{\max2}}{(S_3(t) + K_{D2})(1 + I_2(t)/K_{I2})}$$

$$v_3 = \frac{S_4(t)V_{\max3}}{S_4(t) + K_{D3}}$$

$$v_4 = \frac{S_5(t)V_{\max4}}{S_5(t) + K_{D4}}$$

$$v_5 = \frac{S_7(t)V_{\max5}}{(S_7(t) + K_{D5})(1 + S_3(t)/K_{I3})}$$

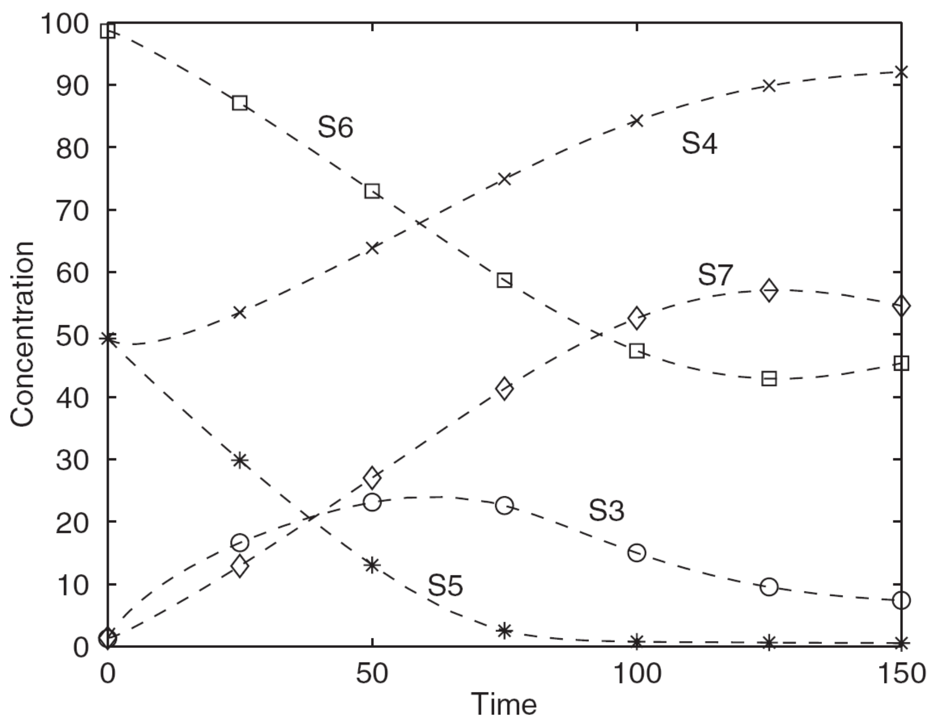
$$v_6 = \frac{S_6(t)V_{\max6}}{S_6(t) + K_{D6}}$$

Input to parameter estimation

ERROR FUNCTION

$$L(\hat{X}_j|\,\mathbf{k}) = -\frac{1}{2}\sum_i \left(\frac{X_j(t_i) - \hat{X}_j(t_i)}{\sigma_j(t_i)}\right)^2$$

DATA



MODEL STRUCTURE

$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$

$$S_4'(t) = v_1 - v_3$$

$$S_5'(t) = v_2 - v_4$$

$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

$$v_1 = \frac{S_3(t)V_{\max 1}}{(S_3(t) + K_{D1})(1 + I_1(t)/K_{I1})}$$

$$v_2 = \frac{S_3(t)V_{\max 2}}{(S_3(t) + K_{D2})(1 + I_2(t)/K_{I2})}$$

$$v_3 = \frac{S_4(t)V_{\max 3}}{S_4(t) + K_{D3}}$$

$$v_4 = \frac{S_5(t)V_{\max 4}}{S_5(t) + K_{D4}}$$

$$v_5 = \frac{S_7(t)V_{\max 5}}{(S_7(t) + K_{D5})(1 + S_3(t)/K_{I3})}$$

$$v_6 = \frac{S_6(t)V_{\max 6}}{S_6(t) + K_{D6}}$$

Basic algorithm

1. Try a parameter set

2. Simulate the model and calculate the error

3. Update the parameters according to some rule and then repeat from 2 until termination according to some criterion, e.g. that the error is sufficently stable

Slow, since a full simulation is required every time the error is evaluated.

Why study the parameter estimation problem assuming completely observed systems?

Important subproblem for structure identification methods, where also the form of the ODEs is unknown. Complete data is more reasonable here.

Significantly faster algorithms on the base case are always of general interest.

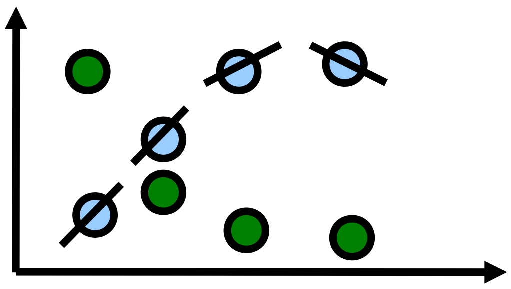
Approximations based on decomposition

$X'(t) = k_1 Y(t) – k_2 X(t)$

Approximations based on decomposition

$X'(t) = k_1\ Y(t) - k_2\ X(t)$

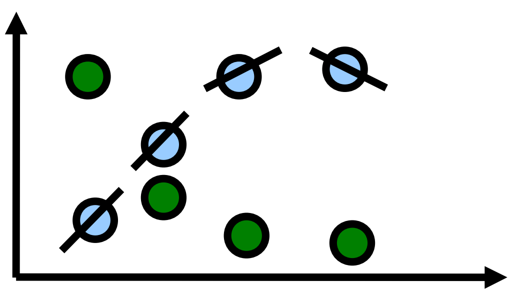
Estimate for each sampling point



Approximations based on decomposition

$X'(t) = k_1 Y(t) - k_2 X(t)$

Estimate for each sampling point



$X'(t_1) = k_1 Y(t_1) - k_2 X(t_1)$

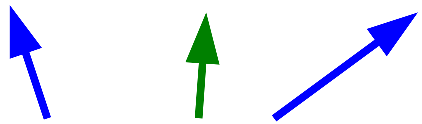
$X'(t_2) = k_1 Y(t_2) - k_2 X(t_2)$

:

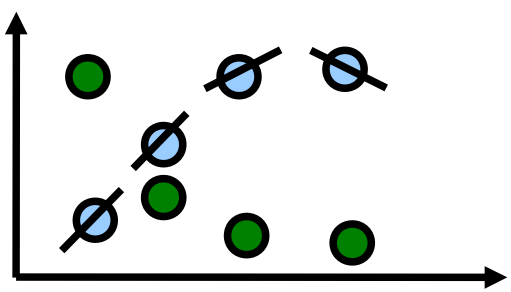
$X'(t_n) = k_1 Y(t_n) - k_2 X(t_n)$

Approximations based on decomposition

$X'(t) = k_1\, Y(t) - k_2\, X(t)$



Estimate for each
sampling point



$X'(t_1) = k_1\, Y(t_1) - k_2\, X(t_1)$
$X'(t_2) = k_1\, Y(t_2) - k_2\, X(t_2)$
:
$X'(t_n) = k_1\, Y(t_n) - k_2\, X(t_n)$

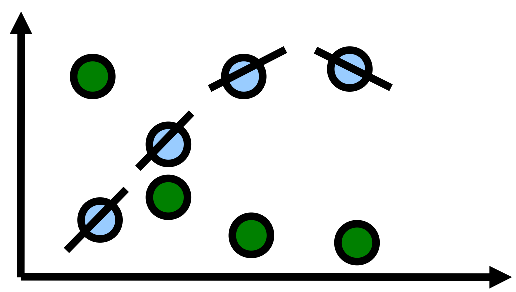
Derivative method

Rough and fast

Approximations based on decomposition

$X'(t) = k_1\, Y(t) - k_2\, X(t)$

Estimate for each sampling point



$X'(t_1) = k_1\, Y(t_1) - k_2\, X(t_1)$

$X'(t_2) = k_1\, Y(t_2) - k_2\, X(t_2)$

:

$X'(t_n) = k_1\, Y(t_n) - k_2\, X(t_n)$

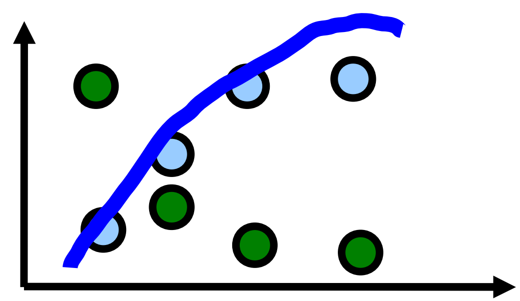
Derivative method

Rough and fast

$X'(t) = k_1\, Y(t) - k_2\, X(t)$

Simulate one ODE

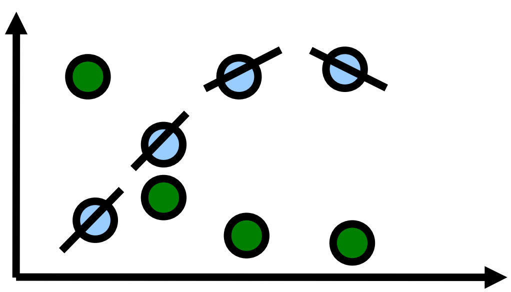
Estimate for each time point



Approximations based on decomposition

$X'(t) = k_1\, Y(t) - k_2\, X(t)$

Estimate for each sampling point



$X'(t_1) = k_1\, Y(t_1) - k_2\, X(t_1)$
$X'(t_2) = k_1\, Y(t_2) - k_2\, X(t_2)$
:
$X'(t_n) = k_1\, Y(t_n) - k_2\, X(t_n)$

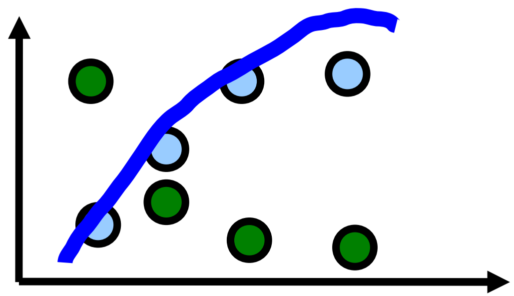
Derivative method

Rough and fast

$X'(t) = k_1\, Y(t) - k_2\, X(t)$

Simulate one ODE

Estimate for each time point



Estimation with basic parameter estimation algorithm in a single equation with simulation

Better precision, still quite fast

How to use the approximations: fast and rough methods initially, slow and accurate methods in the end

1. FOR each equation DO

    (a) Derivative method.

    (b) Estimation for single equation subproblems.

2. Estimation for the entire problem.

More accurate
Slower

Gennemark and Wedelin, *IET Systems Biology*, 1 (2) 2007

Our original parameter estimation algorithm

REPEAT

1. Select input for the other variables to be used in step 2:
Experimental data or simulated data.

2. FOR each equation DO

    (a) Derivative method.

    (b) Estimation for single equation subproblems.

3. Estimation for the entire problem.

UNTIL convergence

Gennemark and Wedelin, *IET Systems Biology*, 1 (2) 2007

Observed problems

If estimation for single equation sub-problems have low precision and estimation for the entire problem is too slow, it can be difficult to find a good mix between the two.

Observed problems

If estimation for single equation sub-problems have low precision and estimation for the entire problem is too slow, it can be difficult to find a good mix between the two.

When the same parameter occurs in two ODEs and this is not taken into account, data is not used optimally and may result in ambiguity and potentially divergent solutions.

$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$
$$S_4'(t) = v_1 - v_3$$
$$S_5'(t) = v_2 - v_4$$
$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

Generalisation of our approach

Why using either a single equation or all equations? We can use any subset!

Generalisation of our approach

Why using either a single equation or all equations? We can use any subset!

We propose two different automatic approaches to select sub-models:

Metabolic/signalling systems. For every ODE, a sub-model is defined by merging those other ODEs that have reactions (and thus also parameters) in common with this ODE.

$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$
$$S_4'(t) = v_1 - v_3$$
$$S_5'(t) = v_2 - v_4$$
$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

Generalisation of our approach

Why using either a single equation or all equations? We can use any subset!

We propose two different automatic approaches to select sub-models:

Metabolic/signalling systems. For every ODE, a sub-model is defined by merging those other ODEs that have reactions (and thus also parameters) in common with this ODE.

$$S_3'(t) = -v_1 - v_2 + v_3 + v_4$$
$$S_4'(t) = v_1 - v_3$$
$$S_5'(t) = v_2 - v_4$$
$$S_6'(t) = -S_7'(t) = v_5 - v_6$$

Genetic networks. The sub-model of each ODE includes the variables which occur on the right-hand side of the equation.

$X'(t) = k_1 - k_2 Z(t) - k_3 X(t)$

:

$Z'(t) = k_6 Y(t) - k_7 Z(t)$

Our modified parameter estimation algorithm

REPEAT

1. Select input for the other variables, to be used in step 2:
Experimental data or simulated data.

2. FOR each equation DO

   (a) Derivative method.

   (b) Estimation for single equation subproblems.

3. Estimation of sub-models defined around each single ODE.

4. Estimation for the entire problem.

UNTIL convergence

Results on benchmark problems

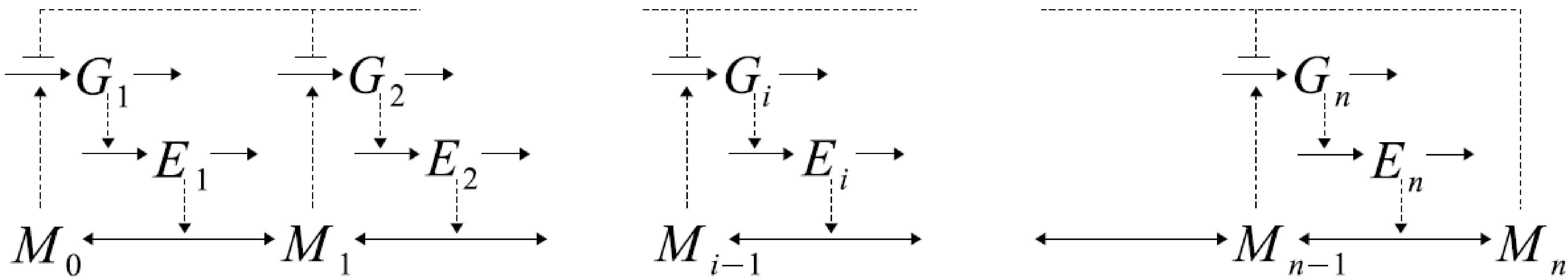
We evaluate our approach on several benchmark problems.

Comapred to our original approach: speed improvment about 2-fold.

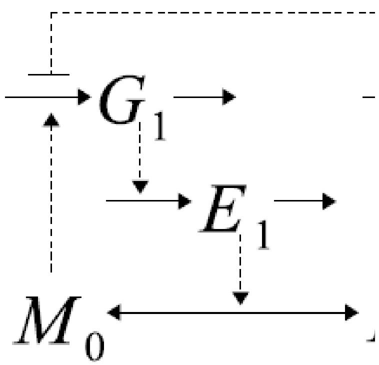
Significantly better than other methods on these problems.

Problem	Reference	Best solution		Performance	
		$-L$	Accuracy	Time (s)	Stability
pe_3genes1	External [1]		<16%	120000	
	External [12]		<0.02%	13000	
	Best external [13]		$<6\text{x}10^{-3}\%$	540	
	This work	0.00	$\approx 10^{-9}\%$	57	6/10
pe_3genes2f	Best external [12]			13000	
	This work	1286		90	5/10
pe_3genes3f	Best external [12]			13000	
	This work	3128		100	1/10
pe_pinene	Best external [13]	9.936		30	
	This work	9.936		0.1	10/10
pe_ss_cascade1	Best external [14]		<112%	660	
	This work	0.00	$\approx 10^{-9}\%$	13	10/10
pe_ss_branch4	Best external [15]	0.00		250	
	This work	0.00	$\approx 10^{-9}\%$	0.5	10/10
pe_ss_30genes2f	Best external [15]			1700	
	This work	3240		210	10/10

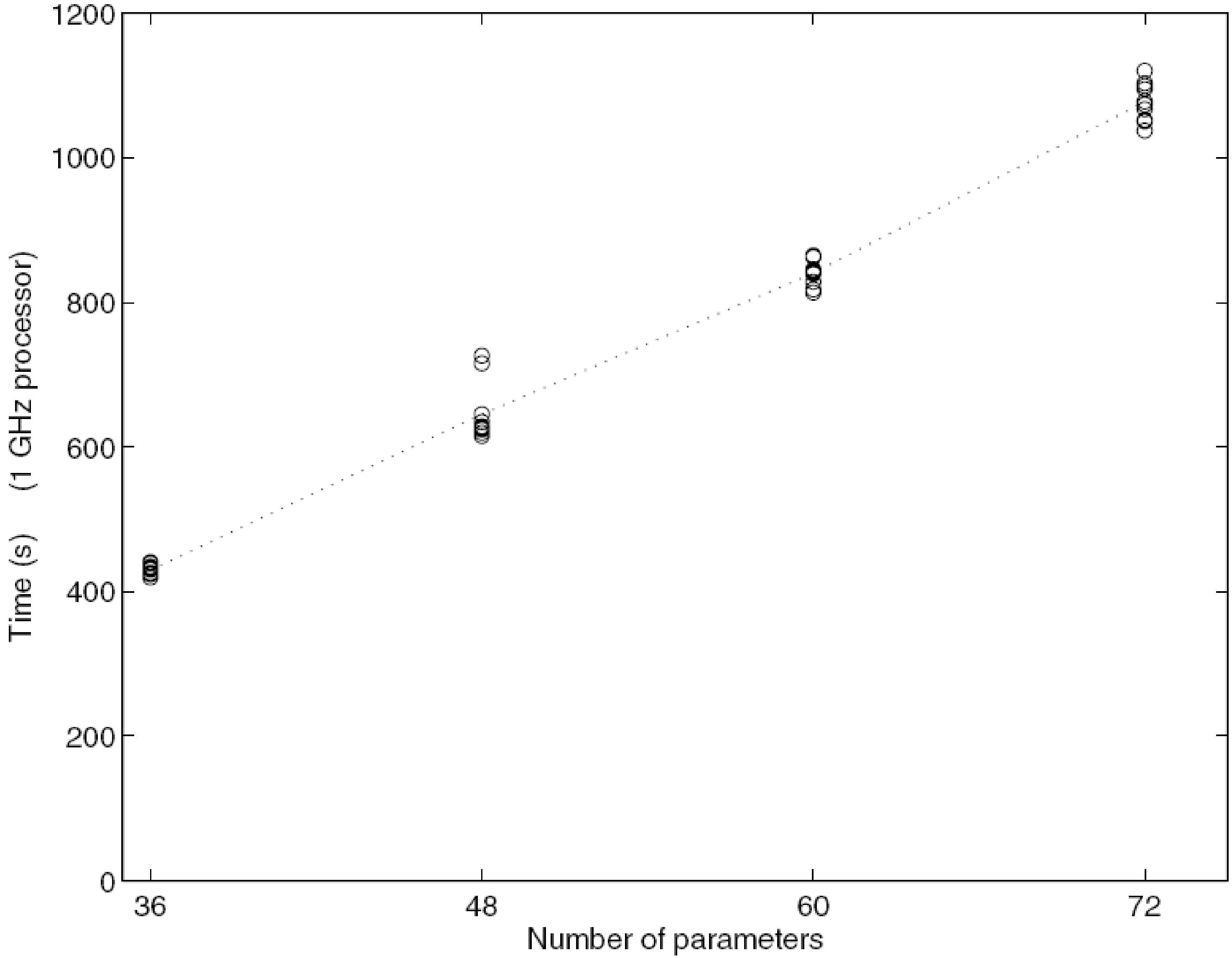
Results on benchmark problems



Results on benchmark problems



Attractive time complexity



ODE identification

All problems and our results are available here:

www.odeidentification.org

Gennemark and Wedelin, *Bioinformatics*, 25 (6), 2009

Thanks for your attention!