Poisson Simulation - Realisation of time continuous dynamic & stochastic processes

Leif Gustafsson

LT, SLU

leif.gustafsson@lt.slu.se

File: Poisson_Simulation.ppt - 000529

ABSTRACT

Poisson simulation is a method to introduce stochastics into continuous system simulation in a realistic way. In e.g. biological modelling you may describe the system in terms of states and flows. The states, representing a number of subjects (animals, plants etc.), change because of in- and outflows. These flows can be handled stochastically by using random numbers. Thereby you can model the stochastics as opposed to just adding input noise. A number of examples of this will be shown.

<u>Two ways to study the real systems</u>

- **Study the real <u>system.</u>**

But:

  - planning
  - understanding
  - interpretation
  - theory
  - ...

Requires a model !

- **Build a <u>model</u> and study it.**

But:

  - problem
  - structure
  - data
  - validation
  - …

requires the system !

In either case you need both the system and the model !!!

Analysis or experiment on the model?

- **Analysis of a model**

+ Gives a closed solution.

- Works only in very simple cases.

- **Experiments on a model**

+ Works also for complex cases.

- The experiment shows only a special case.

Different types of models

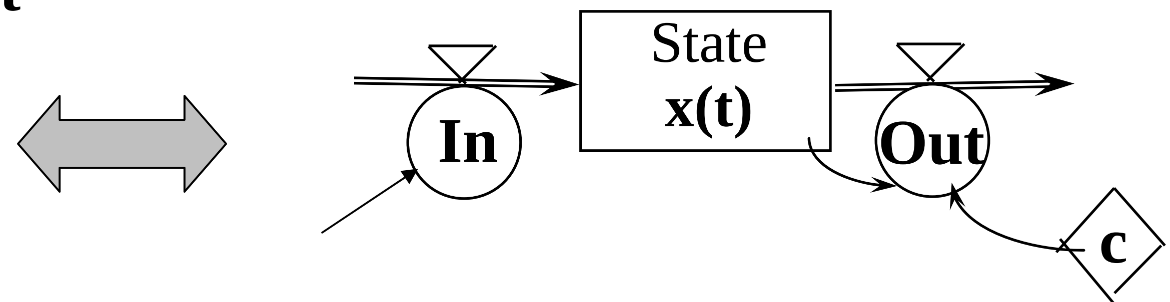
	Static	Dynamic
Stoch-astic	***Statistical models***	***Dynamic & stochastic models*** **!**
Determi-nistic	***Algebraic models***	***Differential. equ. models***

Dynamic (deterministic) models

One type of dynamic models is systems of differential equations, or equivalently state/flow-models (see Appendix).

Differential equation

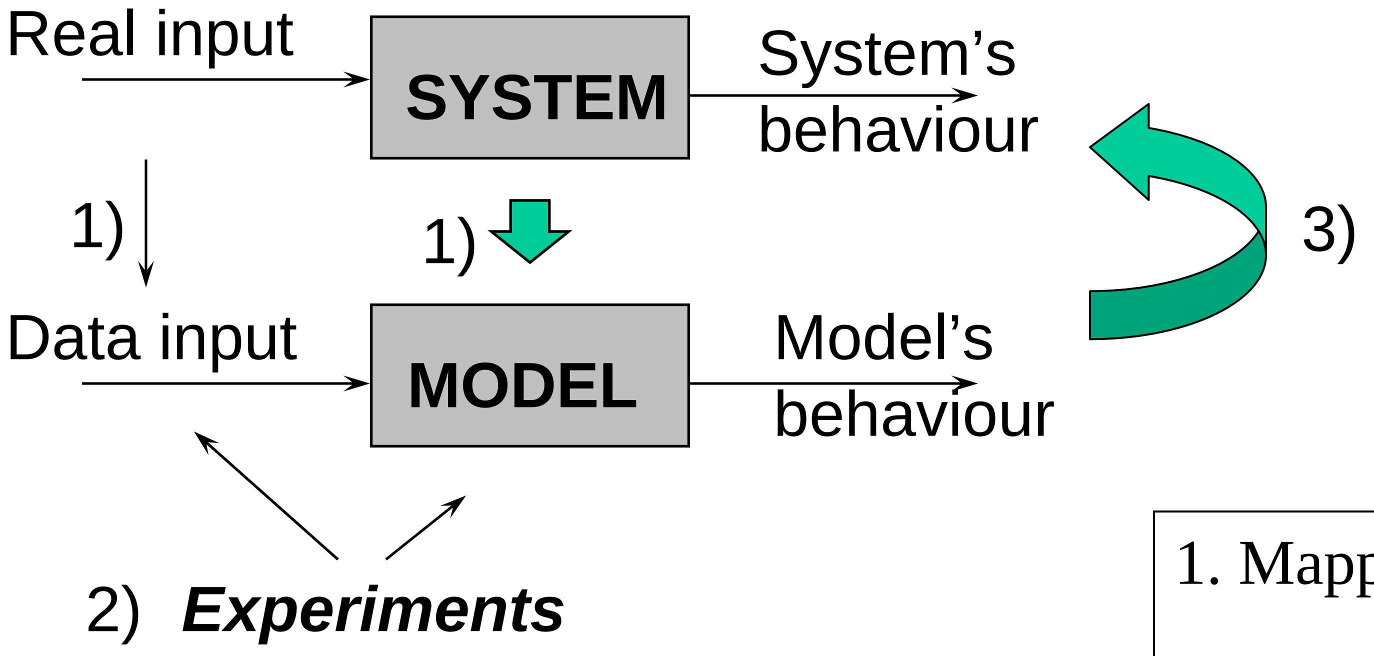
$$\begin{cases} dx(t)/dt = In - Out \\ In = \ldots \\ Out = c \cdot x(t) \\ x(0) = x_o \end{cases}$$



Dynamics originates from in- and out flows that changes the values of states.

A solution can always be obtained by *simulation*.

SIMULATION
= experiments on a model



1. Mapping
2. Experiments
3. Conclusions

Stochastic models

- Regression models of various kinds.
  - Logistic models.
  - Poisson models.
  -

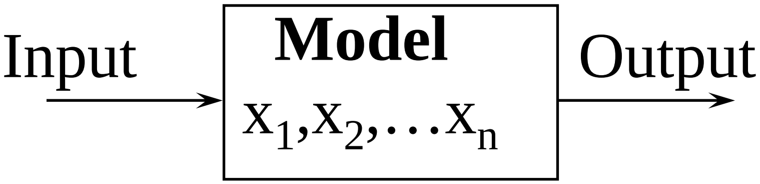
Pick a distribution (NegExp, Normal, Erlang, Weibull etc.). This distribution **D(a,b...)** has a number of parameters.

Find that set of parameter values that minimises the difference between data (describing the system) and the distribution (model) in least square sense. (Or use maximum likelihood methods to do the same.)

WARNING: A "wrong" distribution gives bad estimates! The chosen distribution has to originate from a realistic dynamical structure!

Internal or external description?

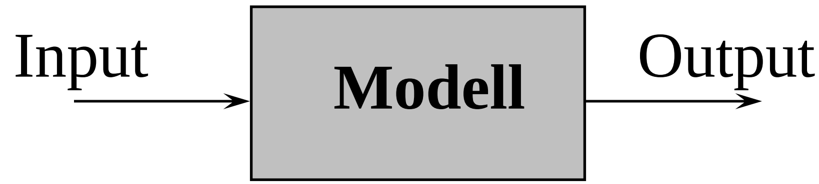
- **Internal model (Structural model)**



Description of the inner structure.

How the components affect each other..

- **External model (“Black box”)**



Describes only the input-output relation.

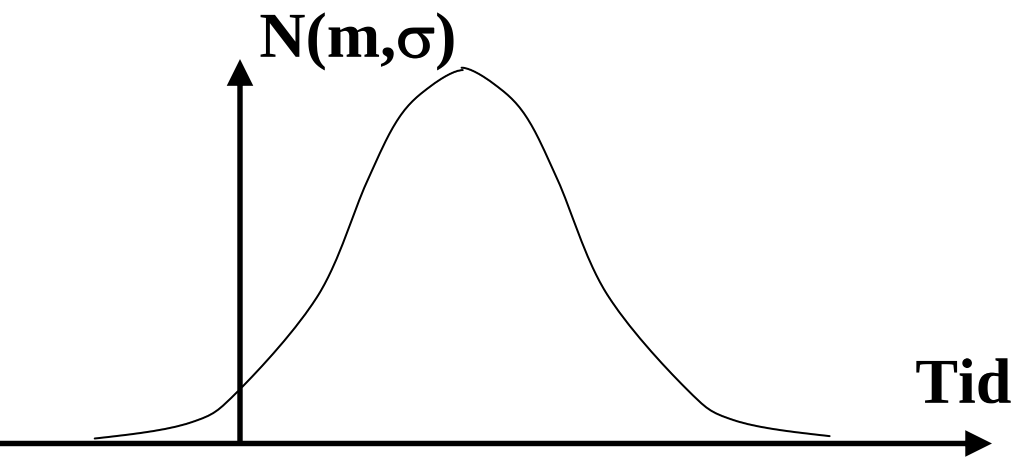
To start from a structure or a distribution?

- **Systems analysis** starts with a ***structure***.
- Focus on the *dynamic process* - but the uncertainty?
- The dynamic model is fitted to data (e.g. *parameter estimation*).

- **Statistics** starts with a ***distribution.***
- Focus on uncertainty - but the dynamics?
- The statistic model is fitted to data by *regression*.

To start from a structure or distribution? - II

- In principle every dynamic structure generates a distribution over time.

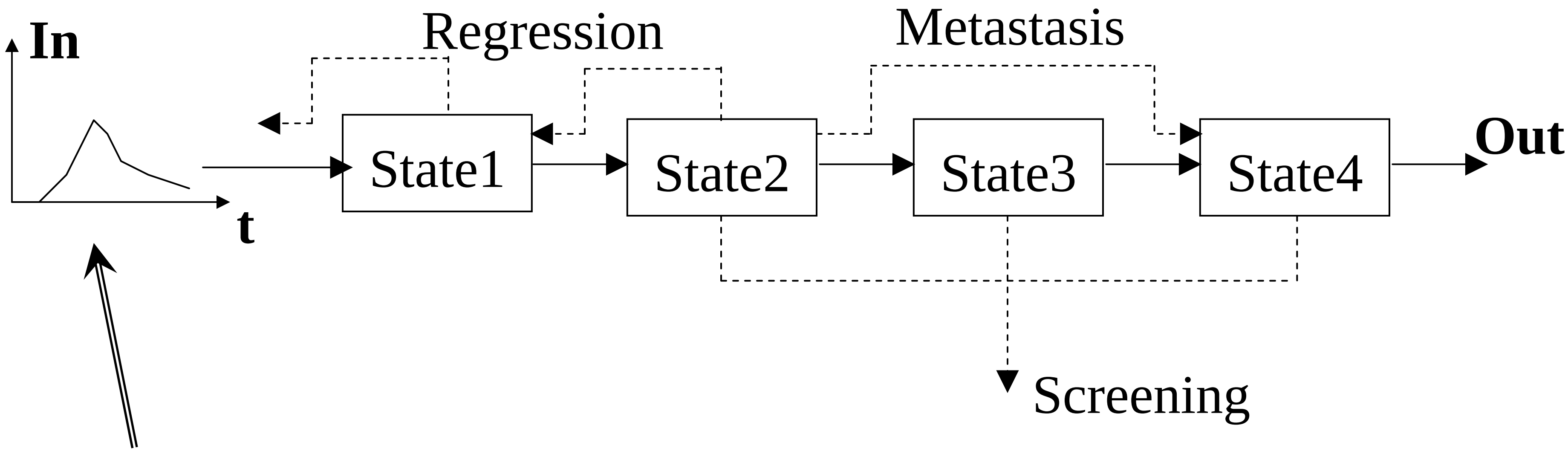


You may assume e.g. a Gaussian distribution and fit it to data.

But what structure would it originate from ??

Reasons for starting with the structure

1. There are no named statistical distributions to most structures.



2. Inputs (and other quantities) may be empirical functions.

3. The initial value problem. The States are usually not zero at the start!

4. When you want to understand/modify/control etc. you need an explanation (structural) model - not a black box.

Could we use dynamical methods instead of statistical ones?

No! Stochastic variations excite the dynamics.

No! We need statistical estimates (mean, variance, confidence intervals, P-values, hypothesis tests etc.)

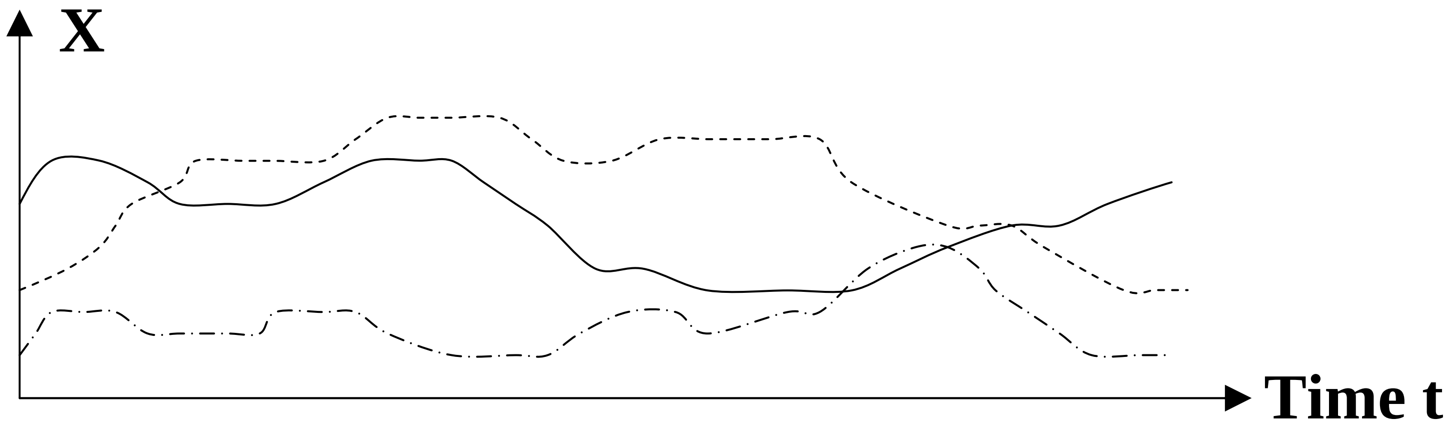
*****

The same model should handle dynamics and stochastics in a correctly integrated way. (You can't just add random noise!)

A dynamic & stochastic process

- A dynamic & stochastic model results in stochastic solutions that vary dynamically.

- This process is really a function of two variables $\{X(t, \omega), t \in T, \omega \in \Omega\}$ where t is time and ω is a solution.



- If we fix t, we get a stochastic variable.
- If we fix ω, we get a realisation (trajectory).

Dynamic & stochastic processes - cont.

- Dynamic & stochastic processes become *mathematically and statistically difficult* when the system becomes complex.
- There are a number of techniques to *realise* a complex stochastic process, e.g *Markov-simulation* and *Discrete Event Simulation.*
- *For time continuous* stochastic processes *Poisson-simulation* may be used.

The Poisson-assumption:

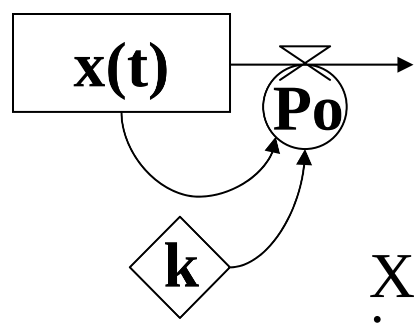
A flow contains a number of entities (events) per time unit. If these events happen *independently* and *one at a time* - we have a Poisson process.

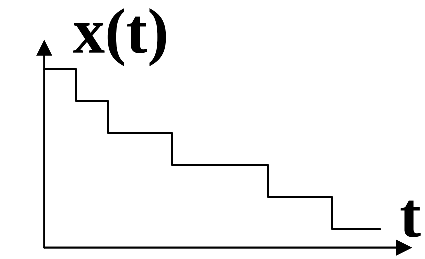
Then the number of events during a short enough time interval (dt) becomes very small (0 or 1).

Further, P(event) is proportional to the time interval dt.
$\Rightarrow$ The number of events over a finite time interval (Δt) becomes Poisson distributed.

We may then write:

$$\begin{cases} x(t+\Delta t) = x(t) - \Delta t \cdot f(x,t) \\ f(x,t) = \text{Poisson}[\underbrace{k \cdot x(t)}_{\text{Expected flow rate.}} \cdot \Delta t]/\Delta t \\ x(0) = x_0 \end{cases}$$





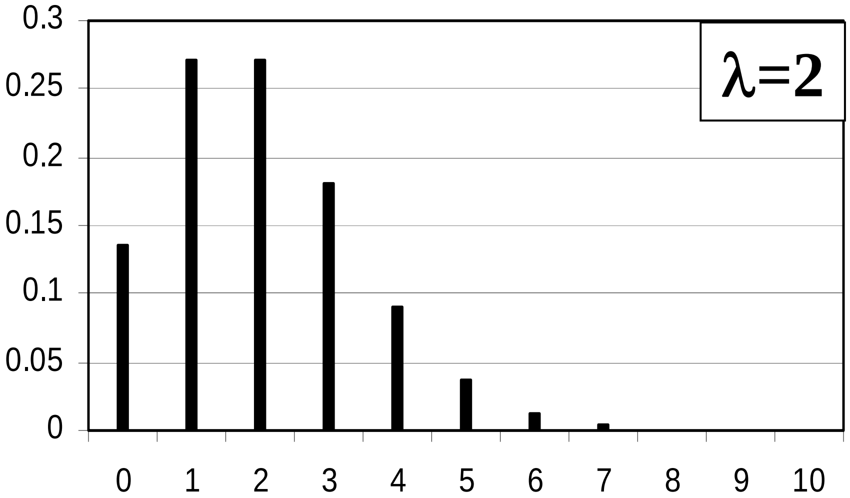
X and f are always integers!

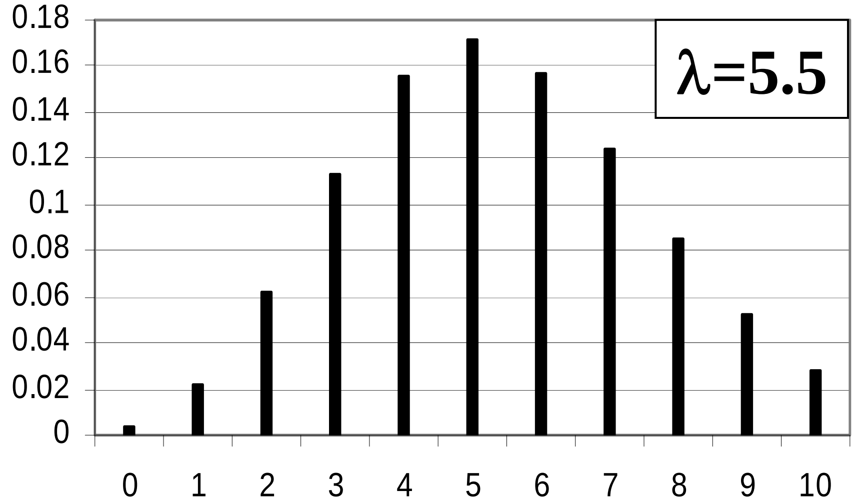
The Poisson distribution

- Number of events that occur in an interval when the events occur independently and one at a time.

- $p(x) = e^{-\lambda} \cdot \lambda^x / x!$ if $x \in \{0, 1, ...\}$ else 0.
(Note that the outcome, x, is always an integer!)

- Po(λ) has only one parameter λ and E(X)=Var(X)=λ.





Poisson distribution or not ?

- The simplest possible case is when events happen *independently* and *one at a time* (Poisson process).

  The intensity λ in Po(λ) may vary over time (non-stationary process).
- If they do not there are an infinite number of possibilities! (Compare: The linearity to the infinite number of ways to be non-linear.)

If it is not a Poisson process you may e.g. use:

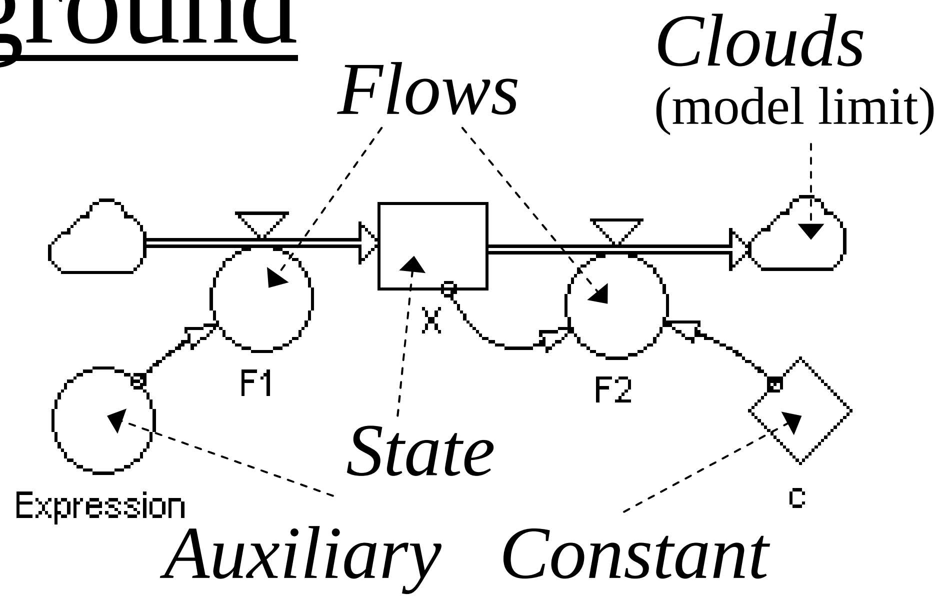
- any distribution D(..).
- sum of distributions D_1(..) + D_2(..).
- product of distributions D_1(..) $\cdot D_2$(..).
- distribution of distributions D_1(D_2(..)).
- D(p_1, p_2) where the parameters can be separately modelled.

Pragmatic view: Do we get a good enough model?

Background

A time continuous process:

$$\begin{cases} dx/dt = f1 - f2 \\ x(0) = x_0 \\ f1 = \textit{some expression} \\ f2 = c^*x \end{cases}$$



This process may be *numerically* described as (Euler):

$$\begin{cases} x(t+\Delta t) = x(t) - \Delta t \cdot (f1(x,t) - f2(x,t)) \\ f1(x,t) = \textit{some expression} \\ f2(x,t) = k \cdot x(t) \\ x(0) = x_0 \end{cases}$$

At the simulation x(t) is first given the initial value x_0.
Then the results are calculated step (Δt) by step.

Demonstrations

- Radioactive decay - (NegExp.sim)
- Logistic growth - (Logistic.sim)
- Gompertz’ growth - (Gompertz)
- Volterra’s equations - (Volterra.sim, Volt_SS.sim)
- Multi-hit model - (MHit2.sim, Mhit2_Po.sim, MHit3.sim)
- Screening - (Screen.sim)
- Study & Ctrl cohorts - (StdyRef.sim, StdyRef2.sim)
- Model fitting - (PEstLogA.sim, PEstLogB.sim, PEstLogC.sim)

Radioactive decay

- Study of N=100 radioactive atoms that will decay with a time constant of T= 50 ($T_{½}$=34.7) time units.
- (When N is large we can neglect the statistical fluctuations.)

Conventional

$$\begin{cases} \Delta N = -\Delta t \cdot f \\ f = N/T \\ N(0) = 100 \end{cases}$$

$N(t)=N(0)\exp(-t/T)$

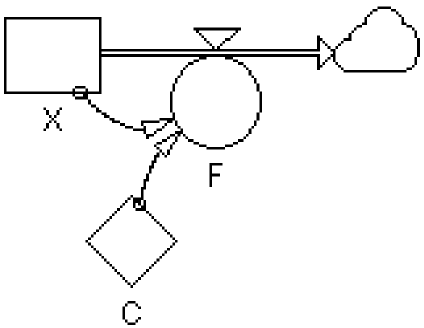
$\Longrightarrow$

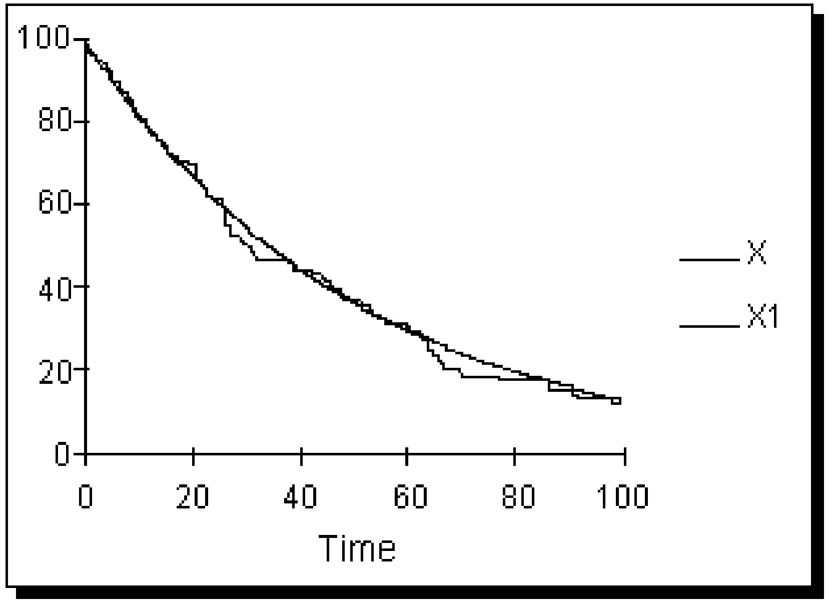
Poisson equivalent

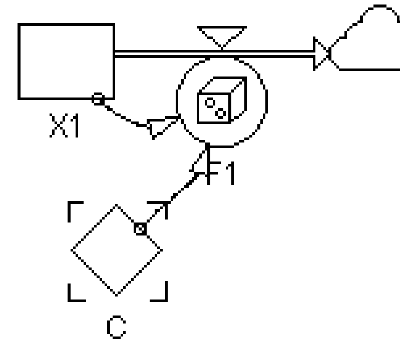
$$\begin{cases} \Delta N = -\Delta t \cdot f \\ f = Po[(N/T)\cdot \Delta t]/\Delta t \\ N(0) = 100 \end{cases}$$

Stochastic exp. solution

Radioactive decay.   File: NegExp.sim.







dX/dt = -F(X)
F = C*X
C = 0.02
X(0) = 100

dX/dt = - F(X)
F = Po(C*X*dt)/dt
C = 0.02
X(0) = 100

C=0.02 <=> T=1/C=50.

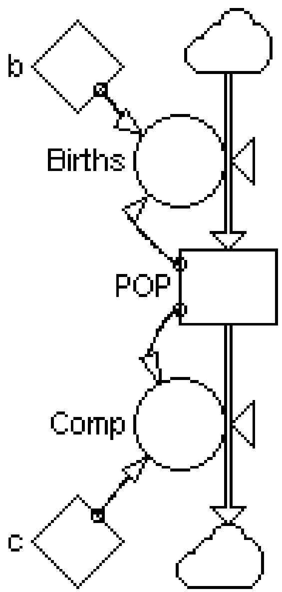
T½=ln 2*T =0.6931*T = 34.66

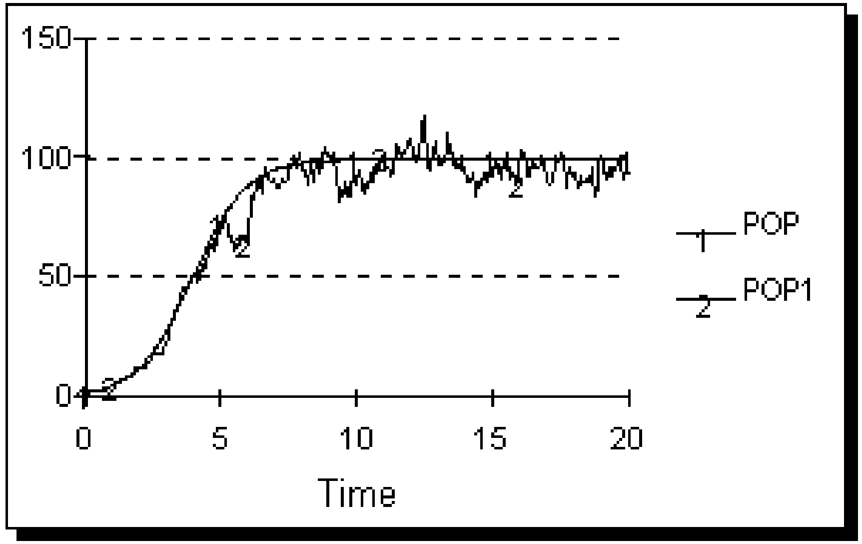
Logistic growth

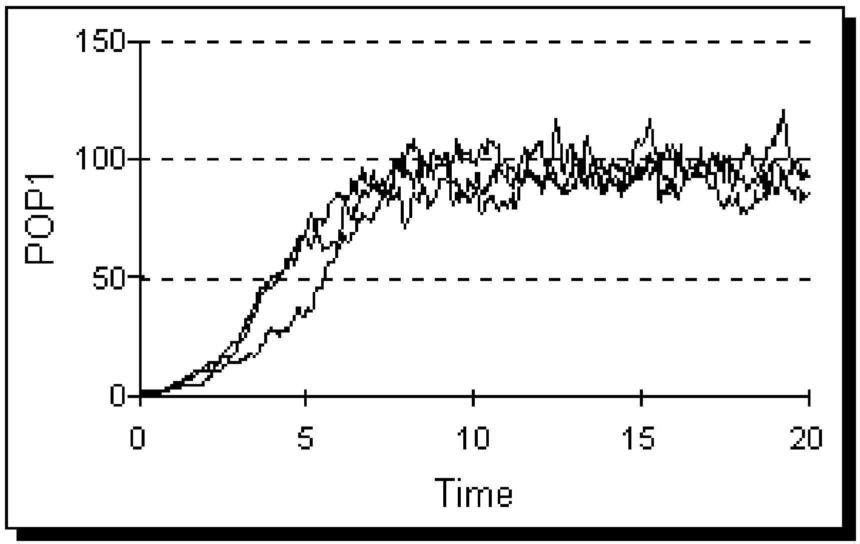
- A species will breed exponentially if there is no limits.
- Sooner or later resources will be scarce - the individuals will compete! (Everyone competes with everyone.)
- A simple deterministic model has the form:

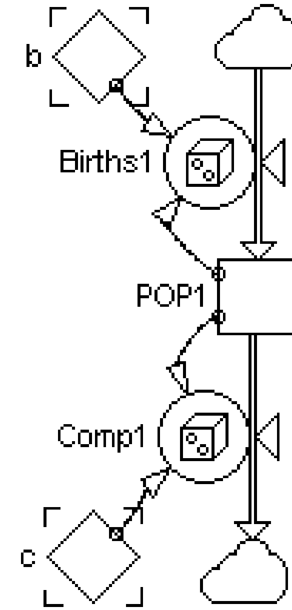
$$\begin{cases} \Delta X = \Delta t \cdot \big(f1 - f2\big) \\ f1 = b \cdot X(t) & \Rightarrow \quad Po[b \cdot X \cdot \Delta t]/\Delta t \\ f2 = c \cdot X(t) \cdot X(t) & \Rightarrow \quad Po[c \cdot X \cdot X \cdot \Delta t]/\Delta t \\ X(0) = \textit{initial value} \end{cases}$$

File: Logistic.sim 000529.









Differential equation model:

$dX/dt = f1 - f2$
$f1 = b*X$
$f2 = c*X*X$
$X(0) = 2$

Poisson simulation model
:

$dX/dt = f1 - f2$
$f1 = Po(b*X*dt)/dt$
$f2 = Po(c*X*X*dt)/dt$
$X(0) = 2$

Same end level.
Large variations around the trajectories.

Gompertz’ growth

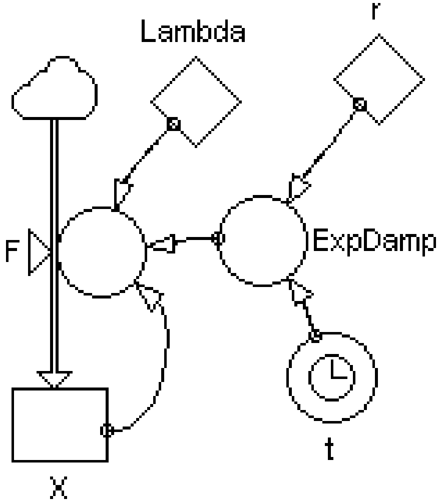
- Another way to model growth is to use Gompertz’ equation: $dX/dt = \lambda X \cdot \exp(-rt)$ where r and λ are constants and t is time.
- (It is sometimes given as: $dX/dt = rX \cdot \ln(K/X)$ where the constants r gives the growth rate and K the maximal size.)
- In both cases solution is: $\mathbf{X(t) = X0 \cdot \exp[(\lambda/r) \cdot (1-\exp(-rt))]}$.

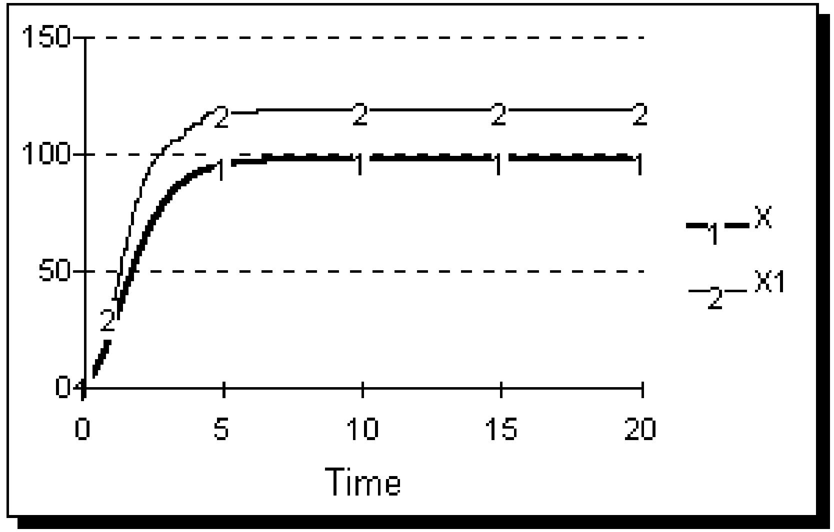
$$\begin{cases} \Delta X = \Delta t \cdot f \\ f = \lambda \cdot X \cdot \exp(-rt) \\ X(0) = \textit{initial value} \end{cases} \quad \Longrightarrow \quad Po[\lambda \cdot X \cdot \exp(-rt) \cdot \Delta t]/\Delta t$$

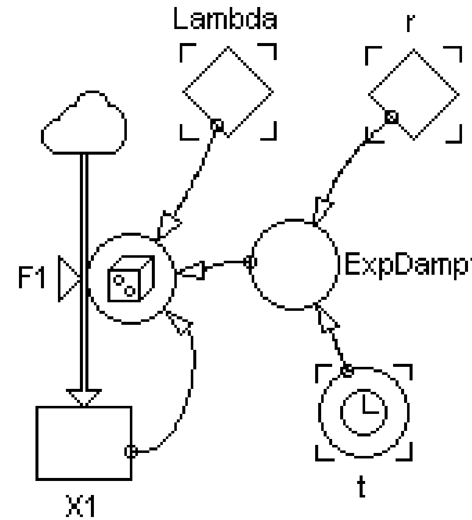
or

$$\begin{cases} \Delta X = \Delta t \cdot f \\ f = r \cdot X(t) \cdot \ln(K/X) \\ X(0) = \textit{initial value} \end{cases} \quad \Longrightarrow \quad Po[r \cdot X \cdot \ln(K/X) \cdot \Delta t]/\Delta t$$

Gompertz.sim   000515.

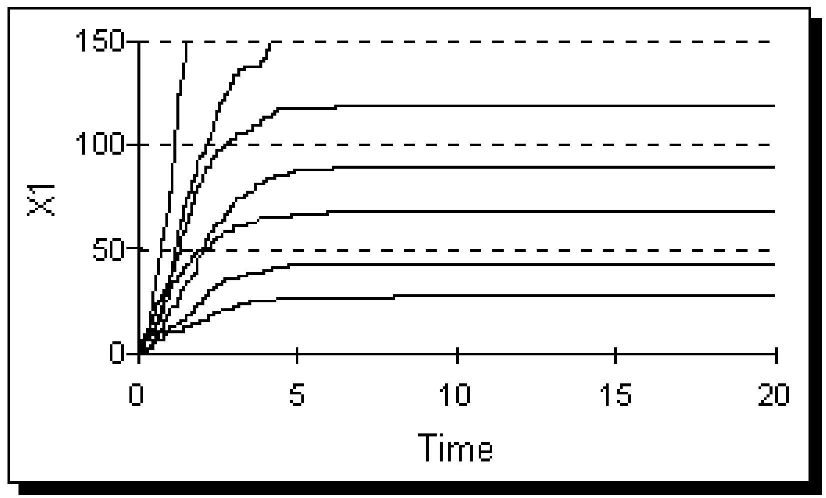






Differential equation model:

$dX/dt = f$
$f = r*X*\exp(-rt)$
$X(0) = 2$



Poisson simulation model:

$dX/dt = f$
$f = Po(r*X*\exp(-rt)*dt)/dt$
$X(0) = 2$

Very different realisation results (end levels).
Small variations around the trajectories.

Predator - prey model

(Lotka -Volterra's equations)

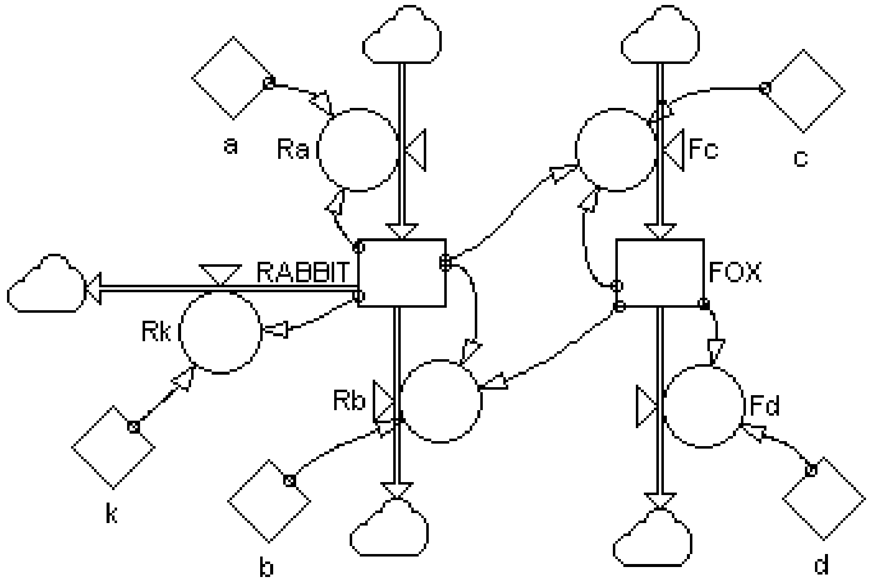
- Two species: The prey and the predator (e.g. Rabbits and Foxes)
- The preys (X) increase by reproduction and decrease by meeting the predators. The preys also have mutual competition.
- The predators (Y) increase by eating preys and a certain fraction dies per time unit.
- There are three possible equilibria: X=Y=0, Only X, and Both X & Y.

Deterministic model

File: Volterr1.sim.

dX/dt = Ra - Rb - Rk
X(0) = 60
Ra = ax
Rb = -bXY
Rk = -kXX

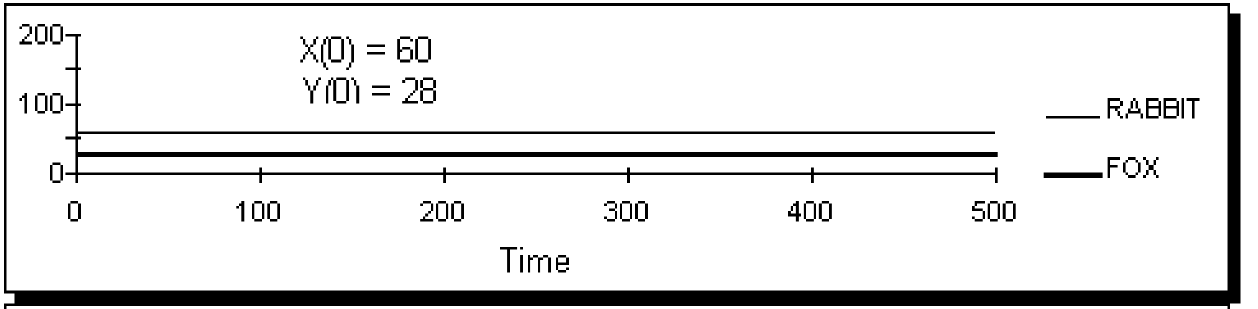
dY/dt = Fc - Fd
Y(0) = 28
Fc = cXY
Fd = -dY

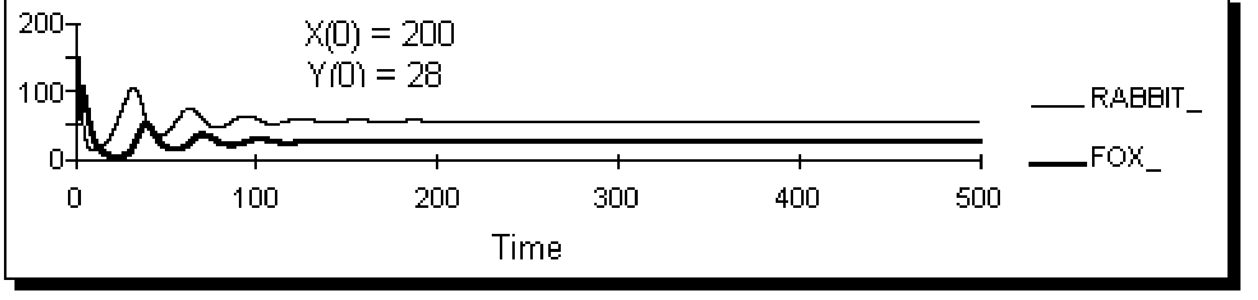


STATIONARY VALUES:
I: X=0, Y=0

II: X=a/k, Y=0

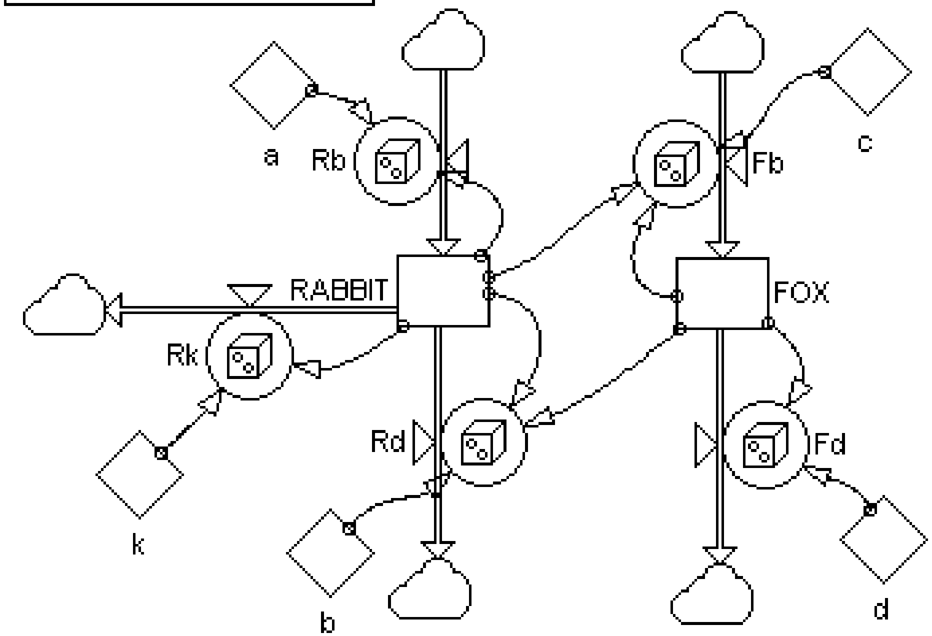
III: X=d/c, Y=(a-kd/c)/b





Poisson model

File Volterra.sim.

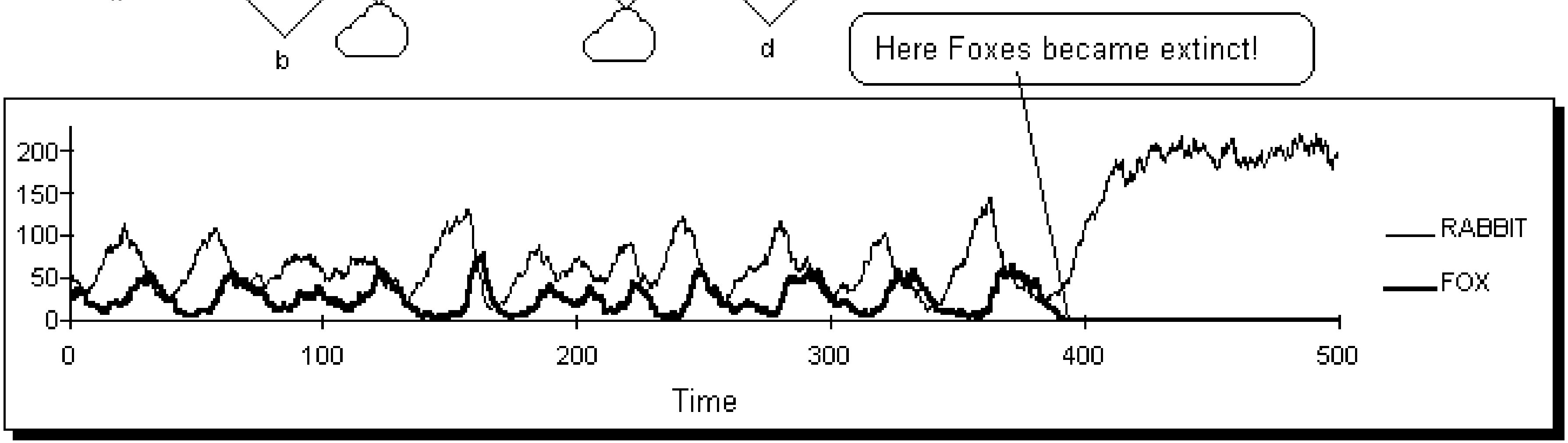


Volterra equations:
a=0.2, b=0.005, c=0.005, d=0.3, k=0.001
gives stat. values: X(0)=60, Y(0)=28

STATIONARY VALUES:
X=0, Y=0

X=a/k, Y=0

X=d/c, Y=(a-kd/c)/b



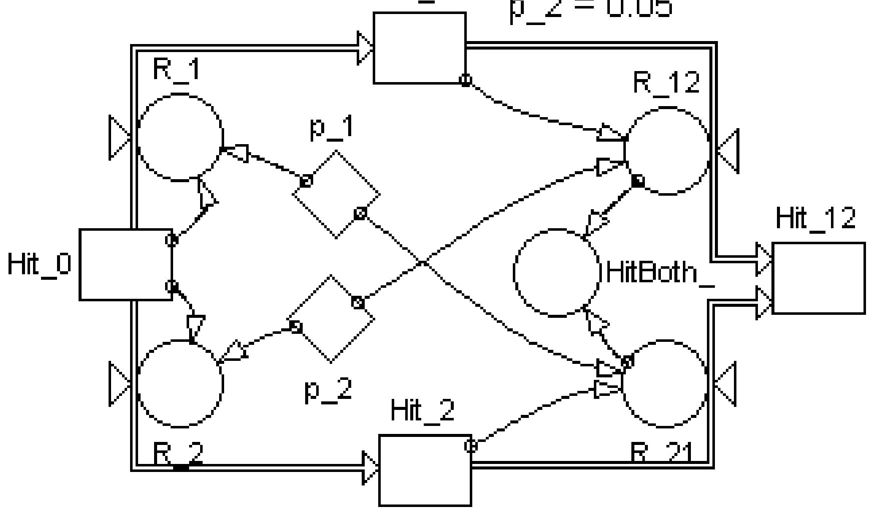
Conclusions: 1) Stochastics excites dynamics!
2) A stochastic model may switch to another dynamic mode!

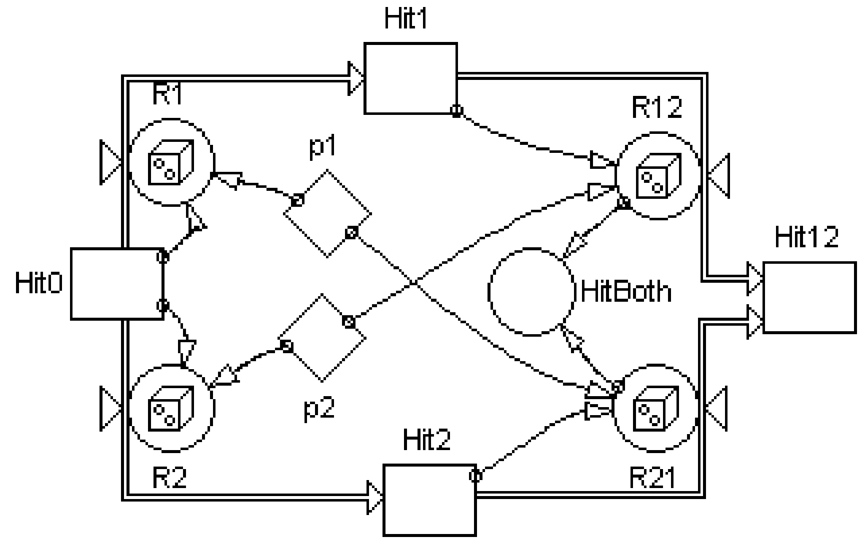
A multi-hit model for e.g. cancer

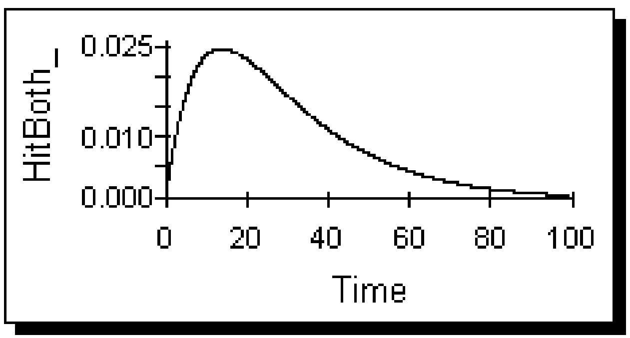
- Many robust systems will not fail until a number of independent events happen. One example is cancer which requires a number of destructive hits in the genome of a cell.
- In a multi-hit model the different hits can happen in any order.
- When a cell has all n destructive hits it can no longer control its growth. It becomes a cancer cell!
- A human has about $5 \cdot 10^{14}$ cells. What is the probability that a cell gets all n hits within a life-time?

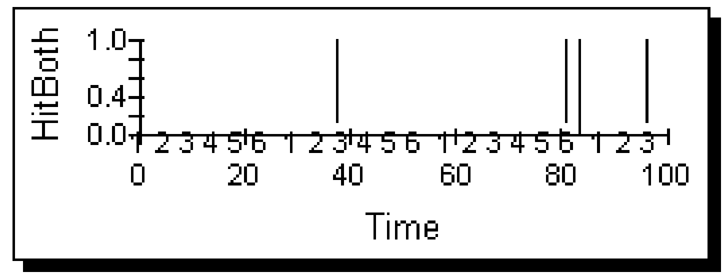
File: MHit2.sim.

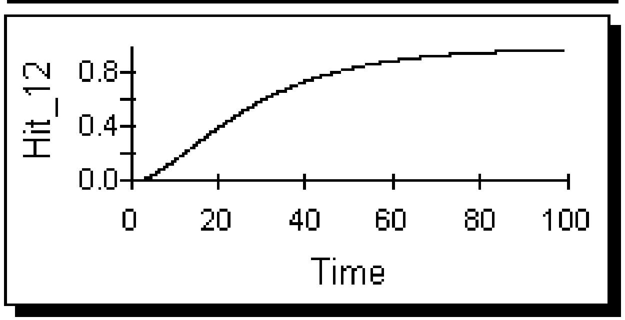
1 cell: p_1 = p_2 = 0.05











Making the model stochastic, we can study the risk that a person with, say, 1000 milion cells of a certain type gets the two hits during lifetime. (p1=p2= 0.0000003.

Above 6 simulations: 4 get the 2 hits.

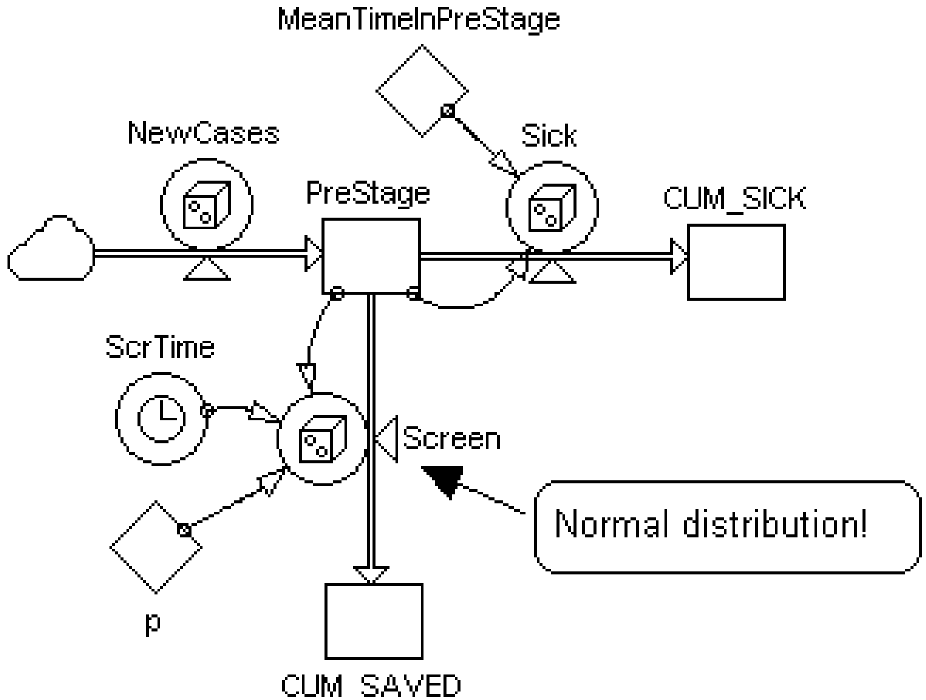
Screening for a disease

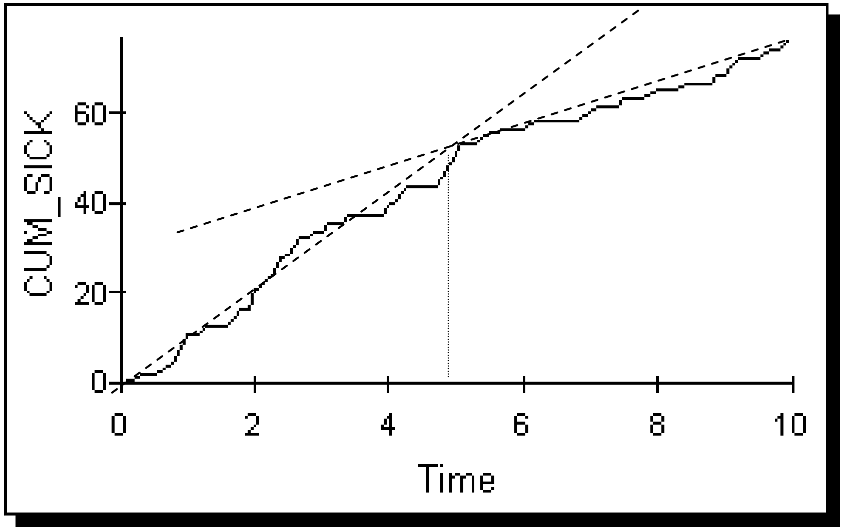
- Suppose you have a population where n individuals have a pre-clinical disease.
- The population is *momentarily**) screened for the disease with a test that has a certain sensitivity p (= probability to detect a true case).
- Since each, of n sick individuals, is detected with the probability p you find (and eliminate) Bi(n,p) individuals.
- For large numbers (say n>15) you may use the normal distribution $N(n\cdot p, \sqrt{n\cdot p\cdot(1-p)})$.

*) Screening may take place during a period of time, but when studying the effects you may adjust the individual processes to "Date of screening".

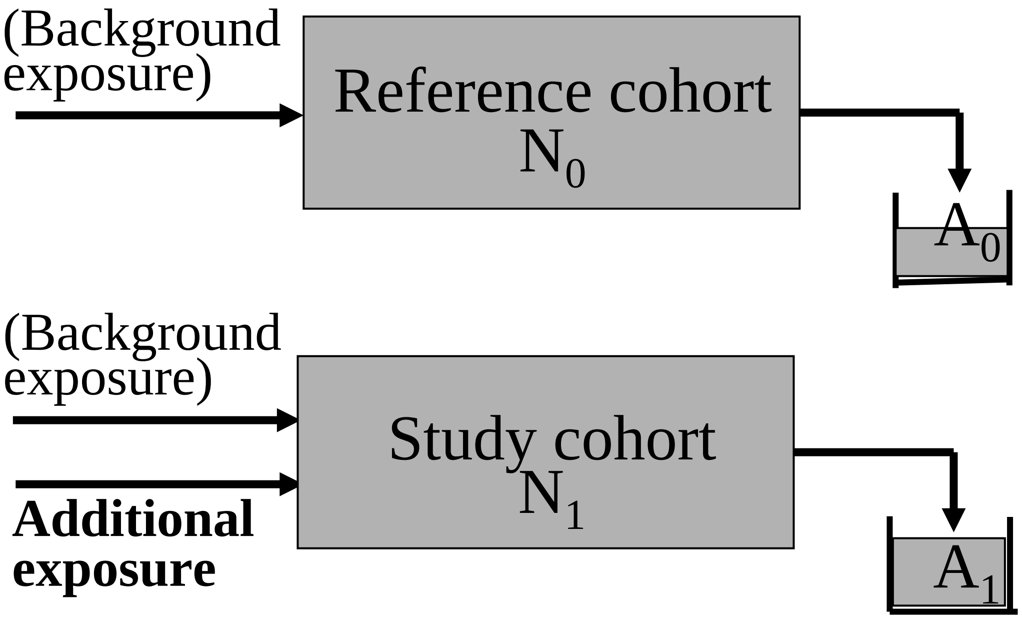
File: Screen.sim
Screening requires Binomial distribution.
TestResult = pos (p) or neg (q) and n testade ger Bi(n,p)
E(X)=np Var=npq

For large n (n>10) the Normal-approximationen
N(np,Sqrt(npq)) is good.





A study and a reference cohort



Static approach: 2 * 2 table

	Exposed	Not exposed
Sick	A_1	A_0
Not sick	N_1-A_1	N_0-A_0
Total	N_1	N_0

Cum. Incidence: $CI_1 = A_1/N_1$

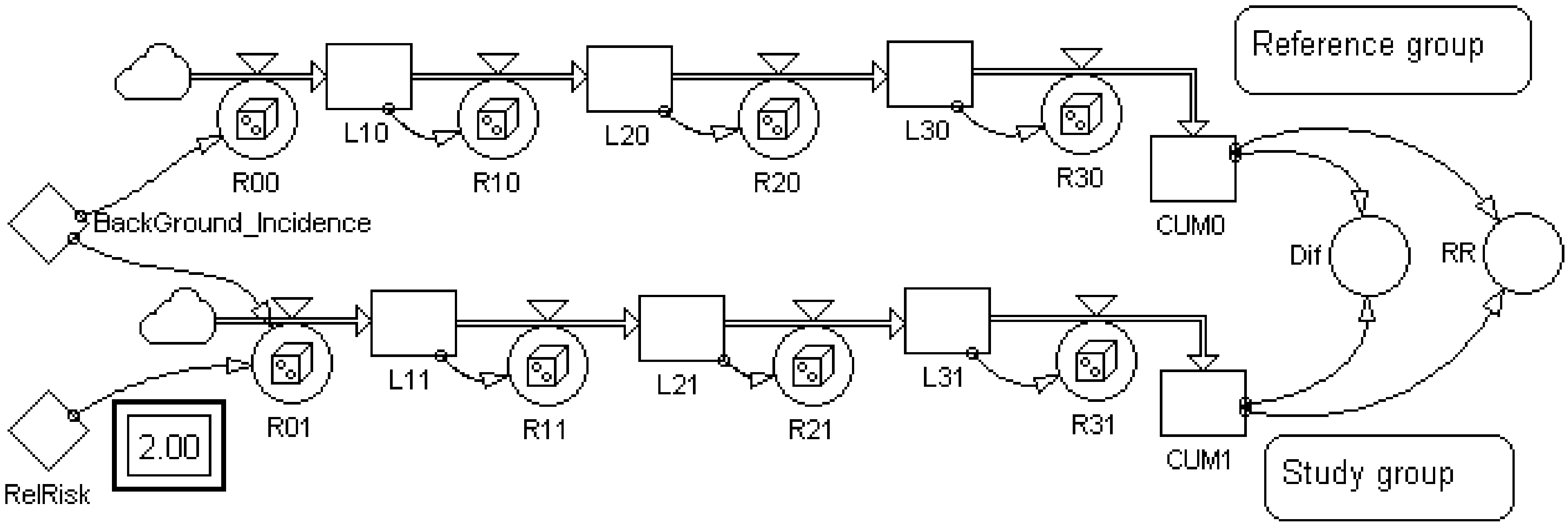
$CI_0 = A_0/N_0$

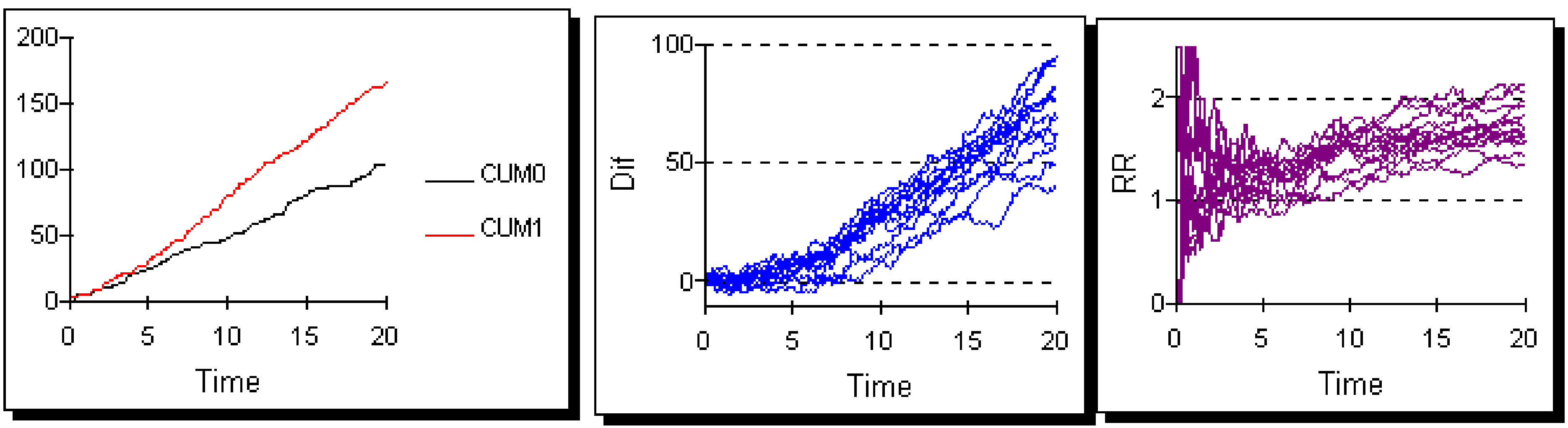
$$\Delta CI = CI_1 - CI_0 \pm 1.96 \cdot \sqrt{CI_1(1-CI_1)/N_1 + CI_0(1-CI_0)/N_0}$$

Problems: Need a dynamic model because:

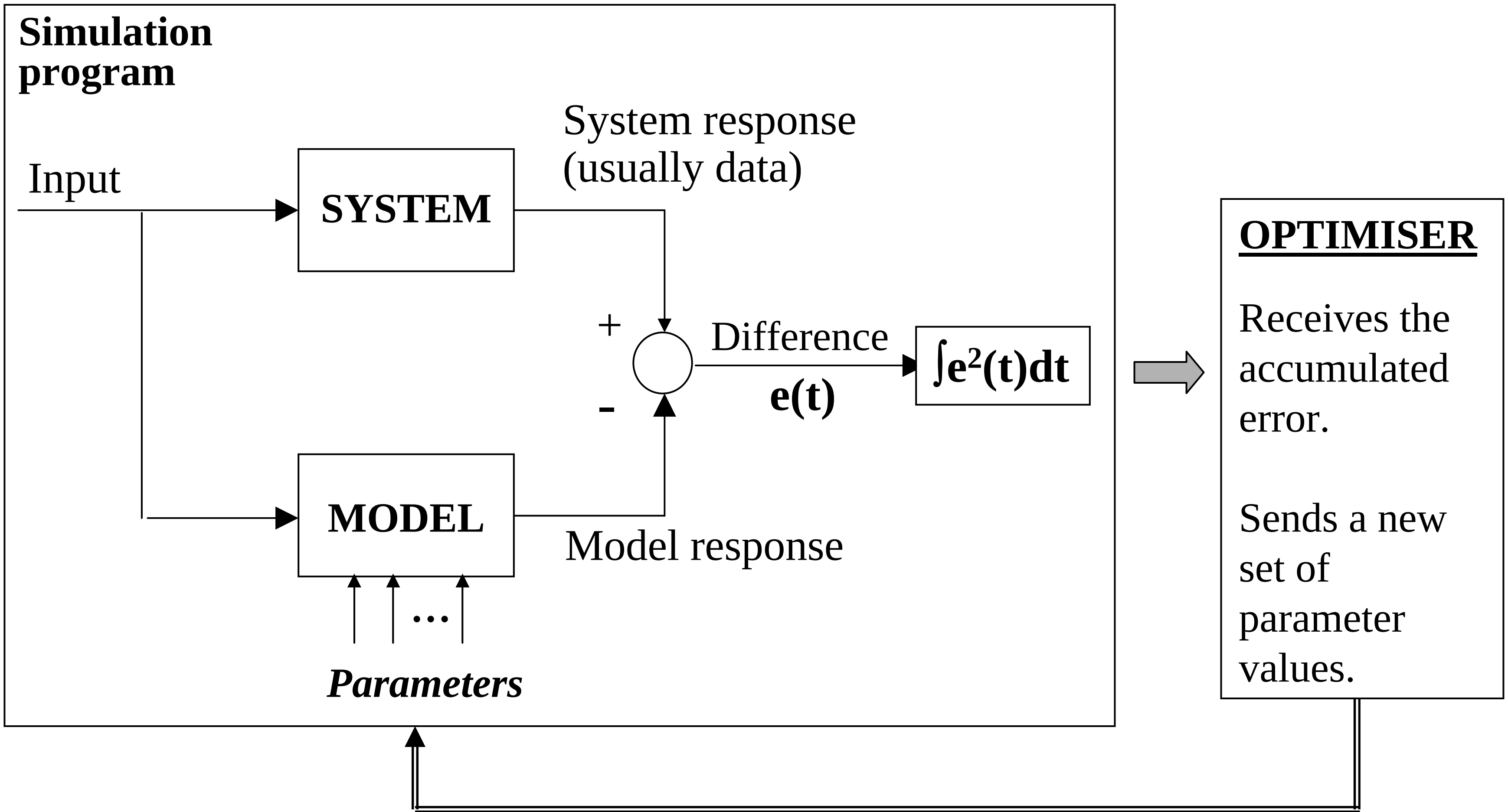
- Want to understand the process.
- Realistic initial values. (The “pipe-line” is not empty at start.)
- Find invariants like Relative Risk of additional exposure & sojourn time.
- Additional exposure may not have reached steady state.
- Follow-up time may be too short. Then biased estimates.

File: StdyRef2.sim 000313.
Poisson simulation of exposure of a study and a referencel cohort. BackGr_Inc=5, T=6, RelRisk=2.





Parameter estimation



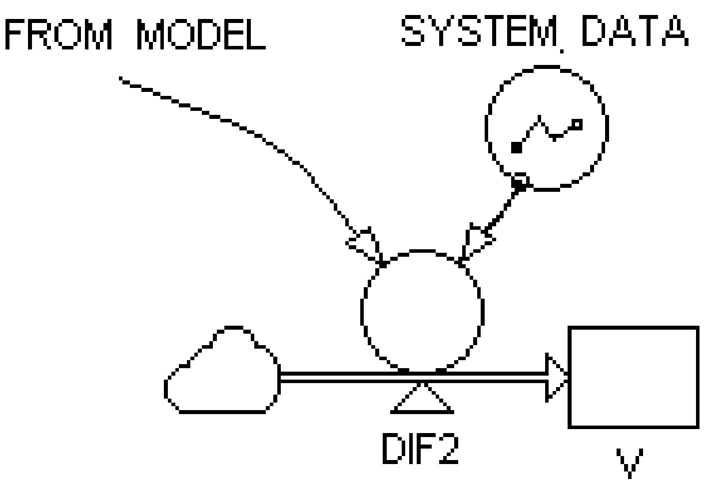
Model fitting is finding that model whose behavior maximally fits the system's.

Working procedure

- Get data of the system behaviour over time.
- Build a realistic structure (use your knowledge of the system).
- Fit the parameters to minimise the difference between system data and model in a least square sense (Parameter estimation).
- [Try different structures (e.g. number of internal states) and fit again.]
- Choose the best [structure and] parameter values.
- Revise the optimal model to its Poisson correspondence. Simulate and calculate statistical estimates

Fitting a logistic model to system data

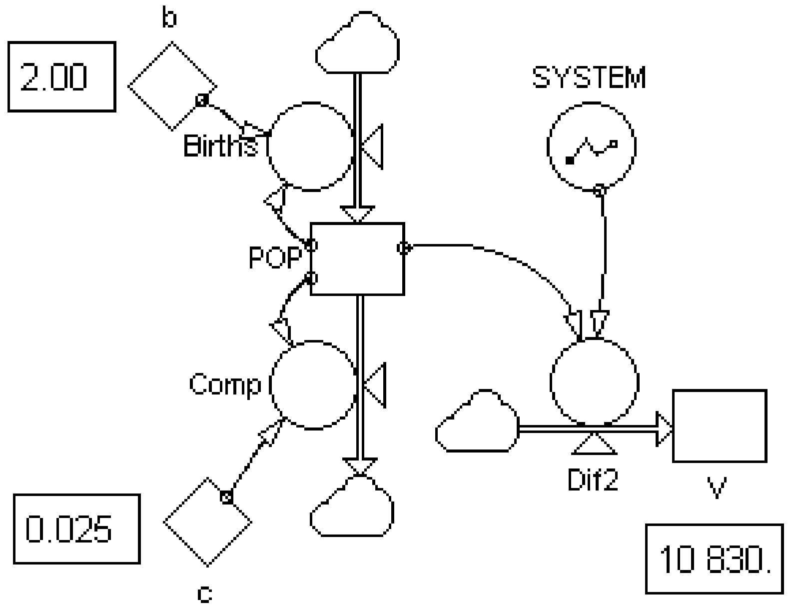
- You have measured a **system**’s behaviour over time.
- A logistic **model** ($dx/dt=b\cdot x-c\cdot x^2$) seems appropriate, but the values of parameters b and c are unknown.
- The difference between system and model is calculated and squared: DIF2 = (SYSTEM-MODEL $)^2$.
- This squared difference (DIF2) is cumulated over time in, an originally empty, state (V).
- Find the combination of b & c that minimises V.

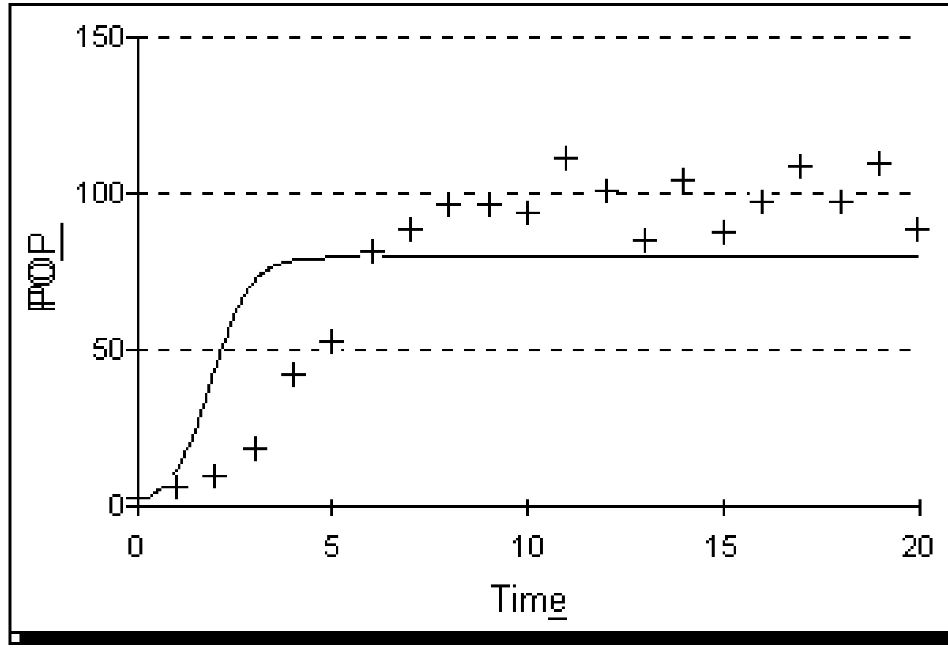


File: PEstLogA.sim

Number of cells counted from the system.

Time	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SYST	2.0	6.0	9.0	18	42	52	81	88	96	96	93	111	100	85	104	87	97	108	97	109	88

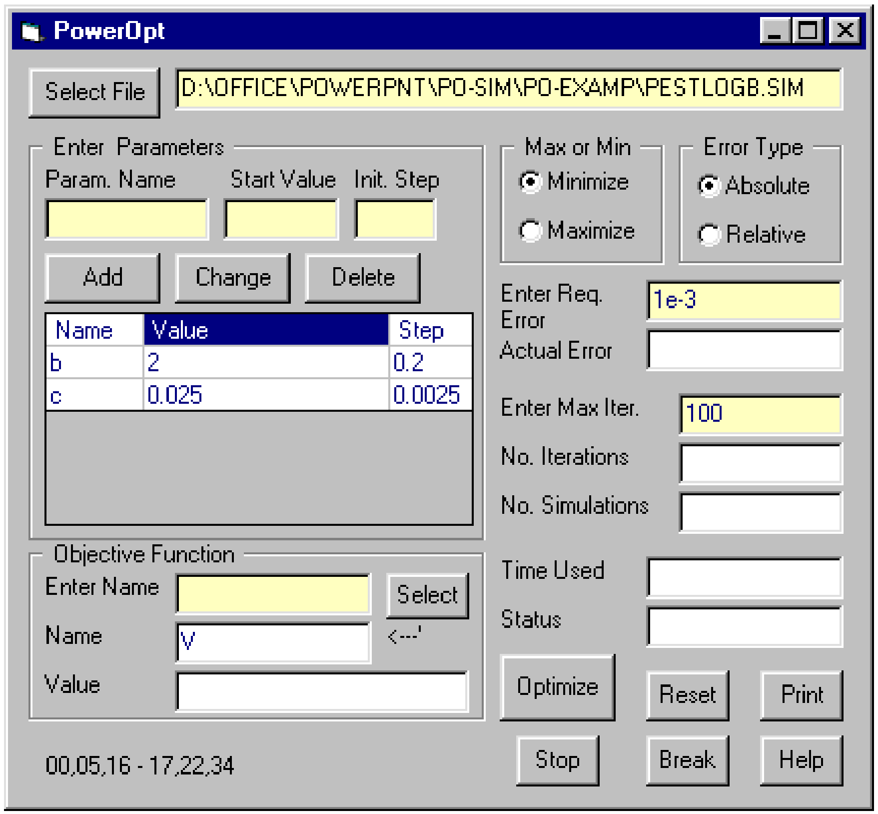


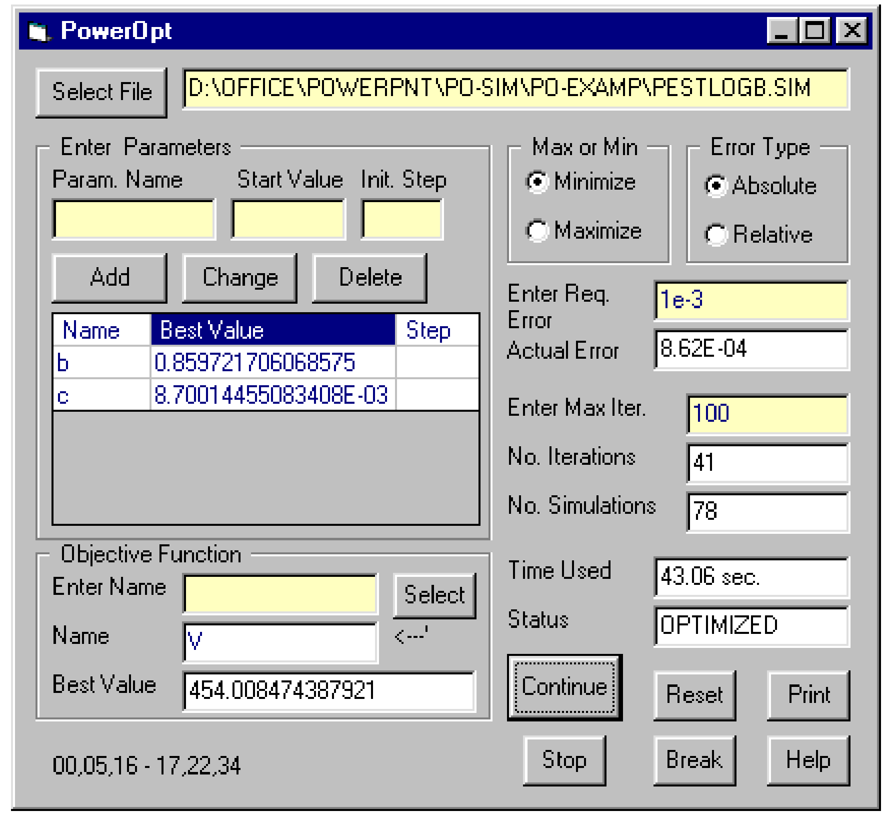


The difference between SYSTEM and model (POP) is calculated and squared. At each time step this squared error (DIF2) is cumulated into V (which was empty at start of simulation). V therefore gives a measure of how well SYSTEM and MODEL fit. We want to find that combination of the parameters b and c that minimizes V.

PowerOpt with starting values b=2 and c=0.025. V will be minimised.

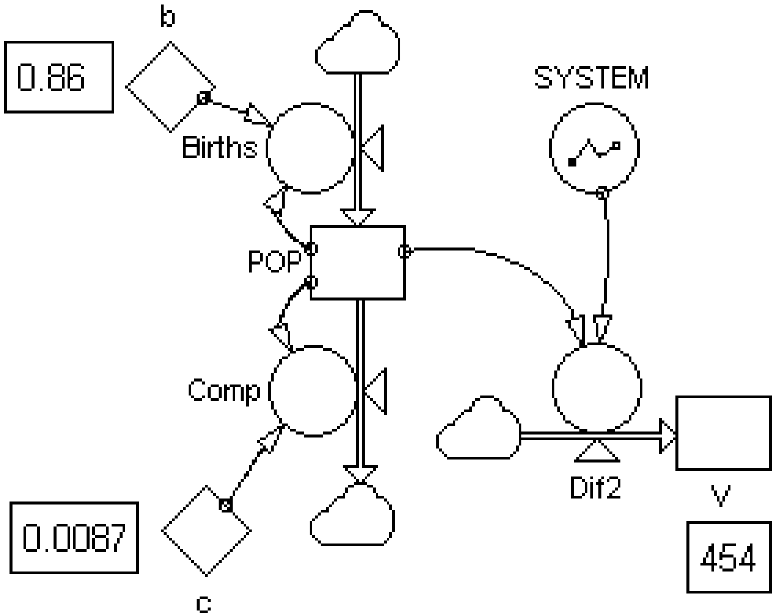
B=0.86 and c=0.0087 will minimise V to 454 after 78 simulations (43 seconds).

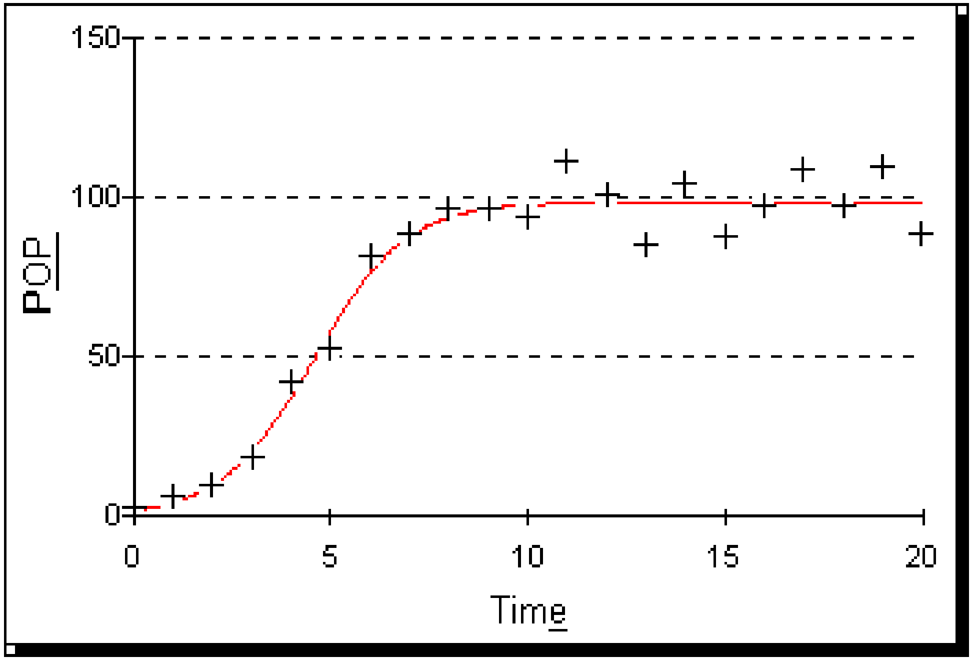




The best fit - V is minimised!

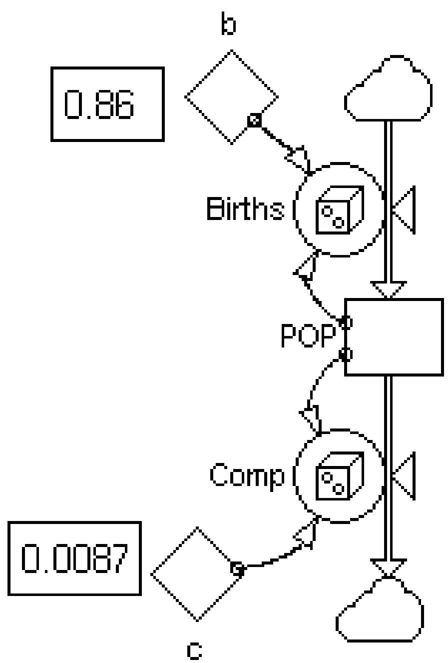
File: PEstLogB.sim

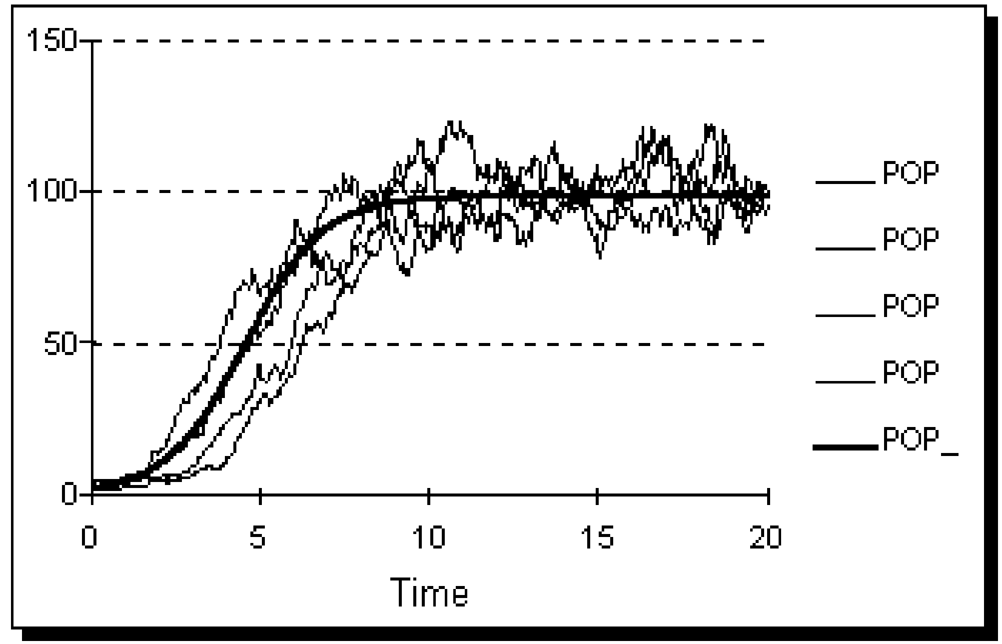




Variations around the best fit model!

File: PEstLogC.sim.





• **By running the Poisson model N times we can estimate the variations around an average.**

• **If we want toe estimate the variations in the parameters b and c we fit b and c to each of the N stochastic curves. From the N estimates of the pairs {b, c} we get their variations.**

Statistical estimates from a Poisson simulation

Estimates

- For linear models you get "unbiased" estimates from the deterministic model. For non-linear models the model can be run N times to estimate a quantity.
- Note: Often an estimated quantity is a function of time rather than a constant. This is clearly displayed when using simulation!

Uncertainty estimates

- Variance, confidence intervals, P-values, hypotheses tests etc. may be estimated from N simulations.

Why Poisson simulation?

- **Dynamics and stochastics together.**
  - Stochastics excites dynamics.
  - The model may switch between modes (cfr. Volterra).
  - Adds statistical estimates to a dynamic model.
- **Complex realistic models.**
  - Structure (not restricted to named statistical distributions).
  - Initial values
  - Input (e.g. from empirical data)
  - Focus on invariants (rather than time varying estimates)
- **Explanation model rather than a black box model.**
- **Gives new possibilities** to discriminate between determi-nistically similar models (cfr. Logistic/Gompertz).
- **Easy to use, easy to understand & fast.**

References

• *Gustafsson L*. Poisson simulation - A method for generating stochastic variations in continuous system simulation. Simulation, May 2000.

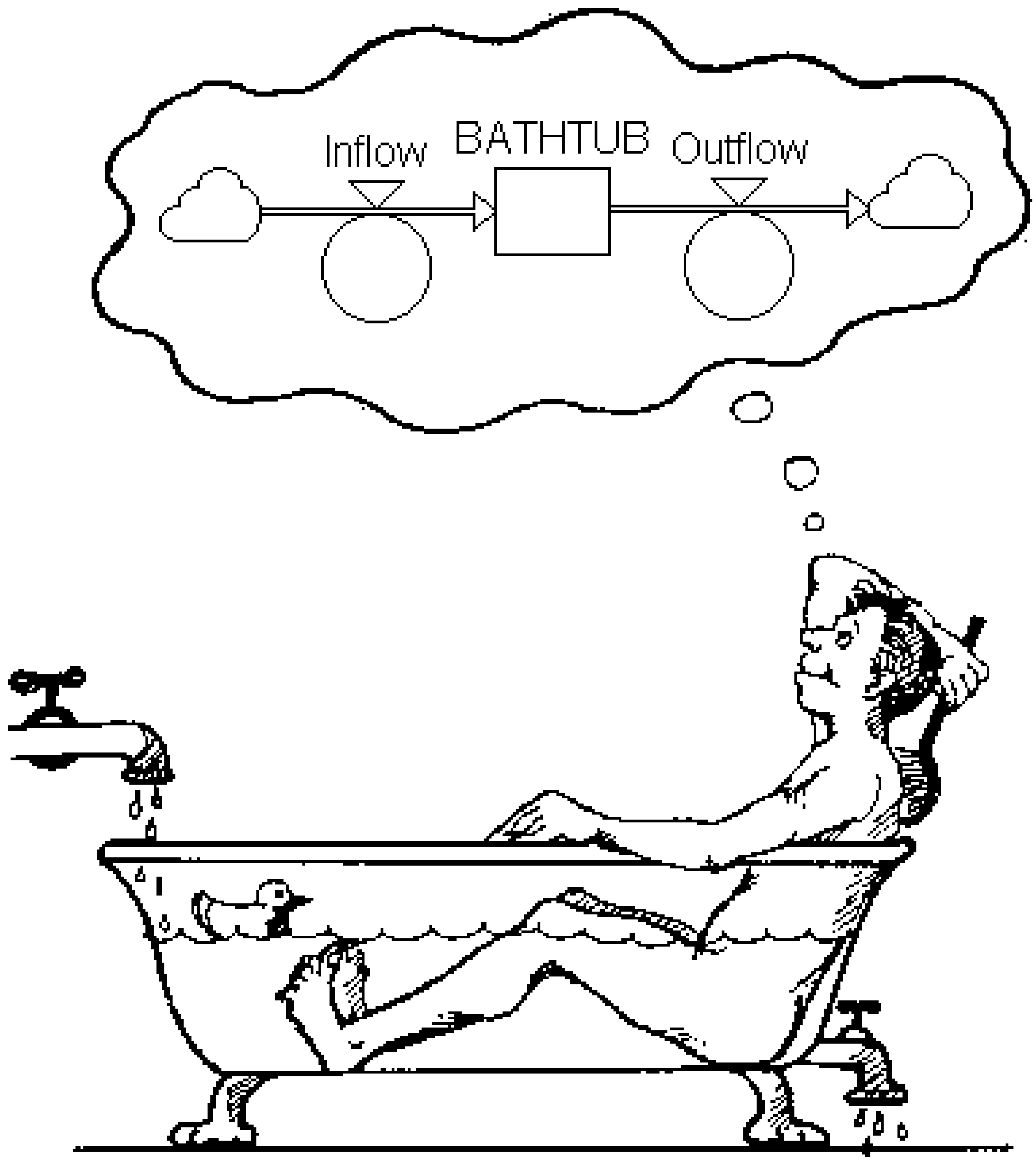
• Powersim© Reference Manual, Powersim Press 1996. Powersim Corporation, 1175 Herndon Parkway, suite 600, Herndon, VA 22170, USA.
(A free demo is found at: **http://www.powersim.com**. Note that Δt is called “TIMESTEP” in Powersim.)

• PowerOpt. An optimiser for Powersim by Leif Gustafsson.

Planned future work

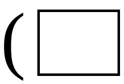
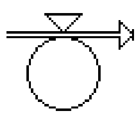
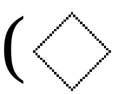
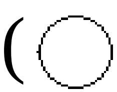
- Wanda - a device to control N simulations, collect, analyse and present statistical results.
- Improved methods for statistical estimates.
- Theory behind the “initial value problem” for dynamical systems in steady state.
- Development of other integration algorithms than Euler for Poisson simulation.
- Need for antithetic random numbers?
- Poisson simulation for DES models [easy to vary $\lambda(t)$ in Po($\lambda(t)$].
- “Extended Poisson simulation” for discrete space (but continuous time).
- A number of practical projects based on Poisson simulation.

APPENDIX. The dynamic diagram



The dynamic diagram

The dynamic diagram is an alternative way to present a system of differential and algebraic equations. It has significant pedagogic merits and is often used by biologists, ecologists, medical people, economists and others where the mathematical background is not as large as for e.g. technicians or physicists.

- A **state** is represented by the "bath tub" symbol () which holds the amount or number of objects.
- The content in the state changes because of **in-** and **out-flows** ().
- Other symbols are **constants** () and **auxiliary quantities** for algebraic computations ().

- To show the limits of the model we let the flows origin from or end in a **cloud** symbol ( ). (We are not interested in where the rabbits go after death!)
- Everything affecting a quantity is displayed by single **arrows** ( ) from the affecting quantity to the affected one.
- A **dice** ( ) in a flow (or auxiliary) symbol denotes that the value of the flow rate (for each time step) is drawn from a probability distribution.
- A clock ( ) in a symbol denotes that it is a time function.
- The **ghost** symbol ( ) is used as a copy of the quantity (instead of using long arrows).
- The exact mathematical expression is defined after double clicking the quantity (but is not seen in the diagram).

The End

Demonstrations

- Radioactive decay - (NegExp.sim)
- Logistic growth - (Logistic.sim)
- Gompertz' growth - (Gompertz.sim)
- Volterra's equations - (Volterra.sim, Volt_SS.sim)
- Demographic model - (Demograf.sim)
- Multi-hit model - (MHit2.sim, Mhit2_Po.sim, MHit3.sim)
- Screening - (Screen.sim)
- Model fitting - (PEstLogA.sim, PEstLogB.sim, PEstLogC.sim)
- Study & Control cohorts - (StdyRef.sim, StdyRef2.sim)
- Life tables - (Life_Tab.sim)
- Epidemic model - (XXX.sim)
- M/M/1-queue - (MM1_que.sim)
- Infection model - (Potatis.sim)
- AIDS model - (AIDS.sim)
- Potatoe virus model - (Potatis1.sim)
- ...

CONTINUOUS versus DISCRETE EVENT SIMULATION

	Continuous Simulation	***Discrete Event Simulation***
VIEW	• *Macro*: Homogenous flows	• *Micro*: Single actors
RANDOM	• Usually no randomness	• Usually crucial
TIME	• Continuous time	• Discrete events
DYNAMICS	• System of differential equations	• Programmed behaviour (Flowchart)
STRUCT.	• States and flows	• Actors come and go
LANGUAGE	• Program for integration of differential equations, Functions ...	• General program lang. + Discrete event handler, Actors, Resources ...
RESULTS	• Time functions x=f(t)	• Cumulated statistics

- **Discrete Event Simulation** - Individual entities (micro view approach).

- **Poisson simulation** - States holding numbers of entities (macro view approach).

You don’t use the transient information.

Studies of **Cause** $\Rightarrow$ **Effect** (e.g. Exposure $\Rightarrow$ Disease) is problematic when the relation is a dynamic process!

The cause at time t_0 gives effects over time.

And Effects at time t_1 originates from causes over time!