Approximation and uncertainty in models of biological systems

Paolo Milazzo

Dipartimento di Informatica, Università di Pisa, Italy

Torino – November 16, 2009

Introduction: Systems Biology

> *"Systems Biology is a comprehensive quantitative analysis of the manner in which all the components of a biological system interact functionally over time."*
>
> Alan Aderem, *Systems Biology: Its Practice and Challenges*. Cell 121, 511-513 (2005)

The aim of current research in Systems Biology is to integrate the knowledge about single constituents of living organisms into *system view*.

The two main approaches to biological systems modelling:

- **Biomath** Models are given as *differential equations* (or recurrence equations), and are studied by applying *analytical* and *numerical* techniques.
- **Bioinfo** Biological systems are modelled as *stochastic concurrent systems* and analyzed by *simulation* and *model checking*.

The application of such tools is limited to small, well known pathways

Introduction: The need of approximations

"Biological processes are profoundly complex, containing hundreds or thousands of component interactions. This leads to uncertainty i.e., precise information about probabilities, pathway structure, rate constants and similar parameters, is often unknown. Further, it is often impossible to assign precise point probabilities to each of the myriad constituents of an intricate biological pathway."

Iyengar M.S., McGuire M.F., *Imprecise and Qualitative Probability in Systems Biology*, ICSB, October 1-6, 2007

The two main problems in biological systems modelling are:

- complexity of the systems
- unavailability of (precise) kinetic parameters

Hence, the need of constructing approximated models

- by means (if possible) of conservative abstractions

Introduction: our approaches

We propose two approaches for the construction and analysis of models with approximations:

- Delay stochastic simulation
  - PhD thesis (in progress) by Giulio Caravagna
    Dipartimento di Informatica, Università di Pisa
- Probabilistic model checking with uncertainty on kinetic rates
  - PhD thesis (in progress) by Guido Scatena
    IMT Lucca Institute for Advanced Studies

Outline of the talk

Delays in models of biological systems

Delays may be used to model events for which the underlying dynamics

- either cannot be precisely observed
- or is too complex to be handled efficiently by analysis tools

A delay σ represents the time necessary for the underlying network of events to produce some result observable in the higher level model.

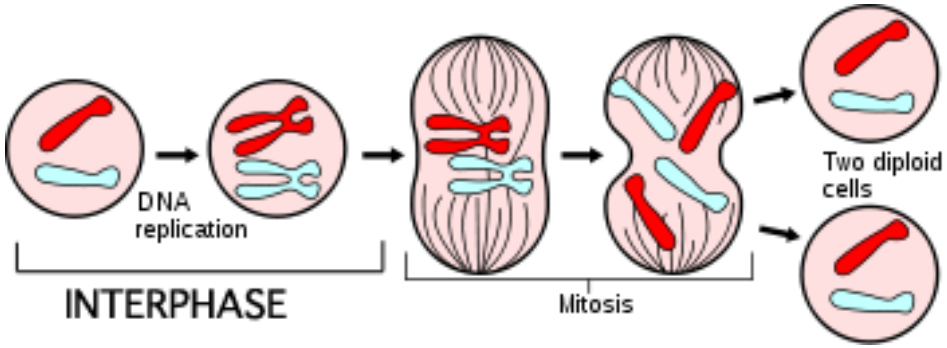
Mathematical modelling of biological systems with delays is mainly based on delay differential equations (DDEs)

- the derivative of the unknown function at time t is given in terms of the values of the function at time $t - \sigma$.

An example: Tumor growth (cell cycle)

Tumor growth is based on cell divisions (or *mitosis*).

The cell cycle, the process between two mitosis, consists of 4 phases :



- I : interphase
  - G_1: pre-synthetic phase
  - S : replication of DNA
  - G_2: post-synthetic phase
- M : mitosis phase

An example: Tumor growth (cell cycle)

We consider a DDE model of tumor growth proposed by *Villasana and Radunskaya*.

Tumor cells are classified in two populations:

- T_I: cells in the interphase (phases G_1, S and G_2);
- T_M: cells in the mitotic phase (M).

The model includes the following events:

1. cell death in any phase (apotosis)
2. interphase $\longrightarrow$ mitosis (one cell in T_I moves to T_M)
3. mitosis $\longrightarrow$ interphase (one cell in T_M becomes two in T_M)

The passage from interphase to mitosis takes much more time than the other events.

An example: Tumor growth (cell cycle)

The DDEs model by *Villasana and Radunskaya* is:

$$\frac{dT_I}{dt} = 2a_4 T_M - d_2 T_I - a_1 T_I(t-\sigma) \qquad T_I(t) = \phi_0(t) \text{ for } t \in [-\sigma, 0]$$

$$\frac{dT_M}{dt} = a_1 T_I(t-\sigma) - d_3 T_M - a_4 T_M \qquad T_M(t) = \phi_1(t) \text{ for } t \in [-\sigma, 0]$$

Let $d = d_3 + a_4$, namely d is the rate at which mitotic cells disappear.

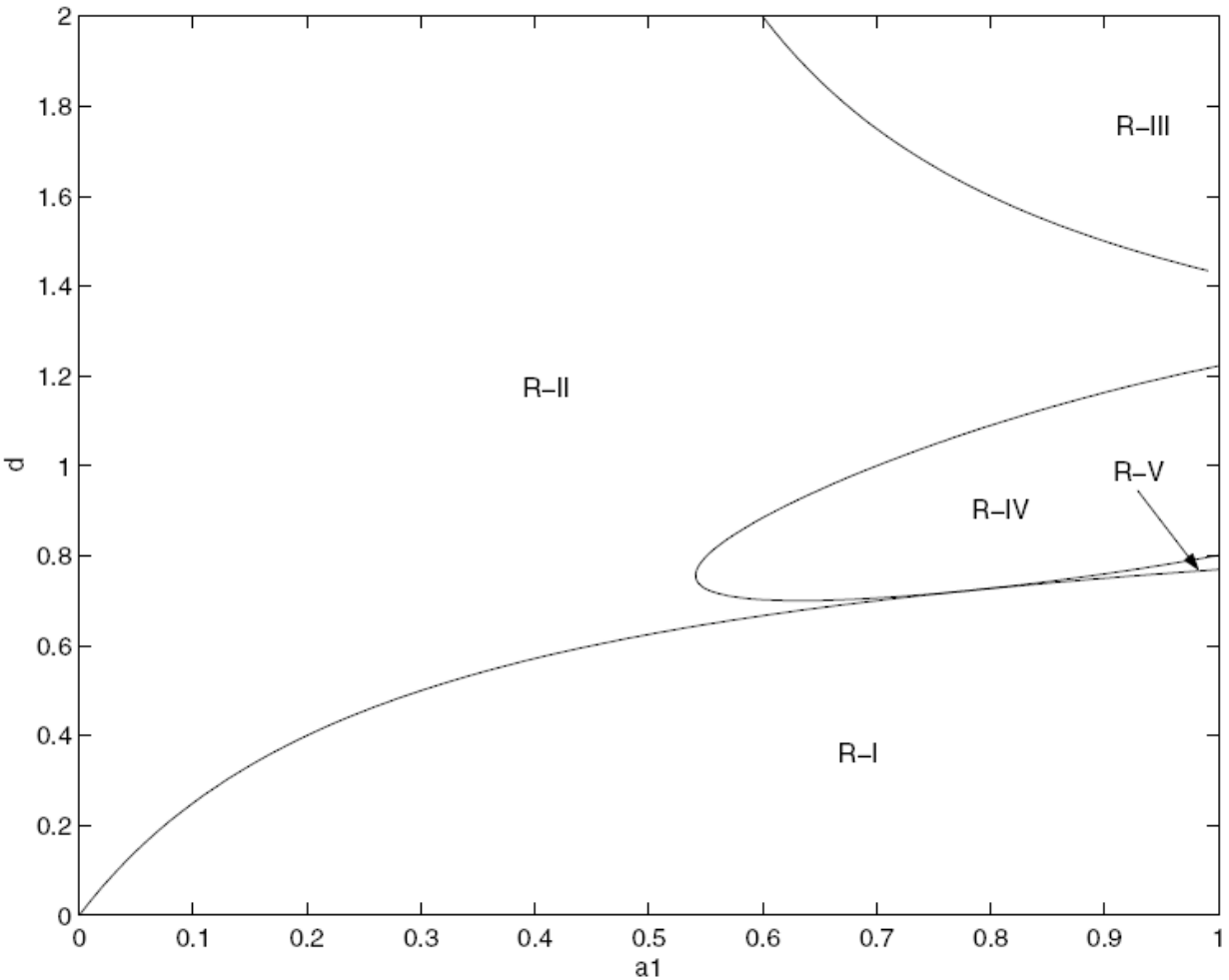
The number of cells that enter mitosis at time t depends on the number of cells that entered the interphase σ time units before, namely $T_I(t-\sigma)$.

This means that the interphase is associated with a duration σ (about one day in human cells).

In DDEs delays are modelled as dependencies form states of the system in the past.

An example: Tumor growth (cell cycle)

Analytical study by varying a_1 and d gives five parameter regions:



When $\sigma = 0$:

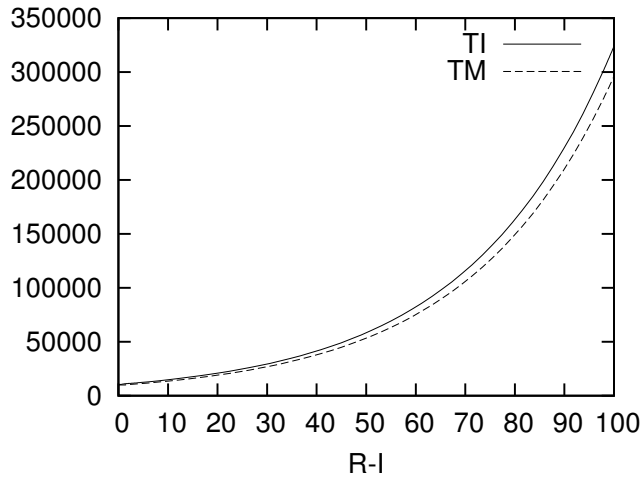
- In R-I the tumor grows
- In other regions the tumor decays

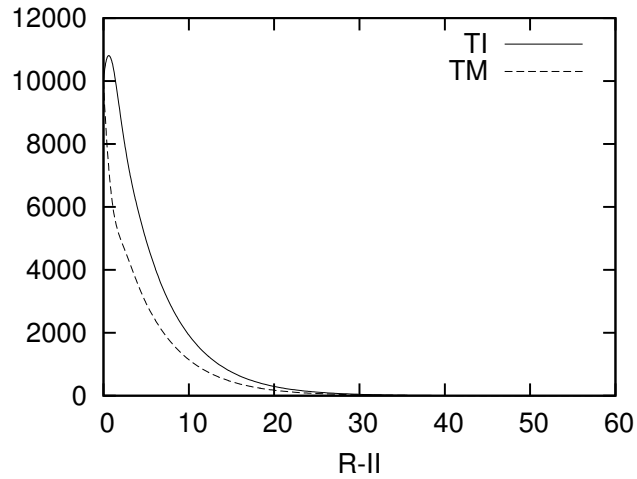
When $\sigma > 0$:

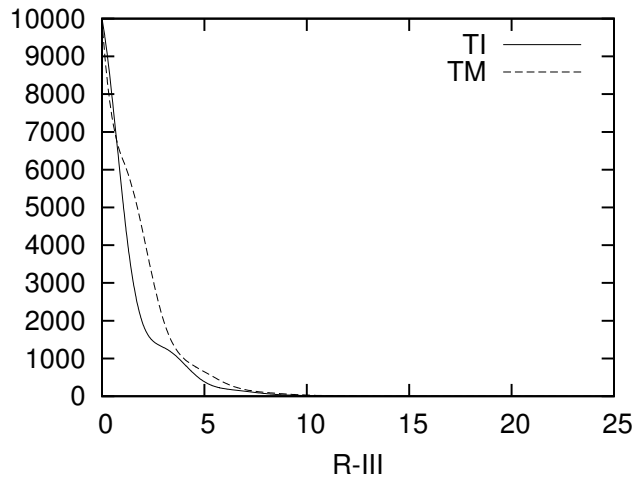
- In R-I the tumor grows
- In R-II the tumor decays
- In other regions the tumor size oscillates

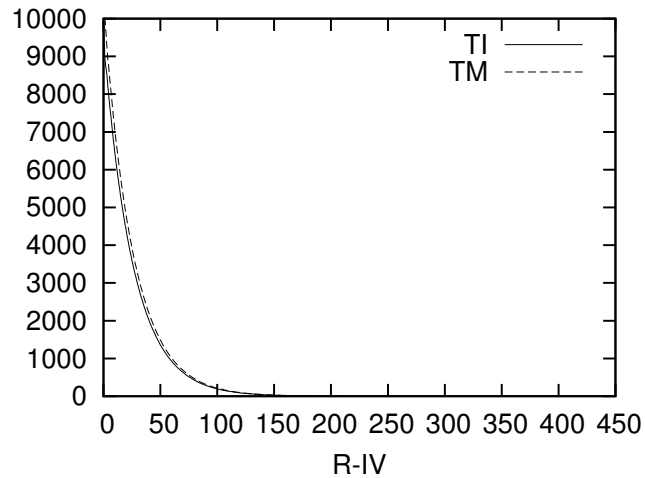
An example: Tumor growth (cell cycle)

These are some results of numerical simulation with $\sigma = 1$.



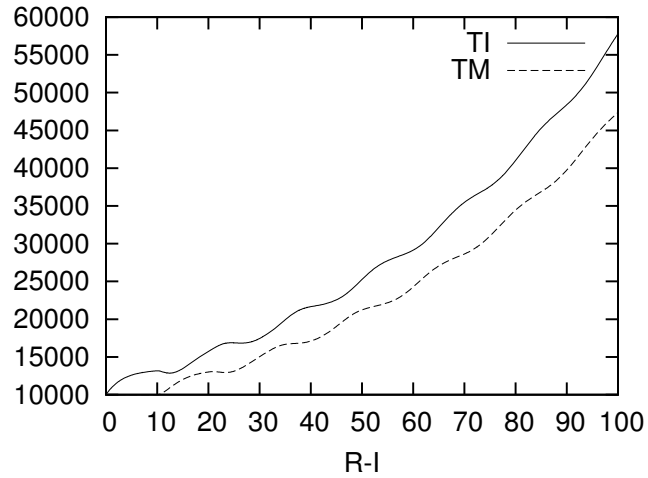


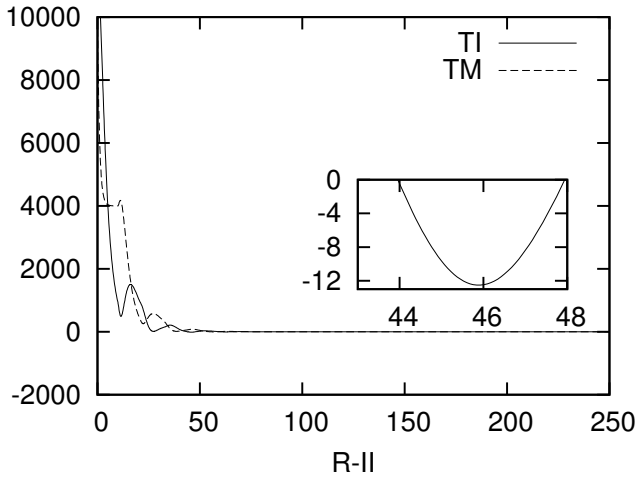


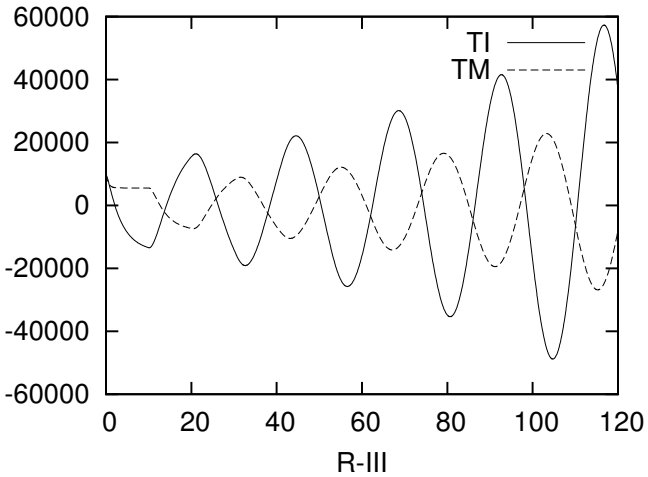


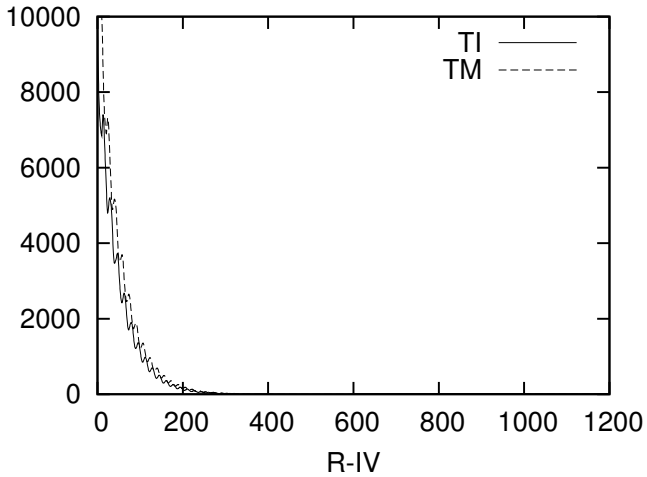
An example: Tumor growth (cell cycle)

These are some results of numerical simulation with $\sigma = 10$.









Stochastic simulation of chemical reactions (no delays)

Usual notation for chemical reactions:

$$\ell_1 S_1 + \ldots + \ell_\rho S_\rho \underset{k_{-1}}{\overset{k}{\rightleftharpoons}} \ell'_1 P_1 + \ldots + \ell'_\gamma P_\gamma$$

where:

- S_i, P_i are molecules (reactants)
- ℓ_i, ℓ'_i are stoichiometric coefficients
- k, k_{-1} are the kinetic constants

Stochastic simulation of chemical reactions (no delays)

Gillespie's stochastic simulation algorithm (SSA):

- represents a chemical solution as a multiset of molecules
- computes the reaction rate a_μ by multiplying the kinetic constant by the number of possible combinations of reactants

Example: chemical solution with X_1 molecules S_1 and X_2 molecules S_2

reaction $R_1 : S_1 + S_2 \xrightarrow{k_1} 2S_1$ rate $a_1 = \binom{X_1}{1}\binom{X_2}{1}k_1 = X_1 X_2 k_1$

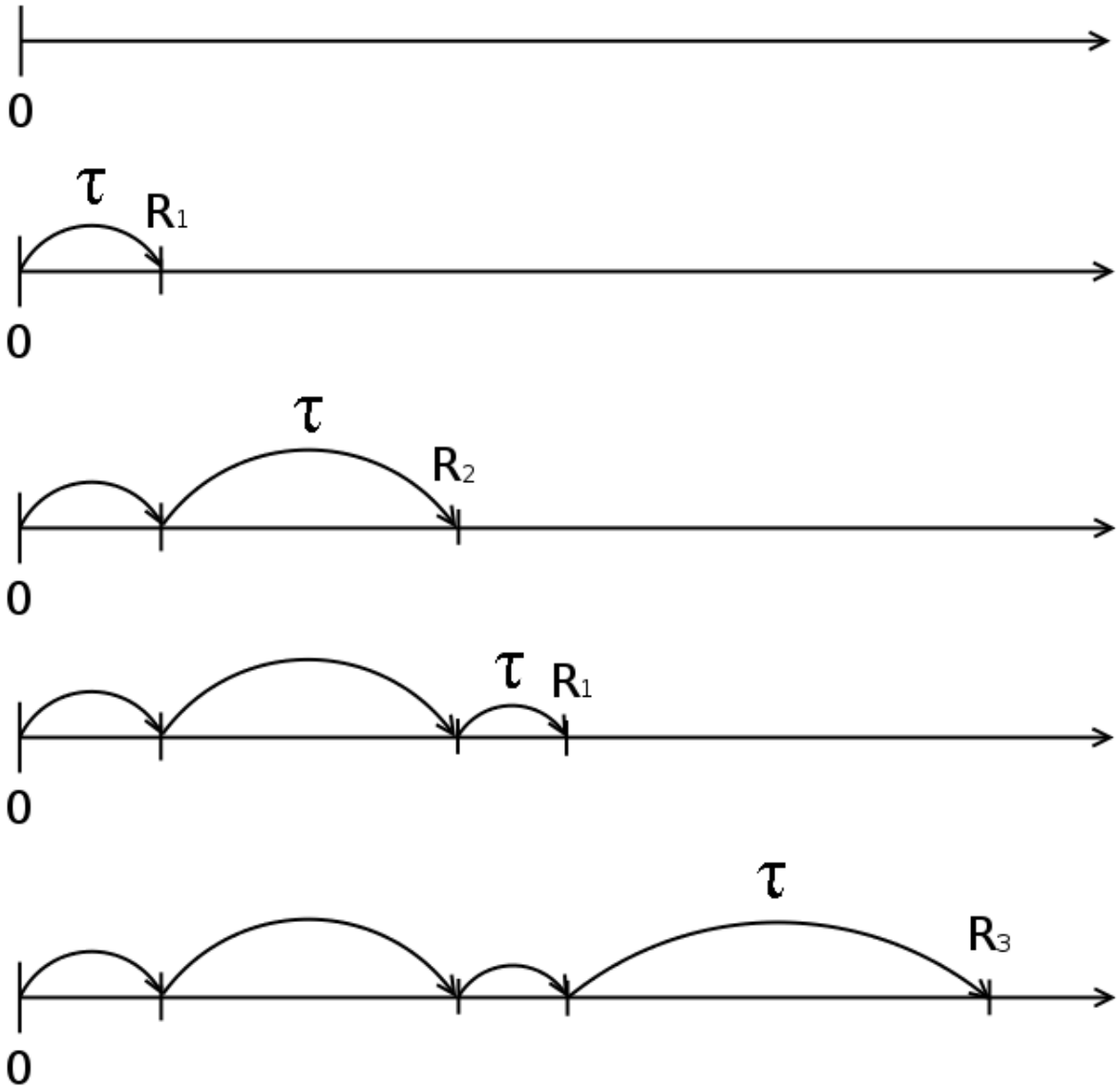
reaction $R_2 : 2S_1 \xrightarrow{k_2} S_1 + S_2$ rate $a_2 = \binom{X_1}{2}k_2 = \frac{X_1(X_1-1)}{2}k_2$

Given a set of reactions $\{R_1, \ldots R_M\}$ and a current time t

- The time $t + \tau$ at which the next reaction will occur is randomly chosen with τ exponentially distributed with parameter $\sum_{\nu=1}^{M} a_\nu$;
- The reaction R_μ that has to occur at time $t + \tau$ is randomly chosen with probability $\frac{a_\mu}{\sum_{\nu=1}^{M} a_\nu}$.

At each step t is incremented by τ and the chemical solution is updated.

Stochastic simulation of chemical reactions (no delays)



Delay stochastic simulation of chemical reactions

Algorithm proposed by Barrio et Al. in 2006.

Chemical reactions may be associated with delays: $S \xrightarrow{k,\sigma} P$

Similar to Gillespie's algorithm, but when a delayed reaction is chosen at time t:

- reactants are removed at time $t + \tau$
- products addition is scheduled for time $t + \tau + \sigma$

The delay σ is actually interpreted as a duration

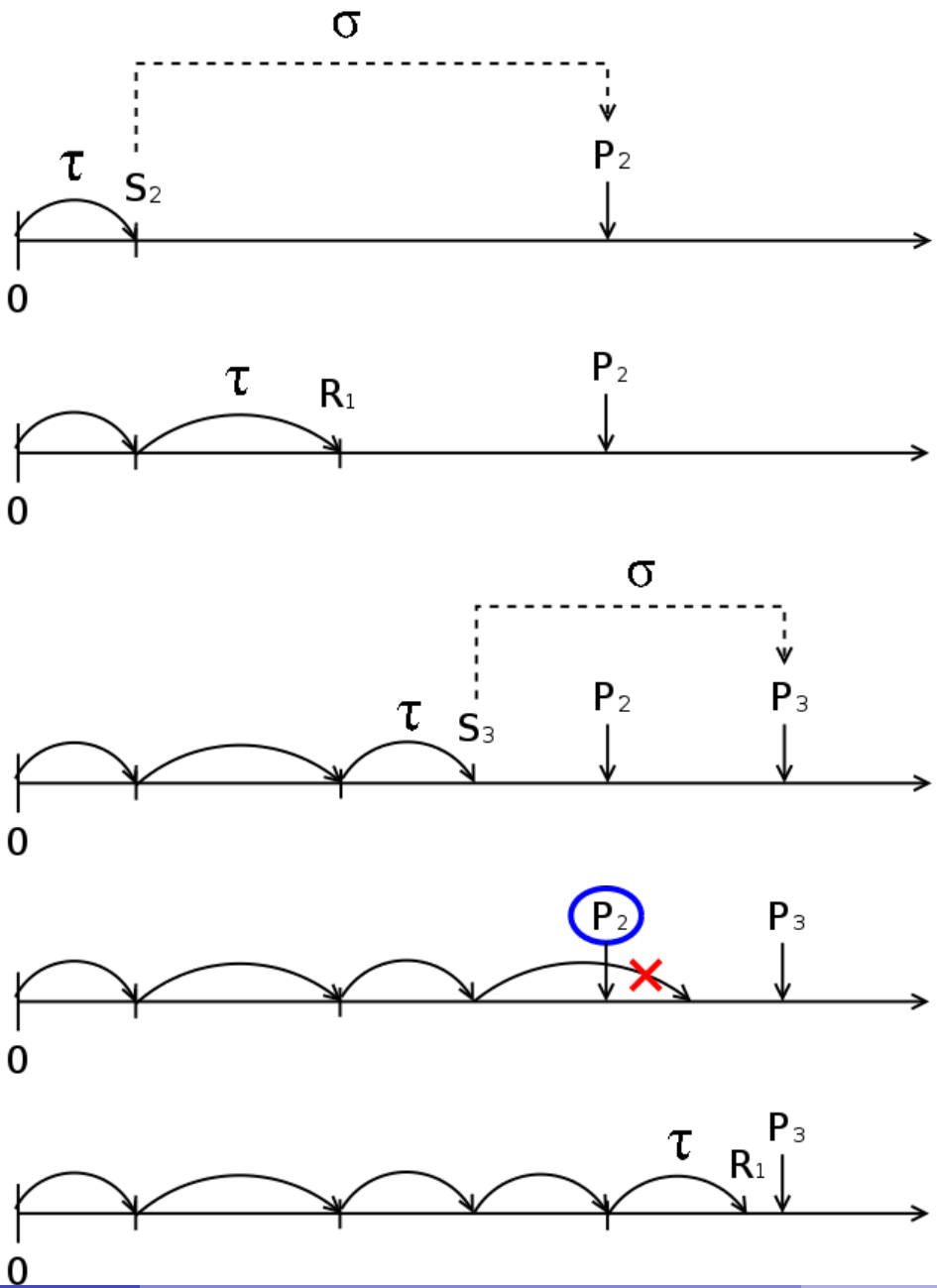
- different interpretation w.r.t. DDEs

Delay stochastic simulation of chemical reactions

Given a set of reactions $\{R_1, \ldots R_M\}$ and a current time t

- The time $t + \tau$ at which the next reaction will occur is randomly chosen with τ exponentially distributed with parameter $\sum_{\nu=1}^{M} a_\nu$;
- If there are no scheduled products additions in $[t, t + \tau]$:
  - Choose reaction R_μ with probability $\frac{a_\mu}{\sum_{\nu=1}^{M} a_\nu}$.
  - If R_μ is associated with a delay σ:
    - ★ remove the reactants and update t to $t + \tau$
    - ★ schedule products addition for $t + \tau + \sigma$
  - Otherwise, execute R_μ as in Gillespie's algorithm and update t to $t + \tau$;
- If there is a scheduled product addition at $t + \tau'$ with $\tau' \leq \tau$:
  - add the products and update t to $t + \tau'$

Delay stochastic simulation of chemical reactions



Delay Stochastic Model of Tumor growth

Let us reformulate the tumor growth example as a delay stochastic model.

Reactions:

$$T_M \xrightarrow{a_4} 2\,T_I \qquad\qquad T_I \xrightarrow{d_2}$$

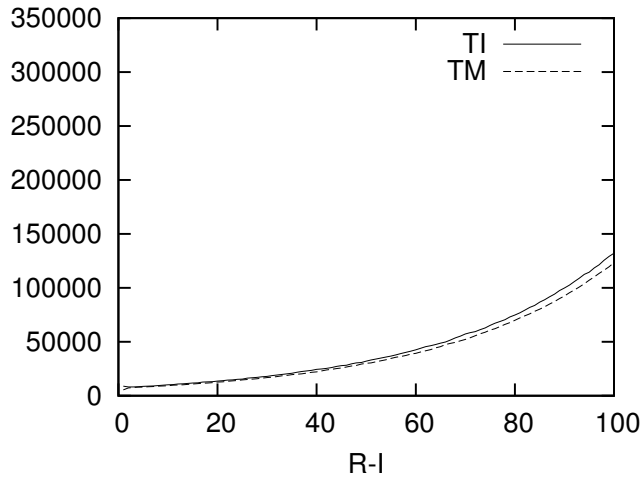
$$T_I \xrightarrow{a_1,\sigma} T_M \qquad\qquad T_M \xrightarrow{d_3}$$

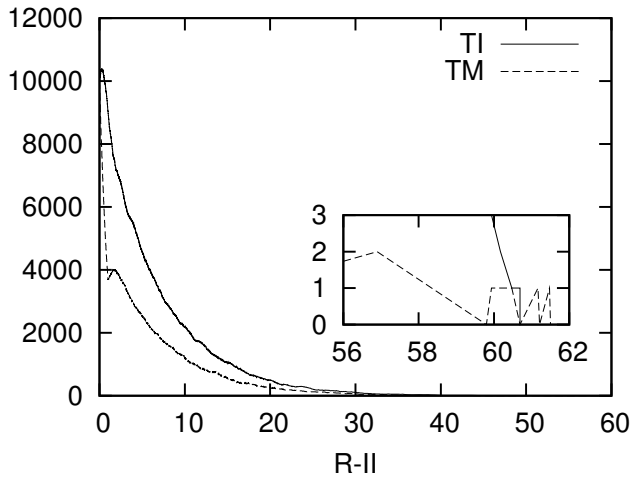
where σ is the duration of the interphase

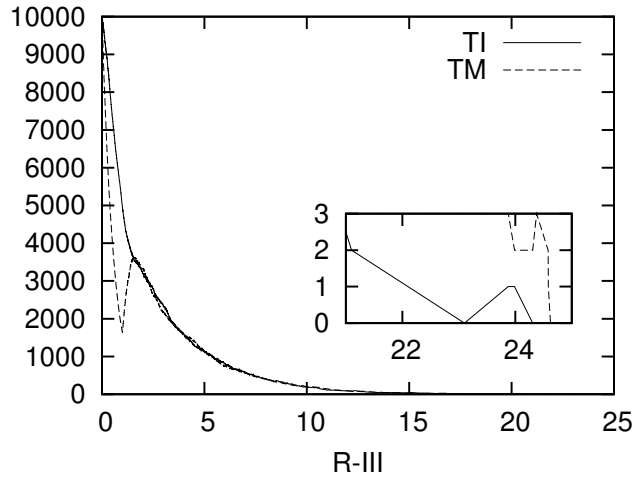
- we will consider $\sigma = 1$ and $\sigma = 10$ as before

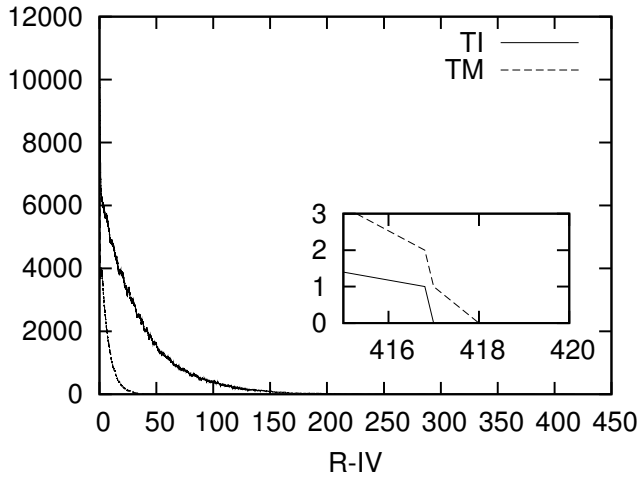
Delay Stochastic Model of Tumor growth

These are some results of delay stochastic simulation with $\sigma = 1$.



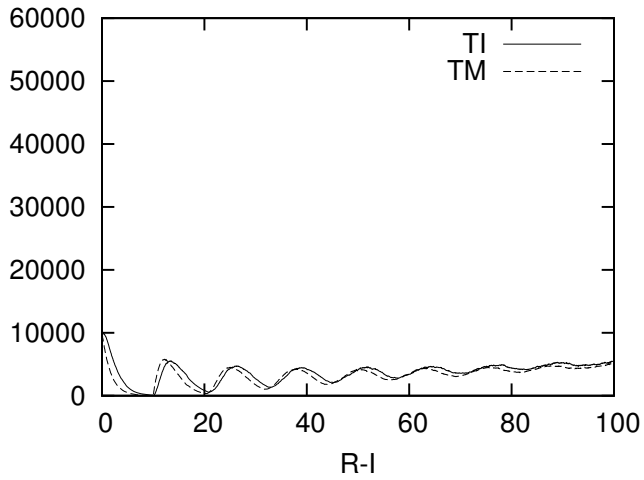


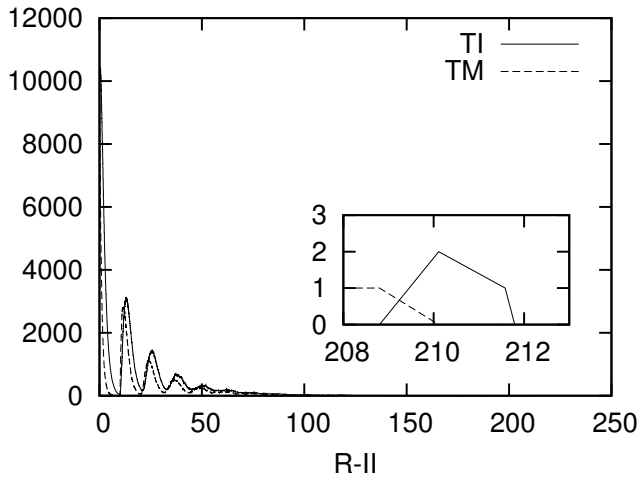


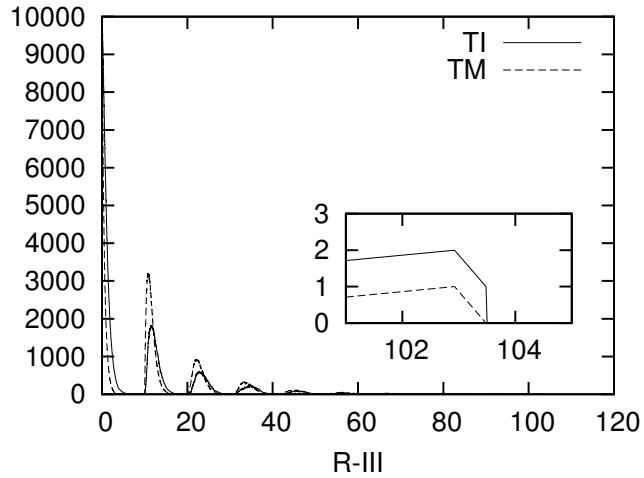


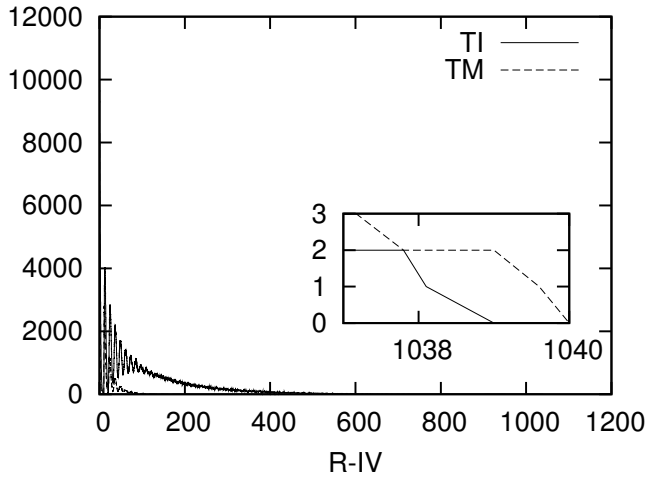
Delay Stochastic Model of Tumor growth

These are some results of delay stochastic simulation with $\sigma = 10$.









Delay Stochastic Model of Tumor growth

Stochastic simulation results are qualitatively similar to numerical simulation results:

- both approaches show tumor growth and eradication with similar parameters

But let us consider average tumor eradication times:

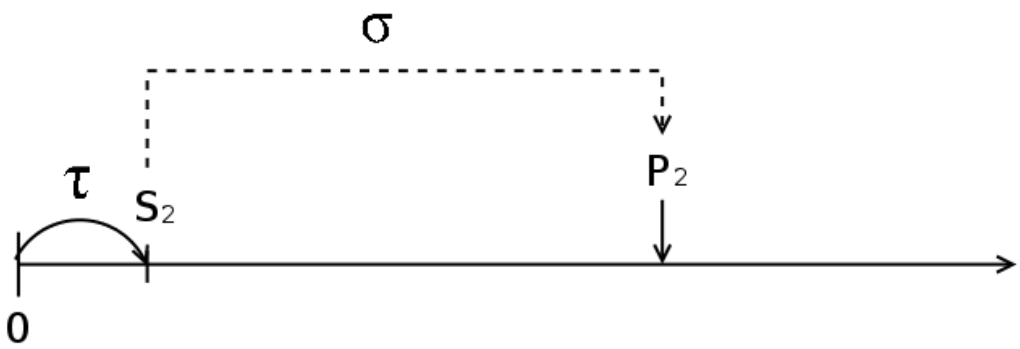
	DDE	DSSA
R-II with $\sigma = 1.0$	50	64
R-II with $\sigma = 10.0$	59	224
R-III with $\sigma = 1.0$	15	29
R-III with $\sigma = 10.0$	12	126
R-IV with $\sigma = 1.0$	238	302
R-IV with $\sigma = 10.0$	440	1072

In the delay stochastic model tumor eradication requires much more time...

Delay Stochastic Model of Tumor growth

Why this difference?

- in the delay stochastic model the tumor cell involved in a delayed reaction cannot die for σ time units!



This motivated us to develop a variant of the approach with a different interpretation of delays

- Delay as duration vs purely delayed

The Purely Delayed Approach

Chemical reactions may be associated with delays: $S \xrightarrow{k,\sigma} P$

Similar to Barrio's algorithm, but when a delayed reaction is chosen at time t:

- the simulation state is left unchanged
- the whole reaction is scheduled for time $t + \tau + \sigma$

The delay σ is actually interpreted as a delay

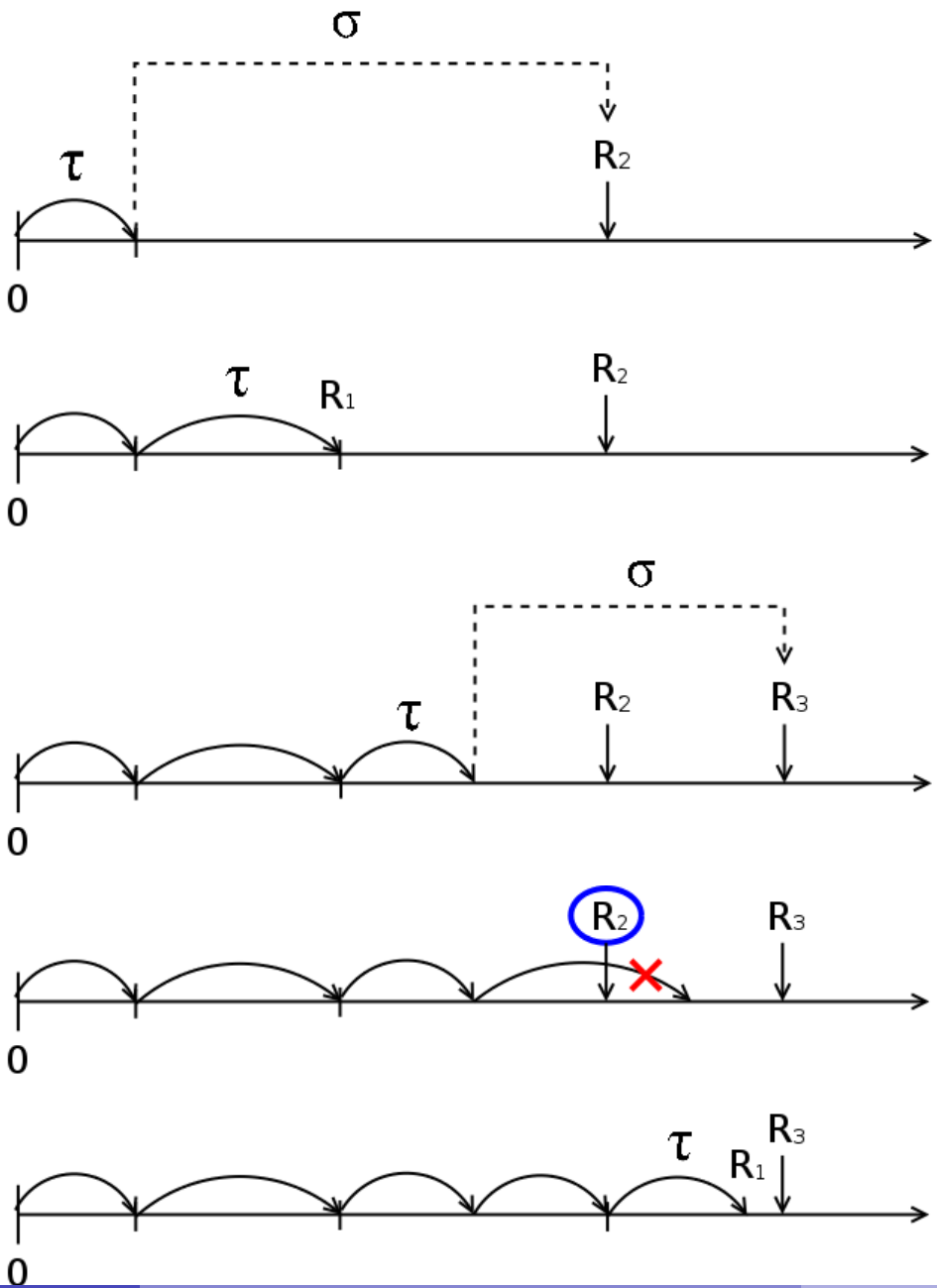
- interpretation more similar to that of DDEs

The Purely Delayed Approach

Given a set of reactions $\{R_1, \ldots R_M\}$ and a current time t

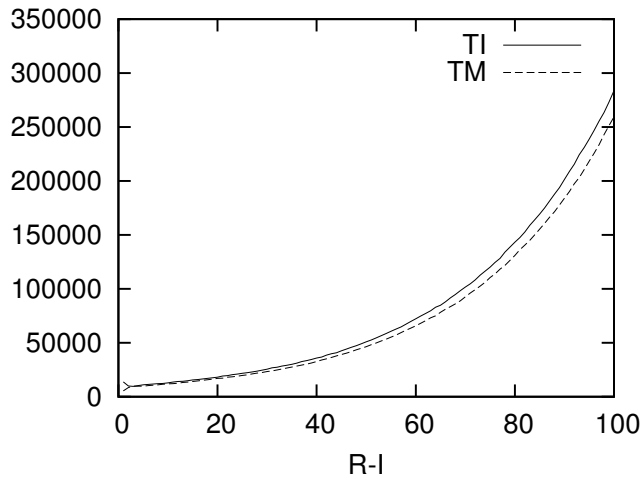
- The time $t + \tau$ at which the next reaction will occur is randomly chosen with τ exponentially distributed with parameter $\sum_{\nu=1}^{M} a_\nu$;
- If there are no scheduled reactions in $[t, t + \tau]$:
  - Choose reaction R_μ with probability $\frac{a_\mu}{\sum_{\nu=1}^{M} a_\nu}$.
  - If R_μ is associated with a delay σ:
    - ★ update t to $t + \tau$
    - ★ schedule reaction R_μ for $t + \tau + \sigma$
  - Otherwise execute R_μ as in Gillespie's algorithm and update t to $t + \tau$;
- If there is a scheduled reaction R_ν at $t + \tau'$ with $\tau' \leq \tau$:
  - if R_ν is still applicable, apply R_ν
  - update t to $t + \tau'$

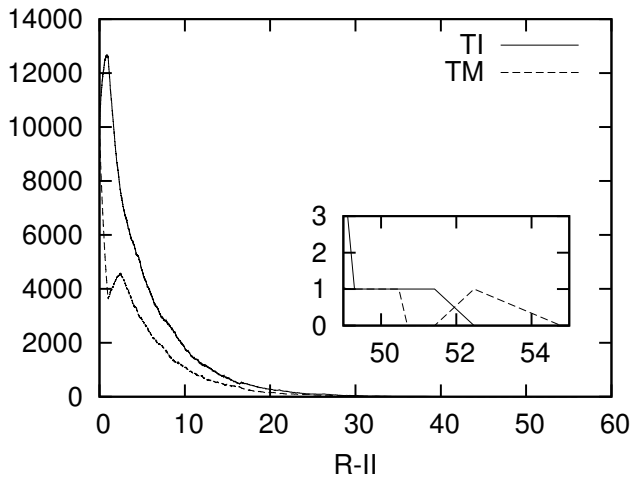
The Purely Delayed Approach

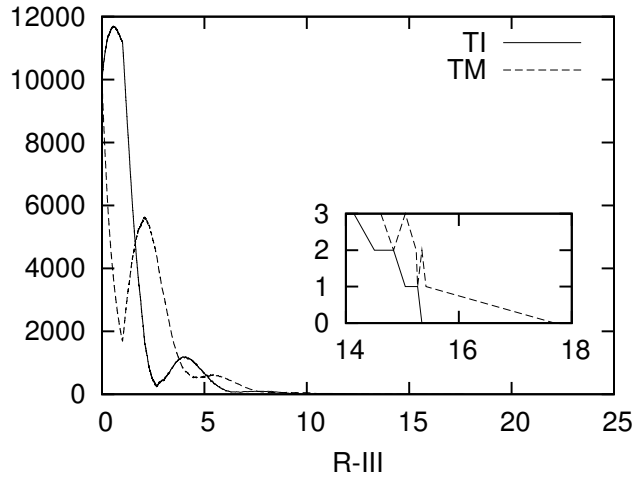


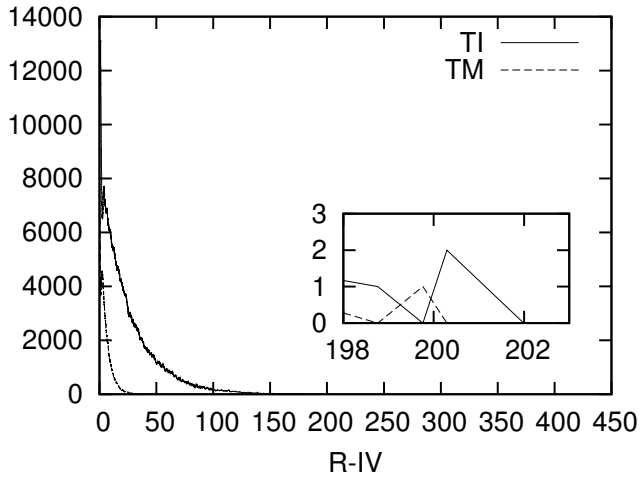
Purely Delayed Model of Tumor growth

These are some results of purely delayed stochastic simulation with $\sigma = 1$.



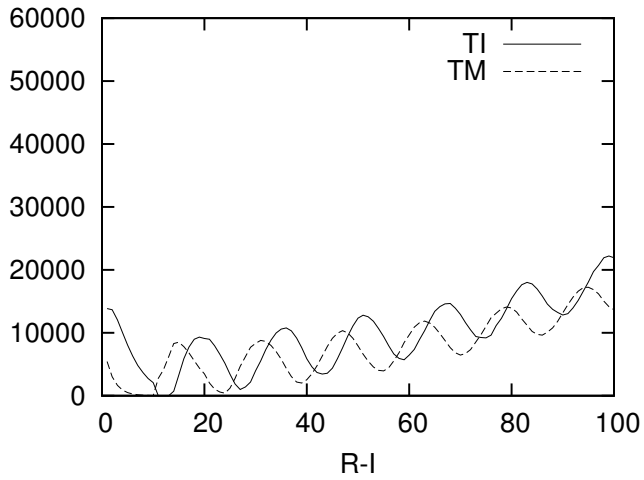


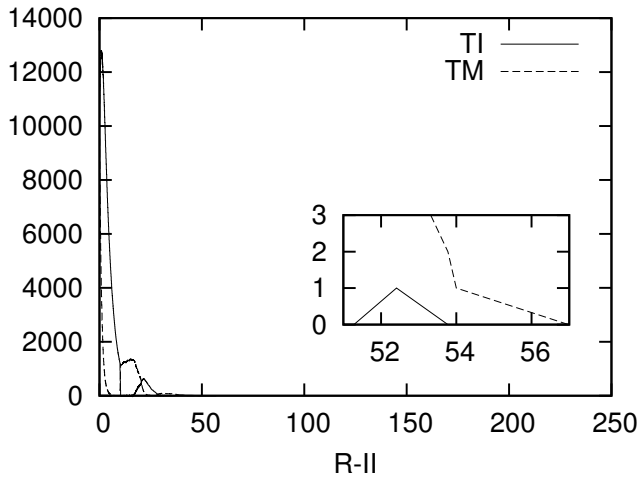


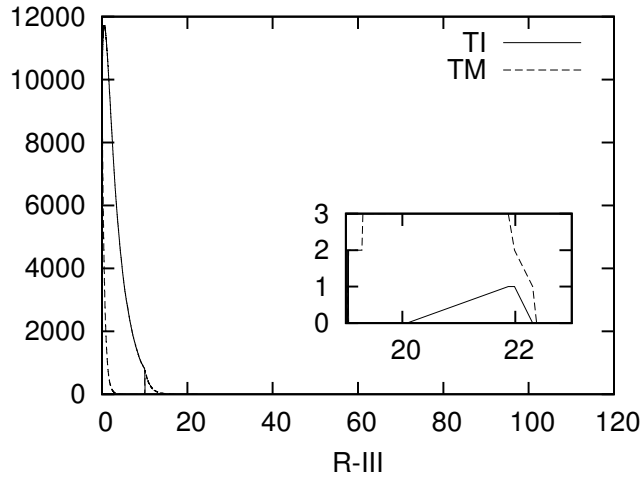


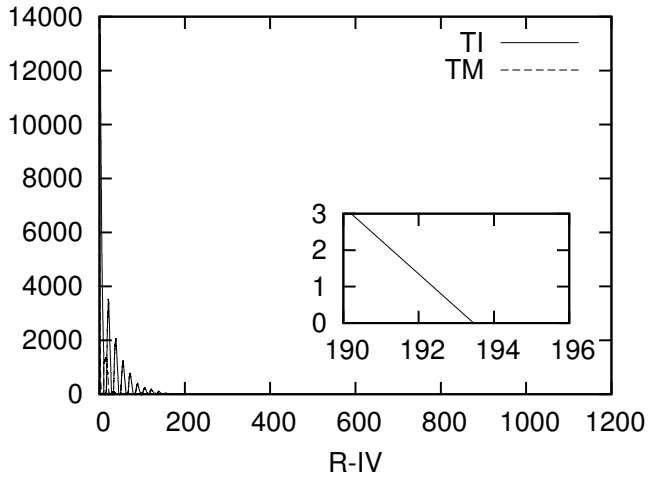
Purely Delayed Model of Tumor growth

These are some results of purely delayed stochastic simulation with $\sigma = 10$.









Purely Delayed Model of Tumor growth

Again, stochastic simulation results are qualitatively similar to previous results:

- all of the three approaches show tumor growth and eradication with similar parameters

But let us consider again average tumor eradication times:

	DDE	DSSA	PureDelay
R-II with $\sigma = 1.0$	50	64	51
R-II with $\sigma = 10.0$	59	224	67
R-III with $\sigma = 1.0$	15	29	17
R-III with $\sigma = 10.0$	12	126	20
R-IV with $\sigma = 1.0$	238	302	214
R-IV with $\sigma = 10.0$	440	1072	248

Some considerations

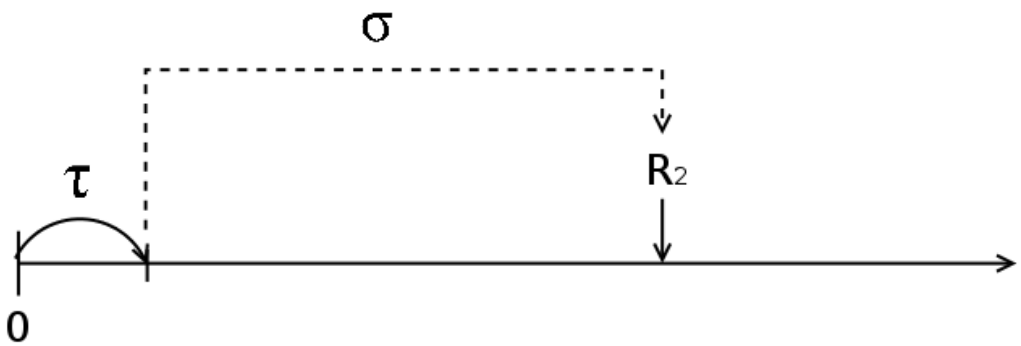
The purely delayed approach is not in general better than Barrio's approach

- it depends on the phenomena to be modelled

Optimal solution: allow both the approaches to be used in models

Moreover, the purely delay approach has to be improved:

- Correctness issue: the reactants of a scheduled reaction may disappear and be recreated
- Performance issue: the same reaction can be scheduled several times on the same reactants



Further Developments

We are developing a CCS-like process algebra that includes stochasticity and delays as in the simulation algorithms

- not ready for presentation...

Outline of the talk

Uncertain kinetic rates

Kinetic parameters of (bio)chemical reactions are often very difficult to estimate precisely

- the rate of a reaction depends many physical parameters: temperature, pH, volumes, etc. . .

Moreover, some parameters cannot be measured at all in laboratory

- inferred (with rough approximations) from similar reactions

The approach we propose consists in:

- replacing kinetic constants with intervals of possible values
- applying probabilistic model checking to obtain conservative upper and lower bounds for probabilistic reachability properties

We expolit abstract interpretation techniques to prove the correctness of our approach

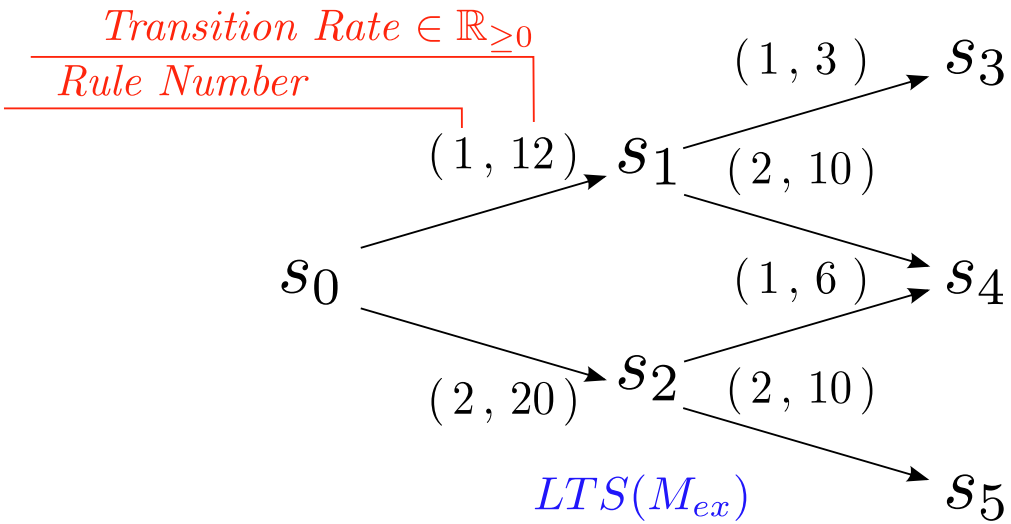
Probabilistic Reachability without Uncertainty

Let us consider the following simple example:

$$M_{ex} = \{ \quad R_1 : X\ Y \xrightarrow{3} Z \qquad R_2 : X\ W \xrightarrow{1} W \quad \}$$

with initial state $s_0 = 2X\,2Y\,10W$.

We can easily construct the following Labelled Transition System (LTS):



where the transition rate is computed as in Gillespie's algorithm.

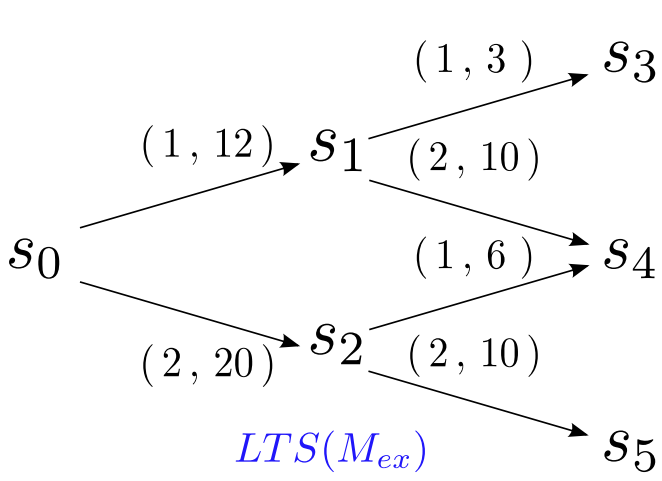
Probabilistic Reachability without Uncertainty

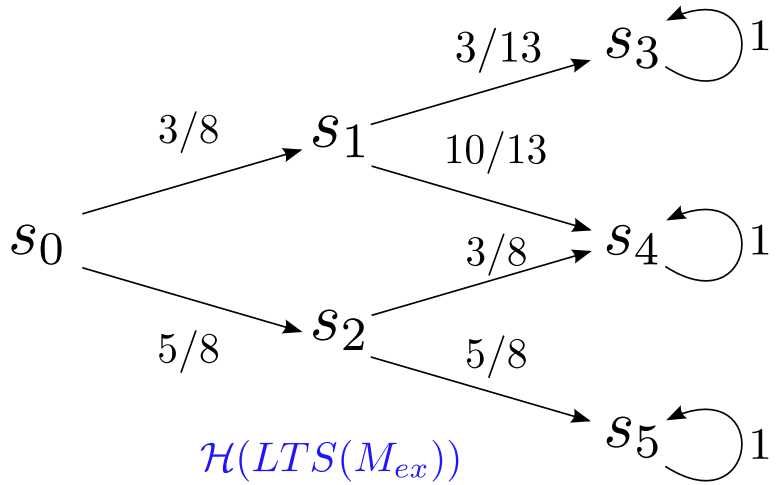
Let us consider the following simple example:

$$M_{ex} = \{ \quad R_1 : X\ Y \xrightarrow{3} Z \qquad R_2 : X\ W \xrightarrow{1} W \quad \}$$

with initial state $s_0 = 2X\,2Y\,10W$.

We can translate the LTS into a Discrete Time Markov Chain (DTMC):





We consider only sequentiality of events and we loose information on the elapsing of time.

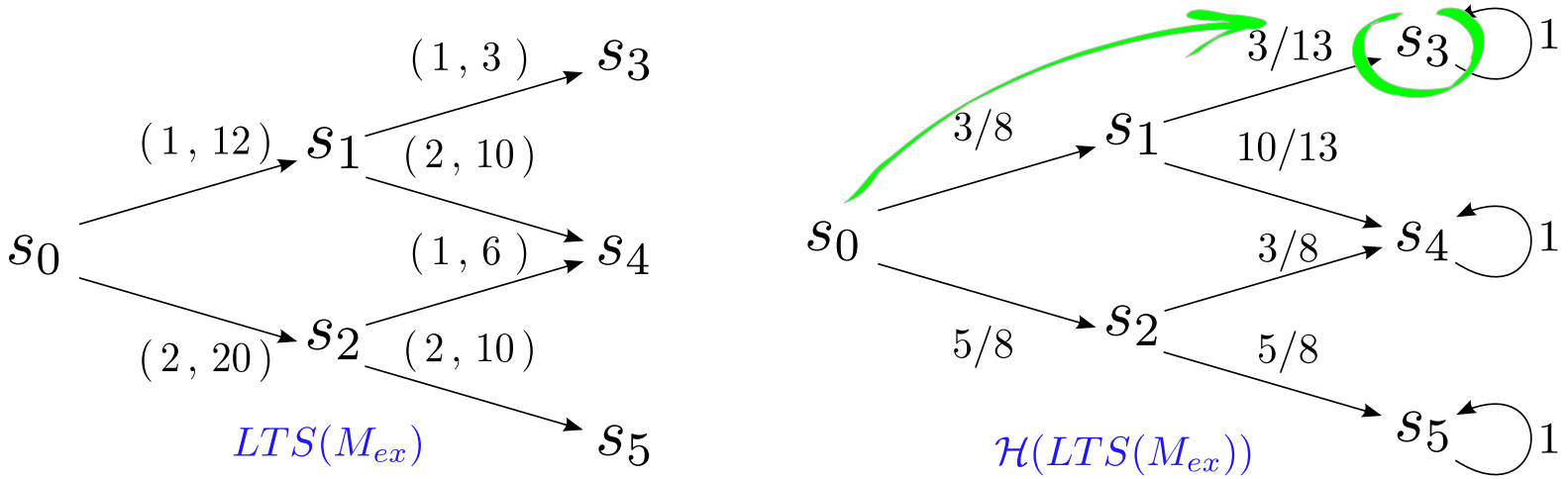
Probabilistic Reachability without Uncertainty

Let us consider the following simple example:

$$M_{ex} = \{ \quad R_1 : X\ Y \xrightarrow{3} Z \qquad R_2 : X\ W \xrightarrow{1} W \quad \}$$

with initial state $s_0 = 2X\,2Y\,10W$.

The DTMC can be used for probabilistic reachability analysis:



Example: $P(\text{obtaining two } Z) = Reach(s_3) = 3/8 \times 3/13 = 9/104$

Probabilistic Reachability with Uncertainty

Our approach:

- we allow intervals of possible values to be used in place of kinetic constants
- a model of chemical reactions with intervals (abstract model) represents an infinite set of models of reactions with kinetic constants (concrete models)

For example, the following abstract model

$$M^{\circ}_{ex} = \{ \quad R^{\circ}_1 : \; X \; Y \xrightarrow{[1,5]} Z \qquad R^{\circ}_2 : \; X \; W \xrightarrow{[1,5]} W \quad \}$$

includes the previously considered concrete model

$$M_{ex} = \{ \quad R_1 : \; X \; Y \xrightarrow{3} Z \qquad R_2 : \; X \; W \xrightarrow{1} W \quad \}$$

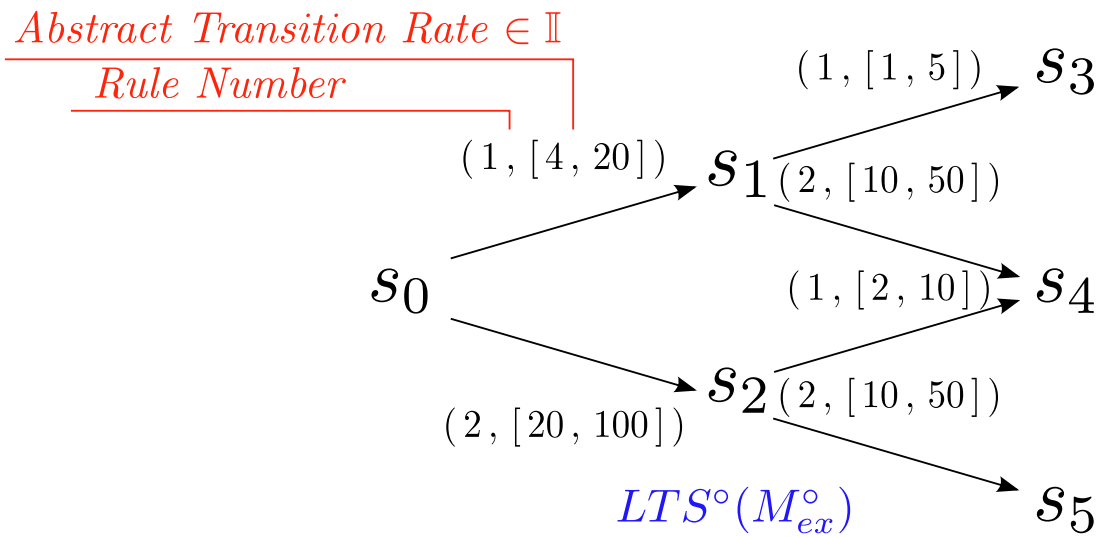
Probabilistic Reachability with Uncertainty

Let us consider the following simple example:

$$M^{\circ}_{ex} = \{ \quad R^{\circ}_1 : X\ Y \xrightarrow{[1,5]} Z \qquad R^{\circ}_2 : X\ W \xrightarrow{[1,5]} W \quad \}$$

with initial state $s_0 = 2X\,2Y\,10W$.

We can easily construct the following Labelled Transition System (LTS):



where the abstract transition rate is computed as in Gillespie's algorithm on the interval endpoints.

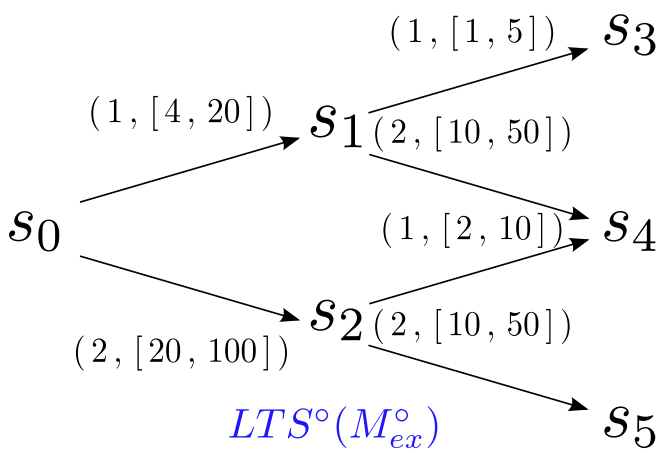
Probabilistic Reachability with Uncertainty

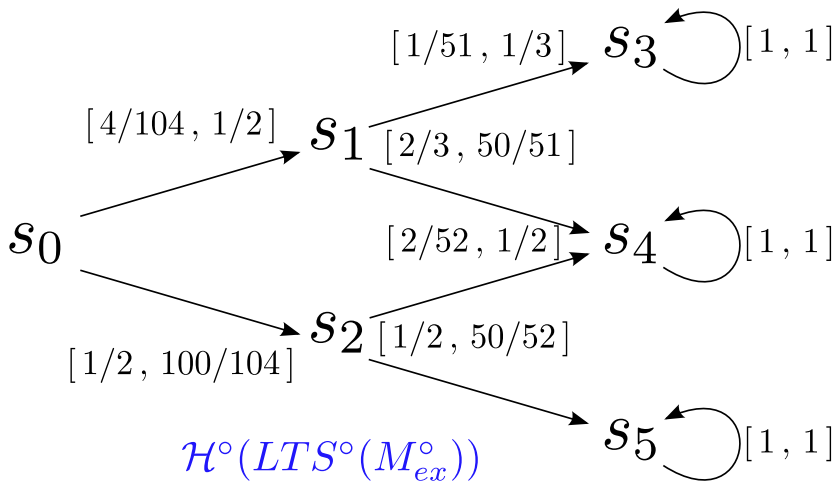
Let us consider the following simple example:

$$M^{\circ}_{ex} = \{ \quad R^{\circ}_1 : X\ Y \xrightarrow{[1,5]} Z \qquad R^{\circ}_2 : X\ W \xrightarrow{[1,5]} W \quad \}$$

with initial state $s_0 = 2X\,2Y\,10W$.

We can translate the LTS into a Interval Markov Chain (IMC):





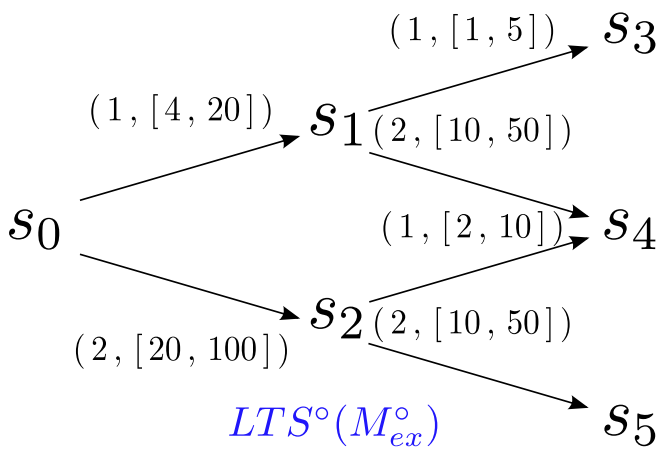
Probabilistic Reachability with Uncertainty

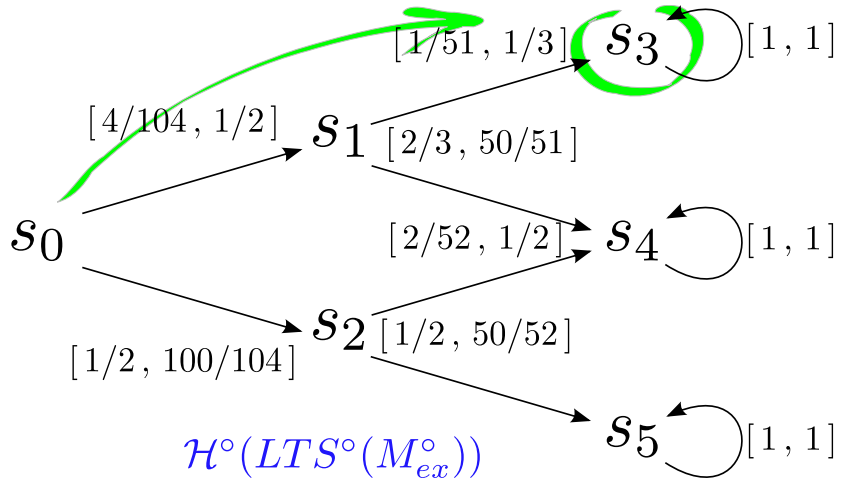
Let us consider the following simple example:

$$M^{\circ}_{ex} = \{ \quad R^{\circ}_1 : \; X \; Y \xrightarrow{[1,5]} Z \qquad R^{\circ}_2 : \; X \; W \xrightarrow{[1,5]} W \quad \}$$

with initial state $s_0 = 2X\,2Y\,10W$.

The IMC can be used for probabilistic reachability analysis:



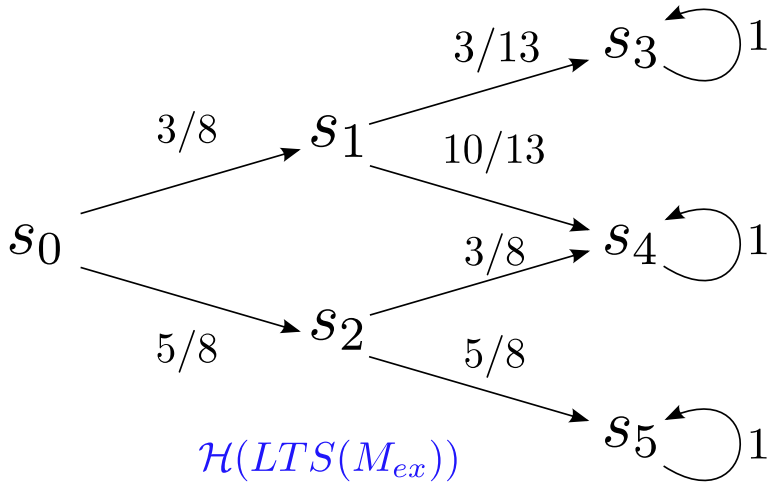


Example: $P(\text{obtaining two } Z) = Reach(s_3) =$

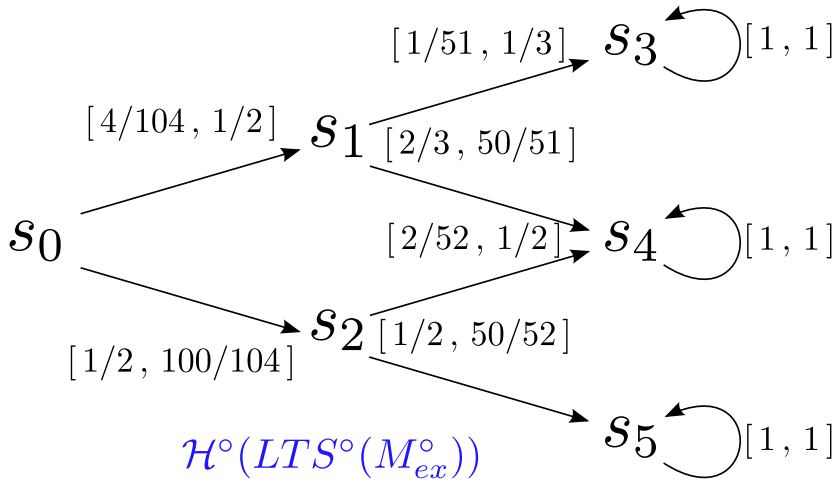
$$= [4/104, 1/2] \times^{Int} [1/51, 1/3] = [1/1326, 1/6]$$

Probabilistic Reachability with Uncertainty

In a DTMC the outgoing transitions of each state are associated with a probability distribution



In a IMC the outgoing transitions of each state may be associated with a infinite number of probability distributions



Probabilistic Reachability with Uncertainty

We have proved that the probability distributions of states of a concrete model M are included in those of the corresponding abstract model M°

- abstract probabilistic reachability gives correct upper- and lower-bounds

We have applied standard abstract interpretation techniques:

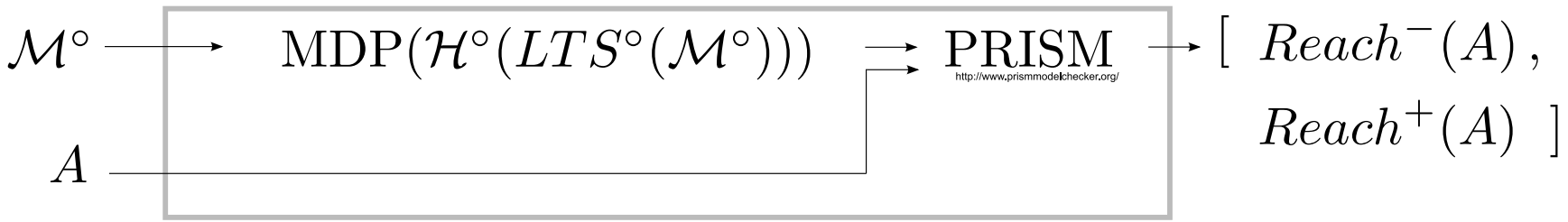
$$\begin{array}{ccccc} \mathcal{M}^\circ & \xrightarrow{LTS^\circ} & \mathcal{LTS}^\circ & \xrightarrow{\mathcal{H}^\circ} & \mathcal{IMC} \\ \uparrow \alpha & & \uparrow \alpha_{\mathcal{LTS}} & & \uparrow \alpha_{\mathcal{MC}} \\ \mathcal{M} & \xrightarrow{LTS} & \mathcal{LTS} & \xrightarrow{\mathcal{H}} & \mathcal{DTMC} \end{array}$$

Probabilistic Reachability with Uncertainty

Probabilistic reachability analysis becomes more complex when the model consists of more than two chemical reactions

- We have followed a standard extreme distributions approach (Fecher et Al.) that requires translation of the IMC into a Markov Decision Process (MDP)

We have developed a translator from chemical reactions with uncertain rates into PRISM input language



- AMSR2PRISM translator,
  `http://www.di.unipi.it/msvbio/`
- PRISM model checker,
  `http://www.prismmodelchecker.org`

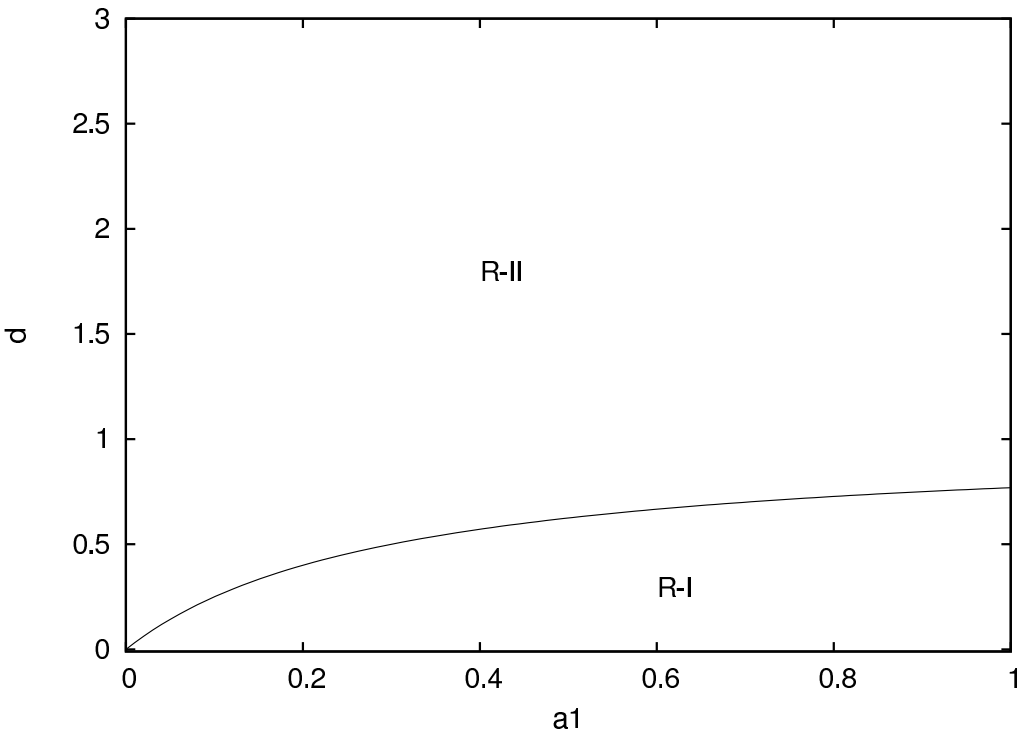
Probabilistic Reachability in the Tumor Growth Model

Let us reformulate the tumor growth example without delays.

Reactions:

$$T_M \xrightarrow{a_4} 2\,T_I \qquad T_I \xrightarrow{d_2}$$

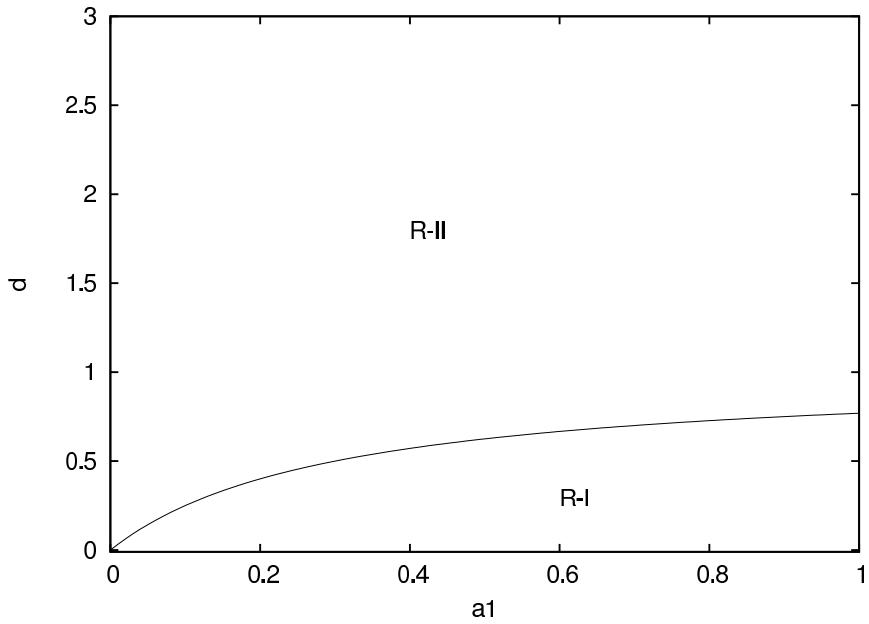
$$T_I \xrightarrow{a_1} T_M \qquad T_M \xrightarrow{d_3}$$



In this case we have only two parameter regions:

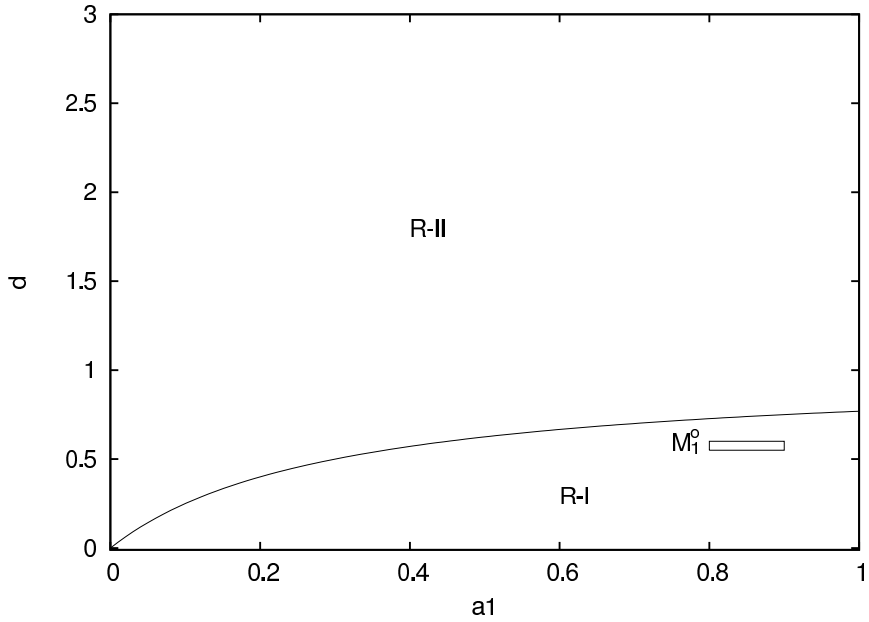
- In R-I the tumor grows
- In R-II the tumor decays

Probabilistic Reachability in the Tumor Growth Model



We consider three abstract models of tumor growth.

Probabilistic Reachability in the Tumor Growth Model



We consider three abstract models of tumor growth.

Abstract model M_1° :

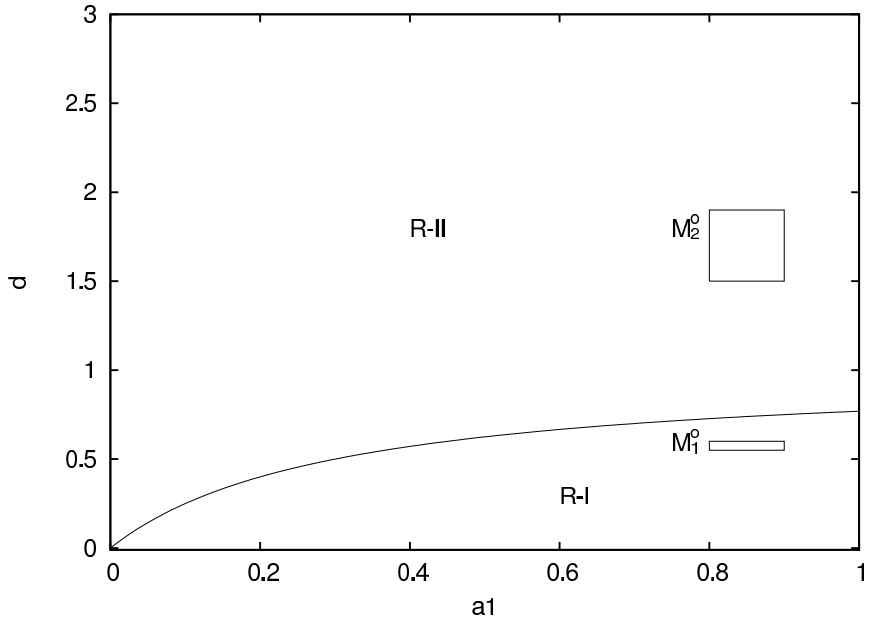
$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[0.05,0.1]}$

Probabilistic Reachability in the Tumor Growth Model



We consider three abstract models of tumor growth.

Abstract model M_1° :

$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[0.05,0.1]}$

Abstract model M_2° :

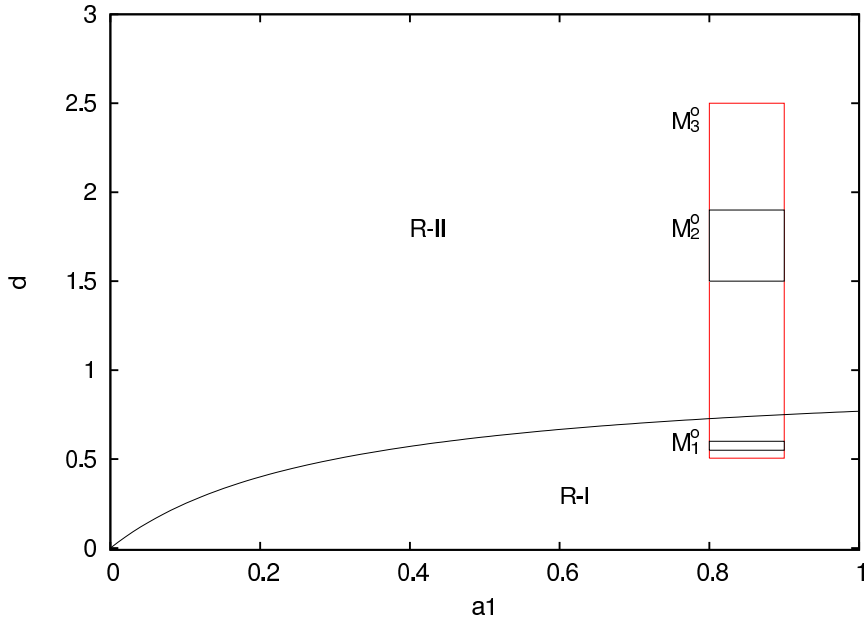
$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[1,1.4]}$

Probabilistic Reachability in the Tumor Growth Model



We consider three abstract models of tumor growth.

Abstract model M_1° :

$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[0.05,0.1]}$

Abstract model M_2° :

$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[1,1.4]}$

Abstract model M_3° :

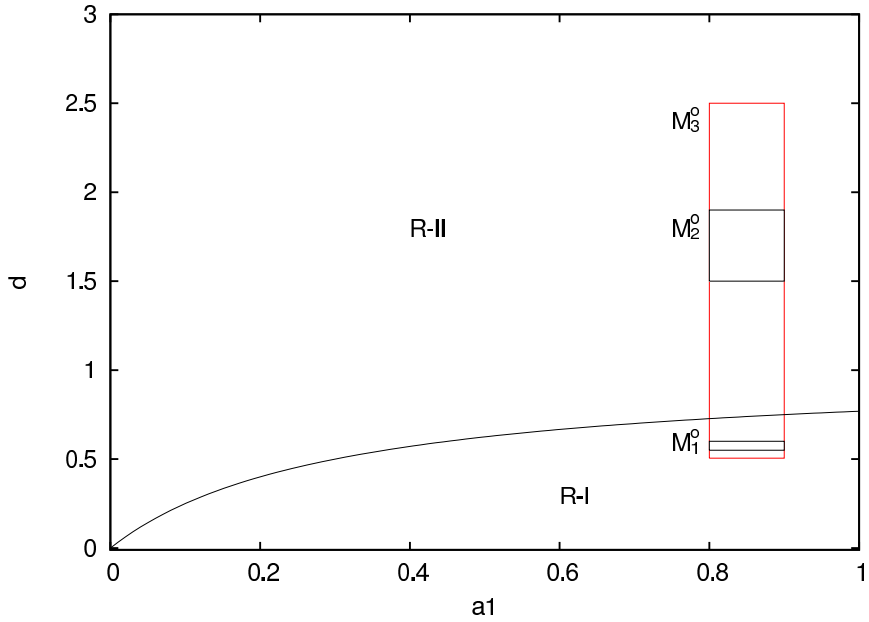
$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[0.005,2]}$

Probabilistic Reachability in the Tumor Growth Model



We consider three abstract models of tumor growth.

We consider an initial population consisting of $10\,T_M$ and $10\,T_I$.

Abstract model M_1° :

$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[0.05,0.1]}$

Abstract model M_2° :

$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[1,1.4]}$

Abstract model M_3° :

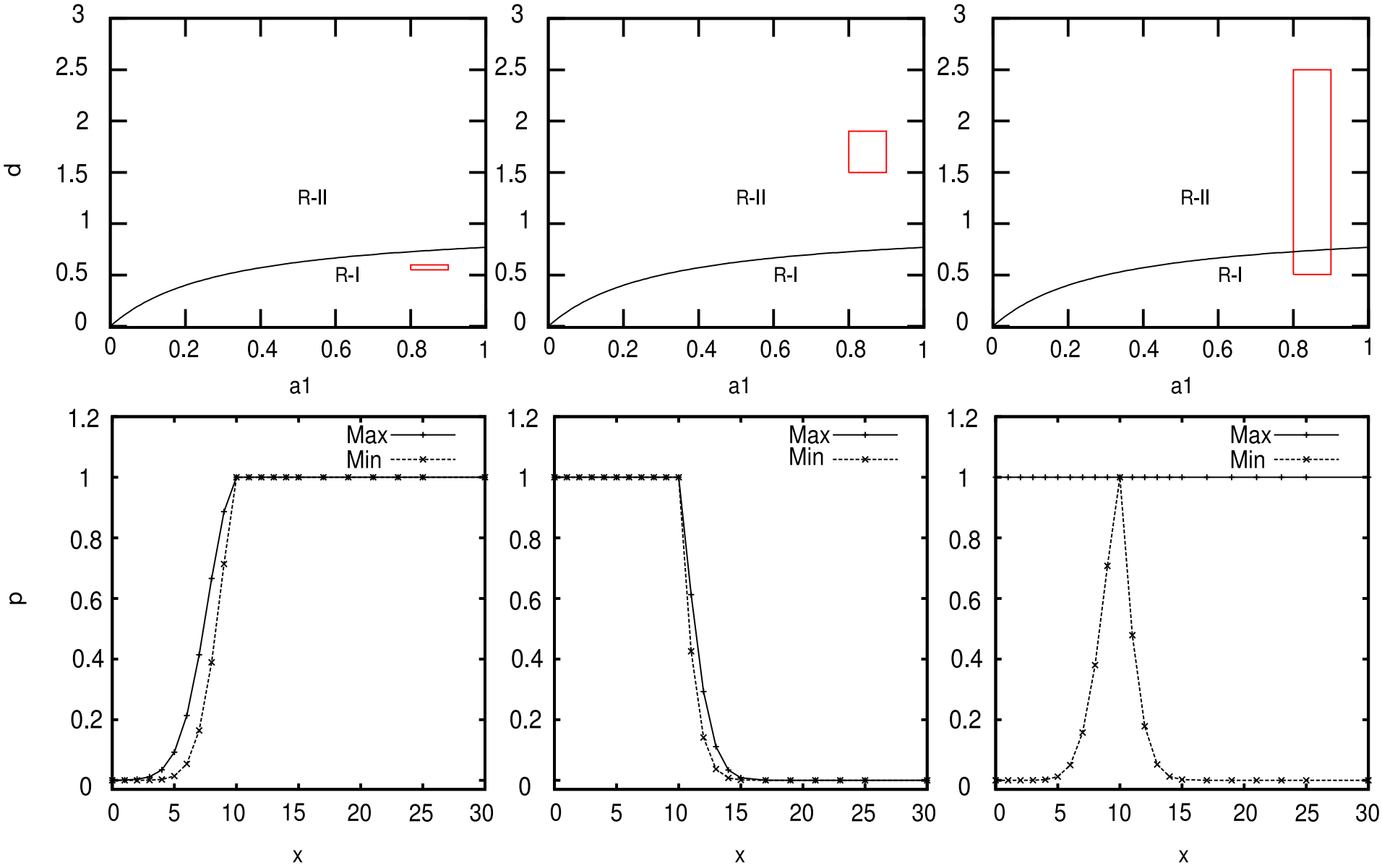
$T_M \xrightarrow{0.5} 2\,T_I$

$T_I \xrightarrow{[0.8,0.9]} T_M$

$T_I \xrightarrow{0.3}$

$T_M \xrightarrow{[0.005,2]}$

Probabilistic Reachability in the Tumor Growth Model



$Reach(T_M = x)$ on $M_1^\circ, M_2^\circ, M_3^\circ$

Some Considerations

Our approach gives meaninful answers when the sensitivity of the system on variation of the uncertain parameters is not too high

The approach can also be used for parameter estimation by iteratively

1. constructing an abstract model with wide intervals
2. checking properties known to hold
3. refine the model until model checking gives [1,1] as result

The efficiency of the approach depends very much on the number of uncertain parameters

- the translation of an IMC into a MDP is exponential in the number of parameter intervals

Further Developments

We are working at a continuous time approach, in which the elapsing of time is taken into account

References

R. Barbuti, G. Caravagna, A. Maggiolo-Schettini and P. Milazzo. **On the interpretation of delays in delay stochastic simulation of biological systems**. Proc. of CompMod'09, EPTCS, in press.

R. Barbuti, F. Levi, P. Milazzo and G. Scatena. **Probabilistic Model Checking of Biological Systems with Uncertain Kinetic Rates**. Int. Conference on Reachability Problems (RP'06), LNCS 5797, pp. 64-78, 2009.