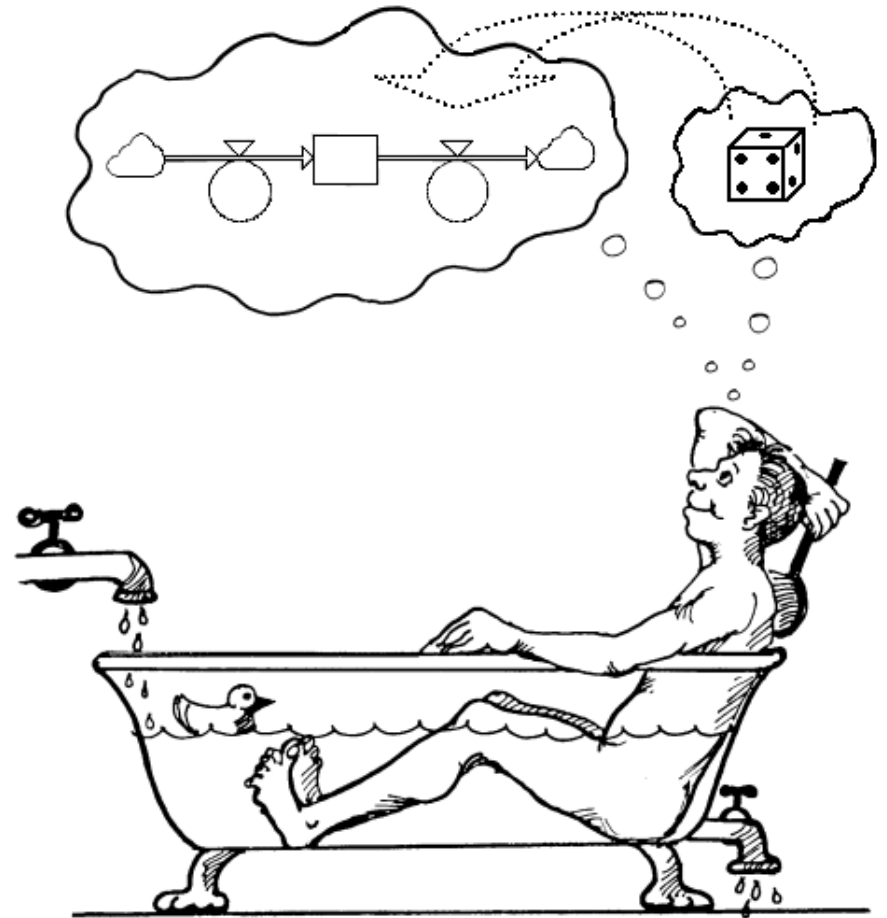# STOCHASTIC POPULATION MODELLING & SIMULATION – especially Poisson Simulation

Part I. Introduction

Part II. Poisson Simulation

Part III. Statistical tools
   & Model fitting

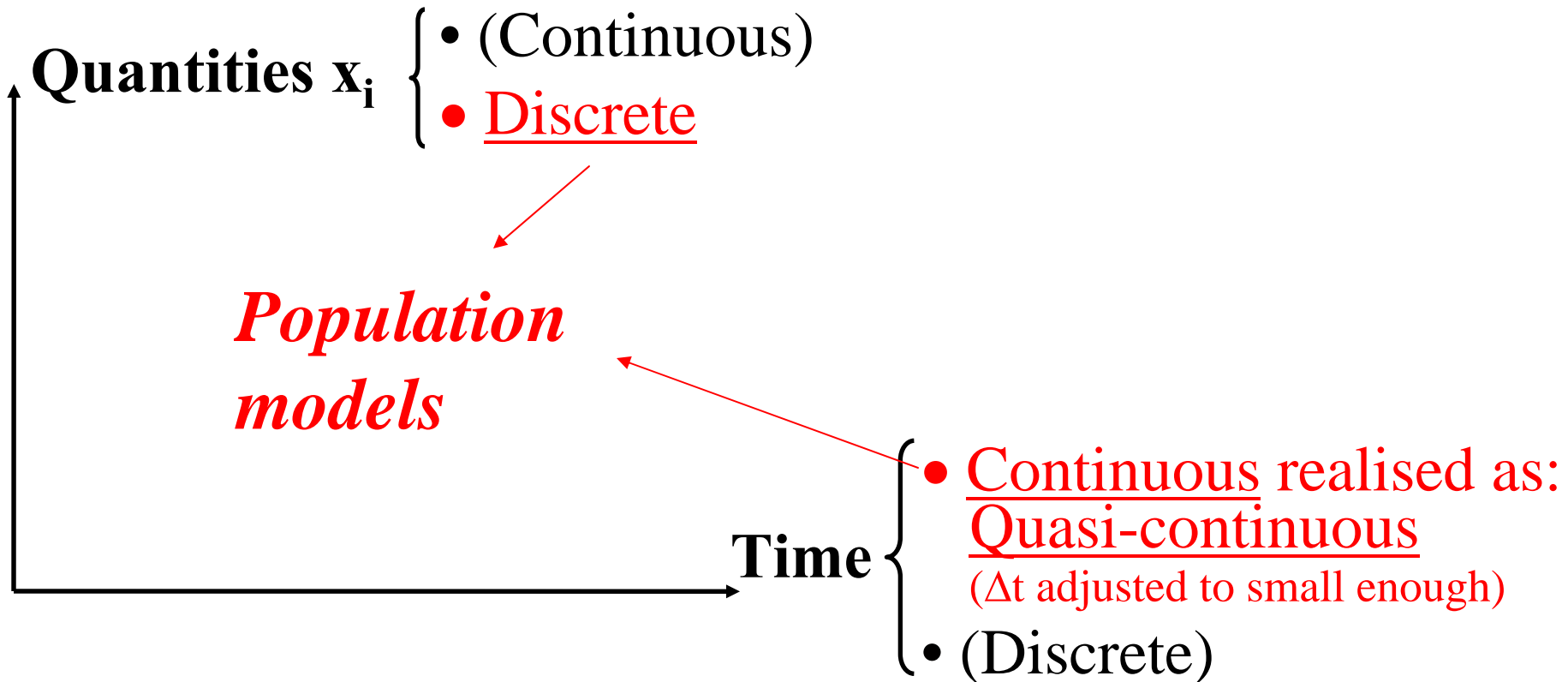**Part IV.** Routine-like modelling
- bad, worse and disastrous examples

Leif Gustafsson©

Part I. INTRODUCTION

Discrete or continuous?

Quantities x_i
- (Continuous)
- Discrete

Population models

Time
- Continuous realised as: Quasi-continuous (Δt adjusted to small enough)
- (Discrete)

Population Models

Population models concern collections of *discrete entities* such as atoms, molecules, genes, cells, humans, animals, plants etc. In population models, the integer number of individuals/entities in a population or sub-population is the main property of interest. The processes with discrete entities develop over *continuous time*.

Population Models

Physics

Biology

Demography

Epidemiology

Ecology

Queuing systems

Production

Etc.

Characteristics: • *Discrete entities* • *Number of entities* ≥ 0

• *Dynamics over continuous time* • *Stochastic events*

• *Results as probability distribution functions*

For complex systems analysis is seldom possible – then use simulation!

SIMULATION _= model experiment_

Influence → **SYSTEM** → System behaviour

1) ↓   1) ⬇   3) ⬆

Input data → **MODEL** → Output
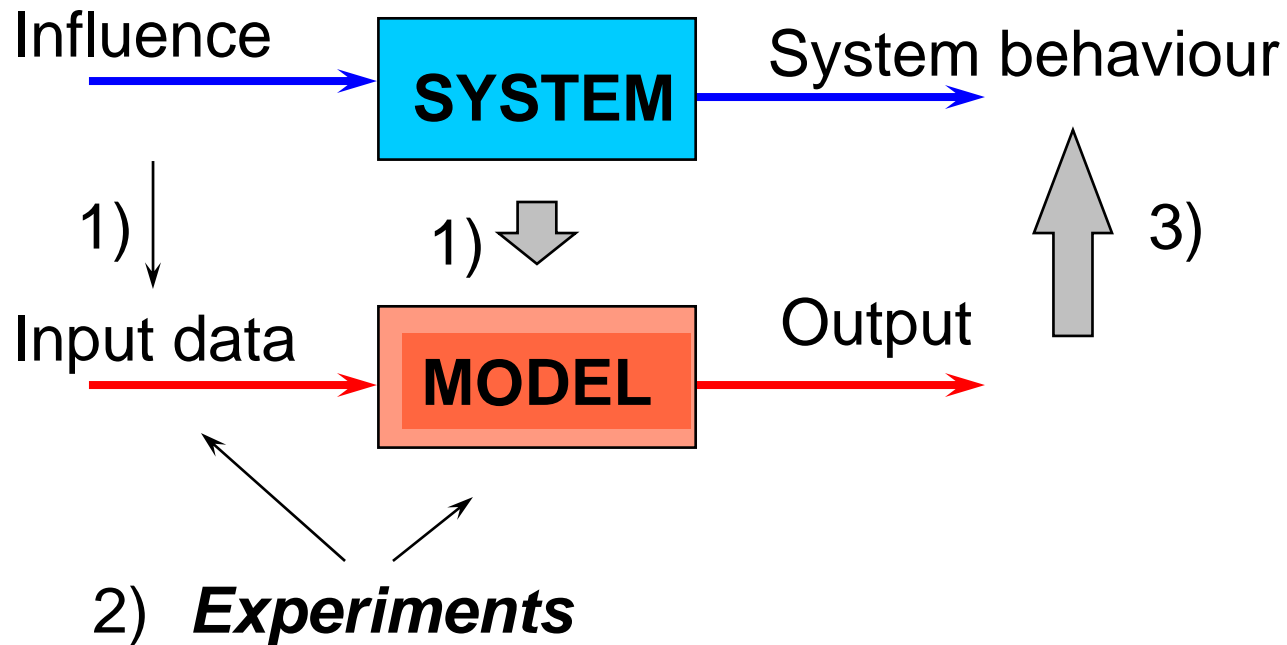
2) *Experiments*

1) Describing the system with a model (and input with corresp. data).
2) Performing the experiments.
3) Conclusions about how the system should behave.

Different types of models

	Static	Dynamic
Stoch-astic	**Statistical models**	**Dynamic & stochastic models**
Determi-nistic	**Algebraic models** **(+ - * /)**	**Differential equ. models**

Many biological, medical, agricultural, etc. systems belong here!

Types of dynamic & stochastic simulation

	Static	Dynamic
Stoch-astic	• **Monte Carlo simulation** (discrete event approach)	• ***Discrete Event Simulation*** • ***Poisson Simulation*** • ***Markov Simulation*** • ***(ARMA models etc.)***
Determi-nistic	• e.g. Experiments on a spreadsheet	• ***Contin. System Simulation*** - Quasi time-continuous approach • *Discrete-time approach*

✓ Real systems develops continuously over time!
✓ A time-discrete model $\Rightarrow$ parameters become func. of the time-step!
✓ Randomness should be modelled – not just added!

Could we use only DYNAMIC or only STATISTC methods? (in e.g. epidemiology)

No! Stochastic variations excite the dynamics! And dynamics changes the conditions for the stochastics!
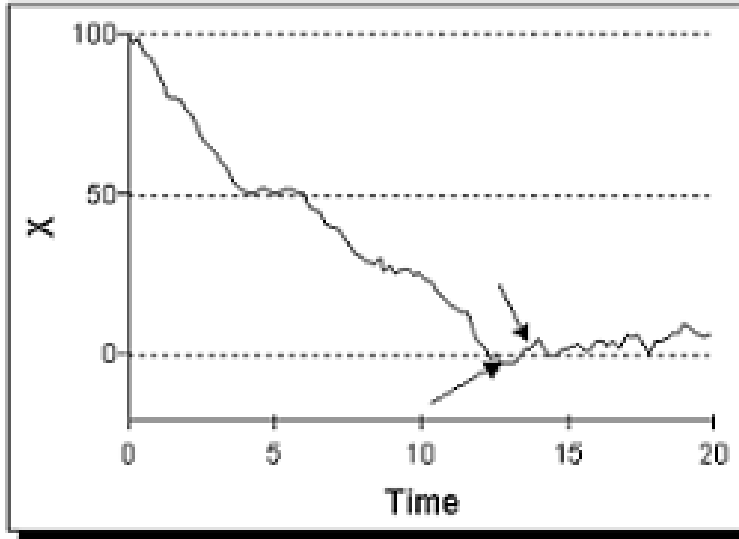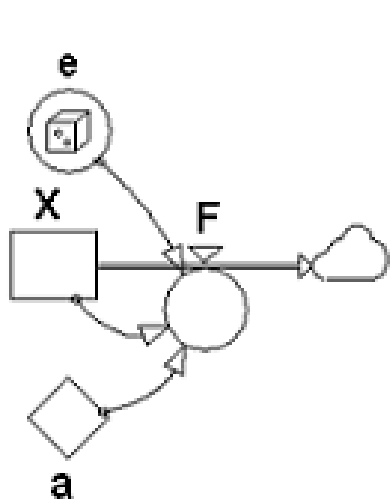
No! We need statistical estimates (mean, variance, correlation, confidence interval, hypothesis test etc.)

*****

- *The same model* should handle *both dynamics and stochastics* in an integrated way.

- Also stochastics must be *modelled* ! (You can't just add random noise! – See next frame!)

- Only for linear models the expected value of the stochastic model equals the results of the deterministic one! (And even then there are dangerous pitfalls!)

Just adding noise ???

Ex: Radioactive decay with dx/dt=−ax+e(t) ➡ **Artefacts:**

(where *e(t) is Normal distributed white noise.*)

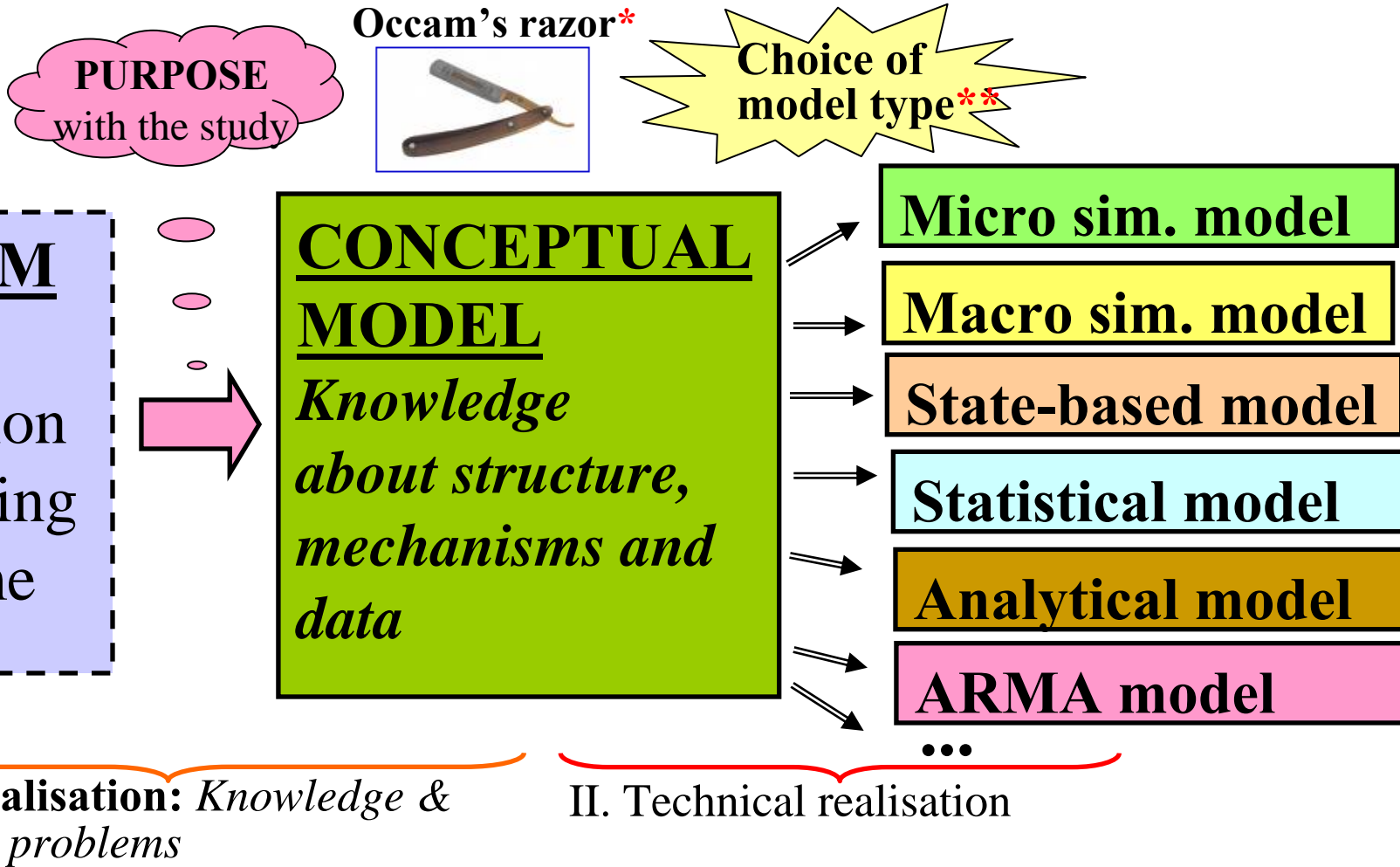- Non integer atoms!
- X never levels out!
- Variations don't decrease!
- New atoms may be created!
- Number of atoms may become negative!
- When a quantity x_i becomes negative – equations including x_i may get mad and produce rubbish!
- Results depend on Δt!
- The shorter Δt the more energy in the noise[8]
- …

Also stochastics must be *modelled* ! (You can't just add random noise!)

System, Conceptual Model and Realisations

Occam's razor*

PURPOSE with the study

Choice of model type**

SYSTEM Studied population developing over time

CONCEPTUAL MODEL *Knowledge about structure, mechanisms and data*

Micro sim. model

Macro sim. model

State-based model

Statistical model

Analytical model

ARMA model

...

I. Conceptualisation: *Knowledge & feasibility problems*

II. Technical realisation

The Conceptual Model *is our best understanding of the system from observations, measurements., deductions, beliefs etc.*

 *) **"A model should be made as simple as possible – but not simpler!"**

**) *You must know what you can do with different types of models. "If you only have a hammer – every problem looks like a nail!"*

Consistency - free from contradiction

SYSTEM
Studied population developing over time

CONCEPTUAL MODEL
Population of individuals. (Stochastic state transfers.)

Micro model

Results

Macro model

Results

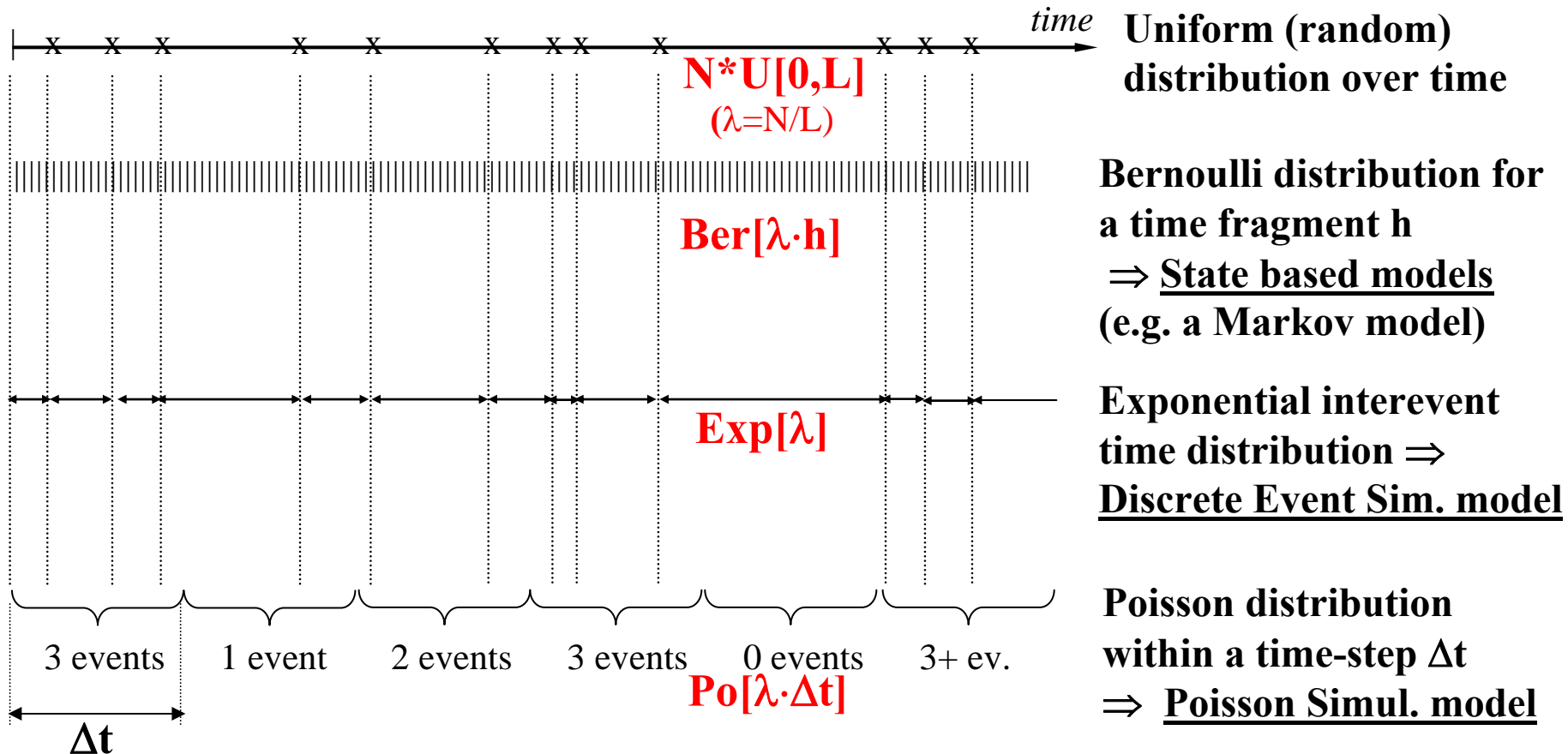**Markov model**

Results

Consistent results ?
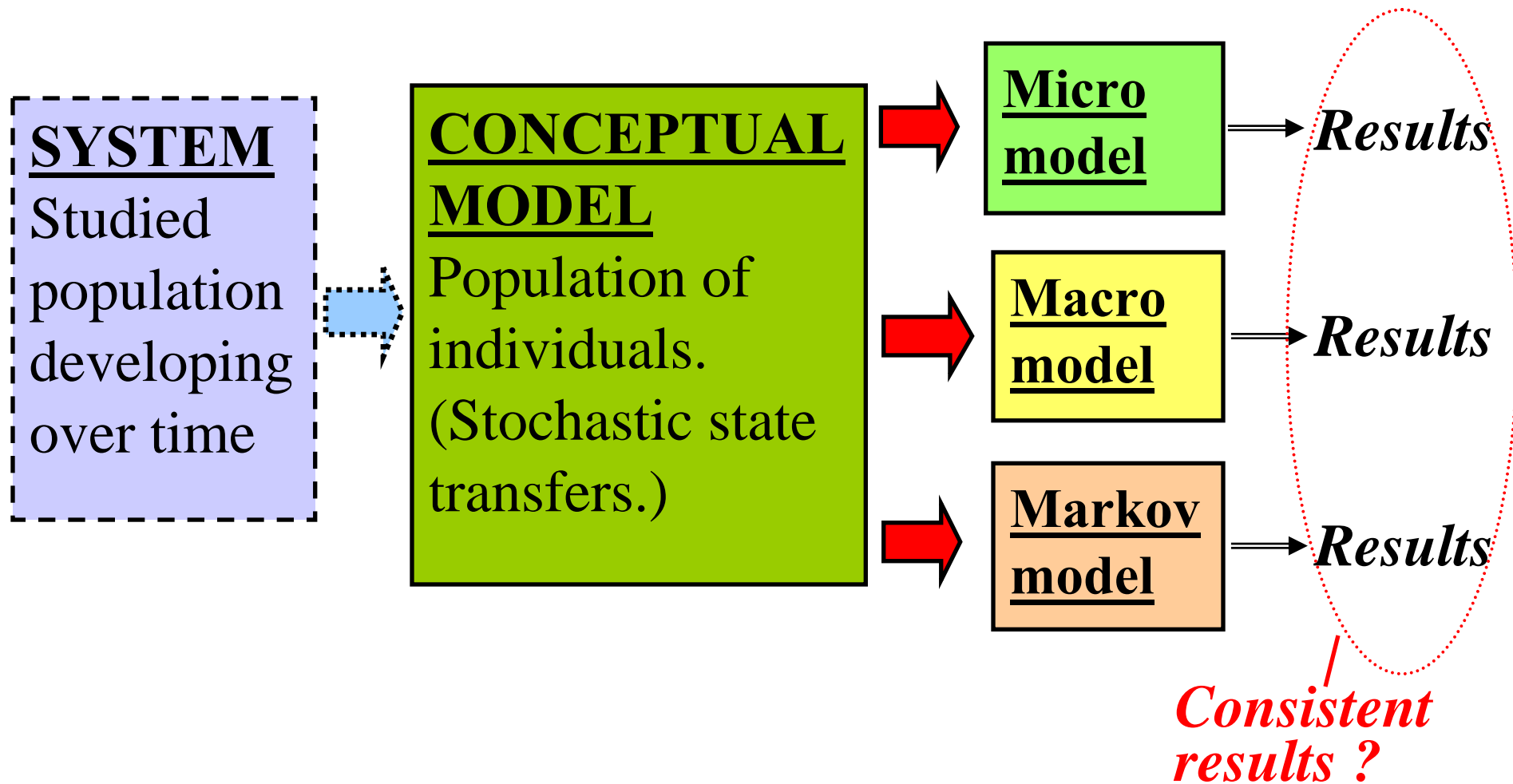
Consistent results from stochastic micro-, macro- & Markov models means that their results should have the same statistics (mainly: probability distribution functions).

How to deal with *EVENTS - 1*

(This leads to fundamentally different approaches!)

Time-handling principles: (Event The intensity is here λ EVENTS/time unit)

N*U[0,L]
(λ=N/L)

Ber[$\lambda \cdot$h]

Exp[λ]

| 3 events | 1 event | 2 events | 3 events | 0 events | 3+ ev. |

Po[$\lambda \cdot \Delta$t]

Δt

Uniform (random) distribution over time

Bernoulli distribution for a time fragment h $\Rightarrow$ State based models (e.g. a Markov model)

Exponential interevent time distribution $\Rightarrow$ Discrete Event Sim. model

Poisson distribution within a time-step Δt $\Rightarrow$ Poisson Simul. model

Basic stochastic mechanism

11

How to deal with *EVENTS - 2*

In the last frame we assumed that the intensity λ was constant.

- However, it is enough if the intensity is close to constant during a time-step Δt. Then we can also handle $\lambda = \lambda(t)$.

- Further, Δt must always be short compared with the rate of the dynamic changes. So $\lambda = \lambda(x_i)$ or $\lambda = \lambda(\underline{x})$; where $\underline{x} = (x_1, x_2, \ldots, x_n)$ are also feasible!

Thus: $\lambda = \lambda(\underline{x}, t)$ works in: **Ber[$\lambda(\underline{x}, t) \cdot h$]**, **Exp[$\lambda(\underline{x}, t)$]** and **Po[$\lambda(\underline{x}, t) \cdot \Delta t$]** !

Three consistent approaches - 1

- ## Discrete Event Simulation (DES)
  – Micro description.
  – Based on entities (e.g. individuals) that may be unique in characteristics and behaviour.
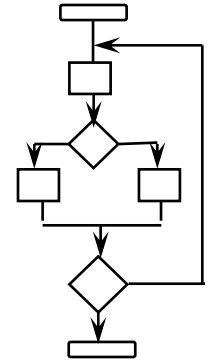
- ## Poisson Simulation (PoS)
  - Macro description. Similar entities are *aggregated* into compartments (state variables). (A stochastic extension of Continuous System Simulation.)
  - Based on compartments and flows between compartments
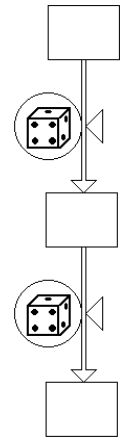
- ## Markov model
  (The mathematical theory is exact when h→0.)
  – Macro description – (individuals are not identifiable).
  – Based on *system states.*
  – *Each and every situation and transition between states must be described.*

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} \dots & a_{2n} \\ \dots & & & \\ a_{n1} & a_{n2} & a_{n3} \dots & a_{nn} \end{bmatrix}$$

13

Three consistent approaches - 2

Never use stochastic modelling if you don't have to!

  If you have to, then:

* **Use DES** when you need to describe *unique* entities that should be identifiable

* **Use PoS** when you can *aggregate* many entities into a number of compartments

* **Never use Markov models for simulation** – There is no case where this technique is superior or even equally good.
  (Markov theory has a great theoretical value in the field of Stochastic Processes.)
    – Gives huge models (unless <u>one</u> entity is studied, like in single channel models)
    – Slow execution
    – Bad mental visualisation
    – Very problematic to modify or extend.
    – Rigid to change (especially awkward for optimisation and model fitting
    – No way of adjusting size of the time-step
    – Illusory simple but full of pitfalls. Attracts people with little mathematical understanding and often results in very bizarre results.   - Etc.

Three consistent approaches - 3

Example: _Compartments_ or _Markov states_ ?

Represent just 10 entities in only 3 possible situations (compartments)!

Poisson Simulation
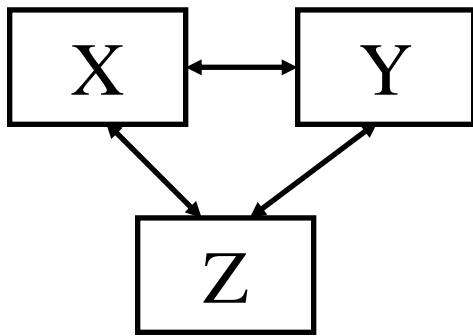
$$dx/dt = F_{yx} - F_{xy} + F_{zx} - F_{zx}$$

$$dy/dt = F_{xy} - F_{yx} + F_{zy} - F_{yz}$$

$$dz/dt = F_{xz} - F_{zx} + F_{yz} - F_{zy}$$

$$x(0)=..,\ y(0)=..,\ z(0)=..$$

Markov model: 66 states:

(The matrix has 66x66 transition probabilities)

(0,0,10)	(1,0,9)	(2,0,8)	(3,0,7)	(4,0,6)	(5,0,5)
(0,1,9)	(1,1,8)	(2,1,7)	(3,1,6)	(4,1,5)	(5,1,4)
(0,2,8)	(1,2,7)	(2,2,6)	(3,2,5)	(4,2,4)	(5,2,3)
(0.3,7)	(1,3,6)	(2,3,5)	(3,3,4)	(4,3,3)	(5,3,2)
(0,4,6)	(1,4,5)	(2,4,4)	(3,4,3)	(4,4,2)	(5,4,1)
(0,5,5)	(1,5,4)	(2,5,3)	(3,5,2)	(4,5,1)	(5,5,0)
(0,6,4)	(1,6,3)	(2,6,2)	(3,6,1)	(4,6,0)	
(0,7,3)	(1,7,2)	(2,7,1)	(3,7,0)		
(0,8,2)	(1,8;1)	(2,8,0)			
(0,9,1)	(1,9,0)				

etc. until (9,1,0), (10,0,0)

(0,10,0

This gives a matrix with **4356** transition probabilities!

A state is a rather abstract concept (not very problem oriented)!
It will only create problems without giving any advantage!

Three consistent approaches - 4

- **Example:** *A small system with n=100 entities that can be in k=20 different conditions.*

- **DES:** A model with 100 entities – each entity separately represented in an internal structure of 20 'states' and rules for change of conditions.

- **PoS:** 20 compartments with flows between compartments.

- **Markov:** (n+k-1 over k-1) states i.e (119 over 19) states ≈ 4900000000000000000000 states. The matrix is the square of this – and each of the matrix elements is a transition probability that must have a value. (Also for a very sparse matrix- this is not fun! Try to visualise it! *(Even with the worlds fastest computer the time of the universe is not enough even to assign transition probabilities to this matrix!*)

A consistent world of power!!!

Micro simulation for unique entities (DES)

Macro simulation (PoS)

Markov simulation (never use in simulation)

Theory of stochastic processes: analytical solutions and powerful results.

Especially Queuing theory for M/M/c (where c=∞) gives powerful theorems!

If and only if:
1) All N_i are large or
2) The model is linear

Deterministic macro models (Ordinary diff. equ's or CSS) are more accessible for analysis.

E.g. by setting $dx_i/dt=0$. Especially valuable for parameter estimation & optimisation for *linear* models.

17

Part II.   Poisson Simulation (PoS)

Random distribution of events over time

|---x--x—x-------x---x-------x---x---x---x----x-------x----x--x--> **time**

Δt Δt

3 events   2 events   2 events   3 events   1 event
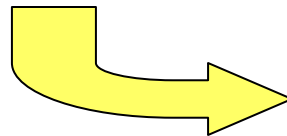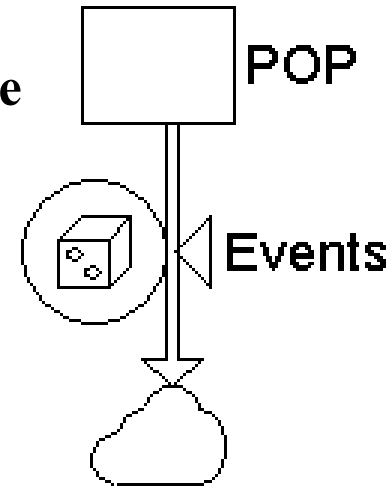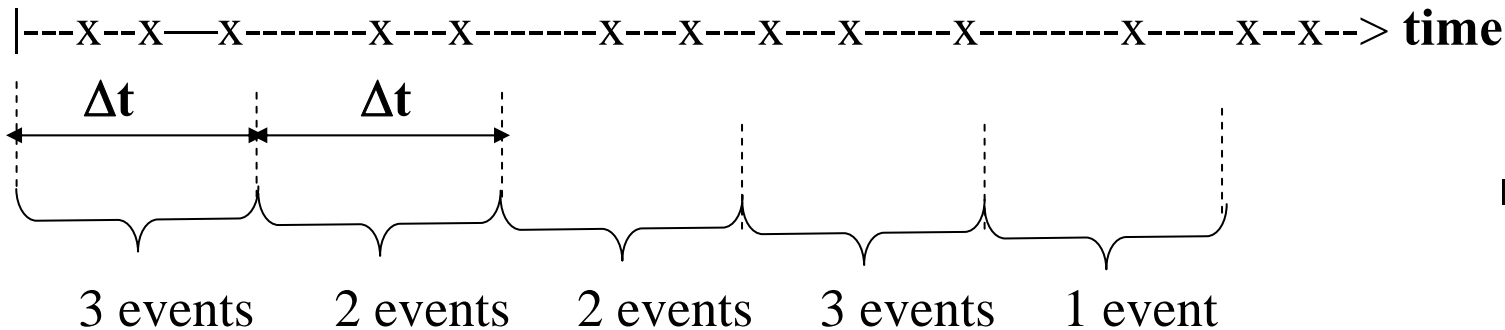
Poisson distribution

(λ**(t)** = Intensity = No. events/time unit)

(Δ**t** = Length of interval)

POP

Events

THUS: **Po[λ(t)·Δt] is the stochastic description & numerical realisation of the number of events during t to t+Δt.**

The Poisson distribution

- <u>The number of events that occur during a time interval Δt</u> when the events occur independently and one at a time with an intensity λ is Poisson distributed. $X \in Po[\lambda \cdot \Delta t]$

- Then: $p_x(k) = e^{-\lambda \cdot \Delta t} \cdot (\lambda \cdot \Delta t)^k / k!$ for $k \in \{0, 1, ...\}$.
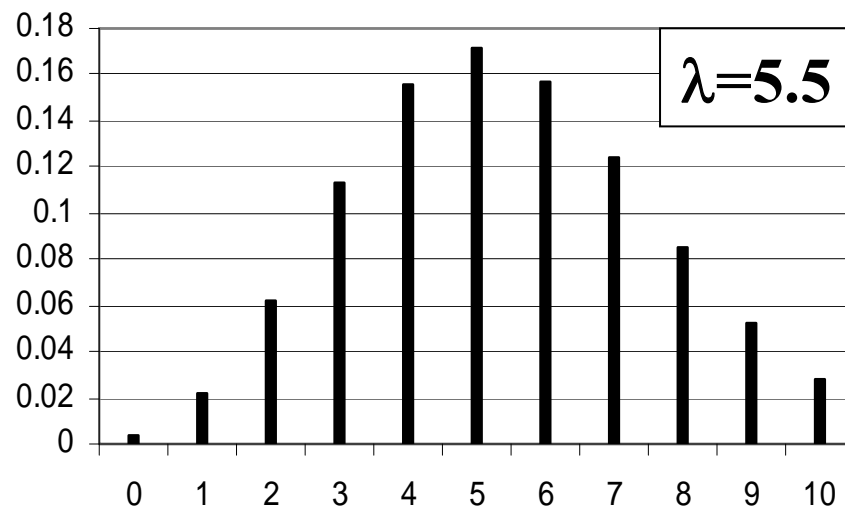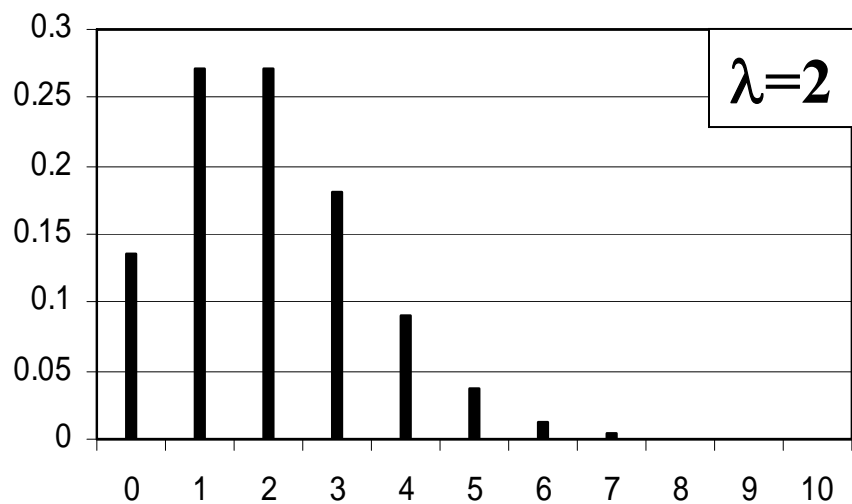
- **$Po[\lambda(\underline{x},t) \cdot \Delta t]$ has only one parameter λ and $E[X]=Var[X]=\lambda \cdot \Delta t.$**

(Δt is not a model parameter! – Δt is the numerical step-size and should just be small enough.)

19

Deterministic and stochastic representation

Deterministic model

$$dx/dt = F1 - F2$$
$$F1 = input$$
$$F2 = output$$
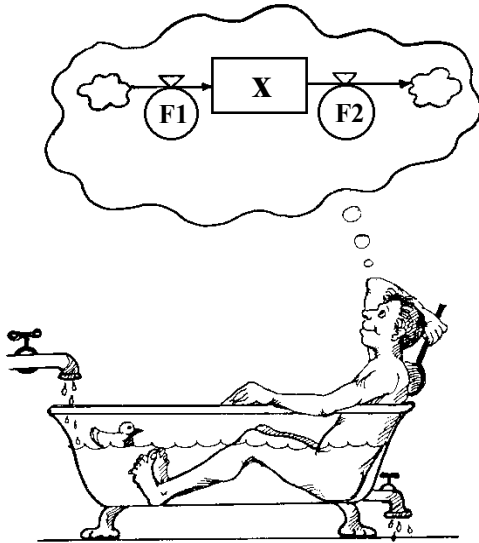
Stochastic model
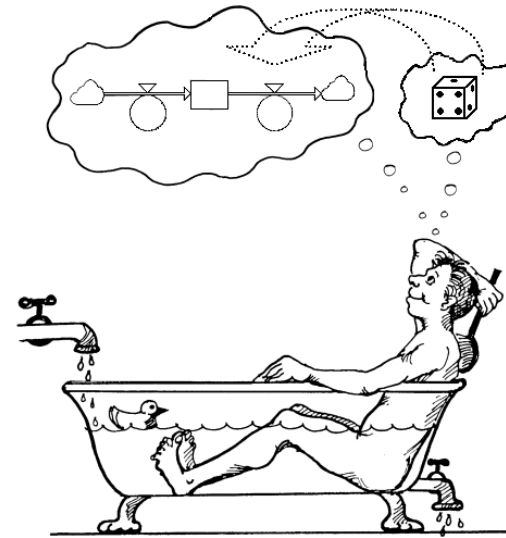
*Differential equation systems are often modelled in terms of **States** and **Flows**.*

Computers can't handle infinitesimally small dx and dt.

Make it numerical with e.g. Euler's method: ($\Delta x/\Delta t = F1 - F2$)

$$x(t+\Delta t) = x(t) + \Delta t \cdot F1 - \Delta t \cdot F2$$
$$F1 = input$$
$$F2 = output$$

$$x(t+\Delta t) = x(t) + \Delta t \cdot F1 - \Delta t \cdot F2$$
$$F1 = \textbf{Po}[\boldsymbol{\Delta t} \cdot input]/\boldsymbol{\Delta t} \quad *$$
$$F2 = \textbf{Po}[\boldsymbol{\Delta t} \cdot output]/\boldsymbol{\Delta t} \quad *$$

[*] **A Poisson distributed sample is drawn for each time-step.**

20

The general population equation in PoS
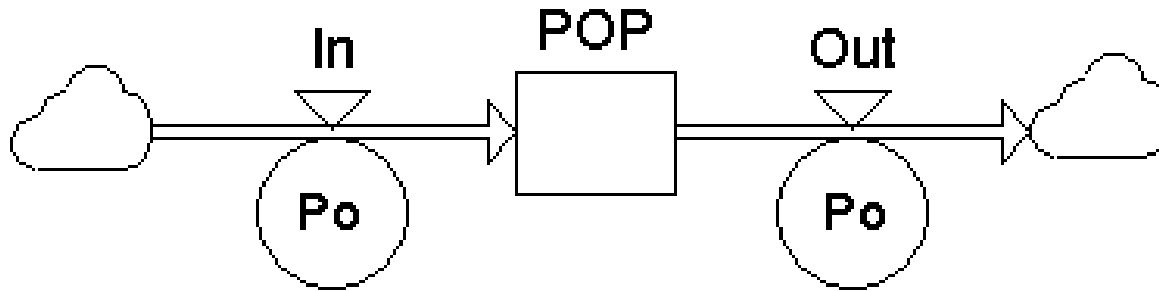
- The general (dynamic & stochastic) equation for aggregated population models in numerical form:

$$\Delta x_i / \Delta t = Po[\Delta t \cdot f_{in,i}(\underline{x},t)]/\Delta t - Po[\Delta t \cdot f_{out,i}(\underline{x},t)]/\Delta t$$

where i=1…k.

Three principles for Poisson Simulation

Poisson Simulation is a modeling method based on stochastic difference equations where:

1. The stochastics are (almost entirely) introduced in the *flow rates*.

2. The stochastics are *modeled* by coupling the parameters of the statistical distribution to other quantities of the model.

3. The stochastic is implemented so that *the integration step-size* (Δt) is adjustable without rebuilding the model. (Which is not the case for e.g. a Markov model.)

Example 1: Radioactive decay

Study of x=100 radioactive atoms that decay with a time constant of T=10 time units. (a=1/T=0.1)

Deterministic model

$$x(t+\Delta t) = x(t) - \Delta t \cdot F(t)$$
$$F = x \cdot a$$
$$x(0)=100$$
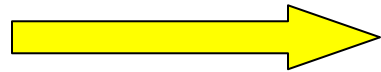
PoS model

$$x(t+\Delta t) = x(t) - \Delta t \cdot F(t)$$
$$F = Po[(x \cdot a) \cdot \Delta t]/\Delta t$$
$$x(0)=100$$

Gives: x(t)=100·exp(-t*c)

Gives a new stochastic solution for each simulation

Example 1: *Radioactive decay*
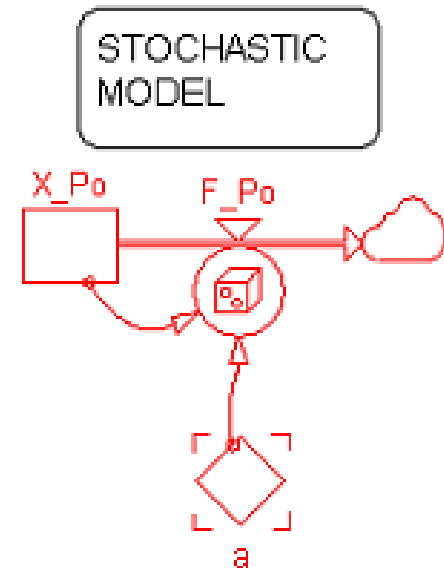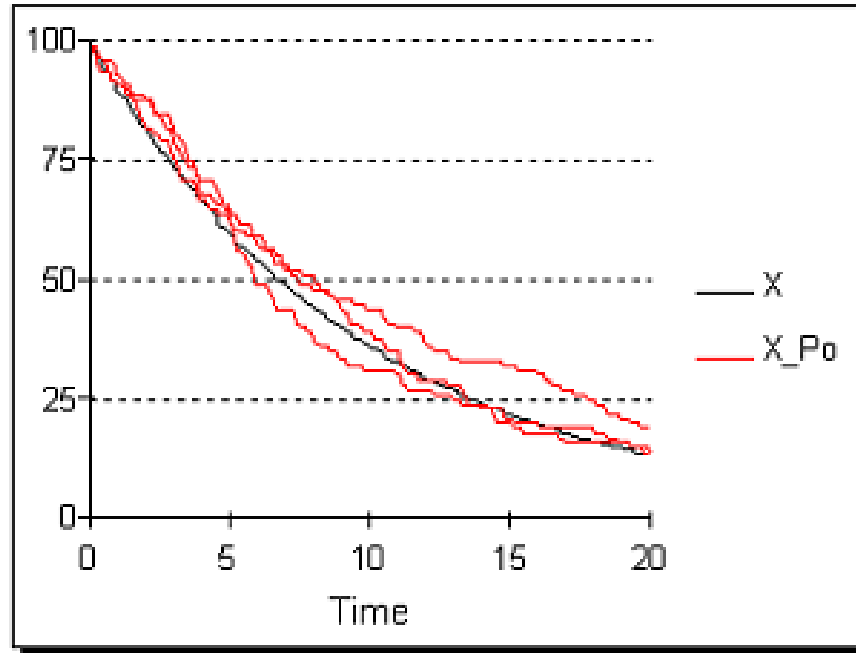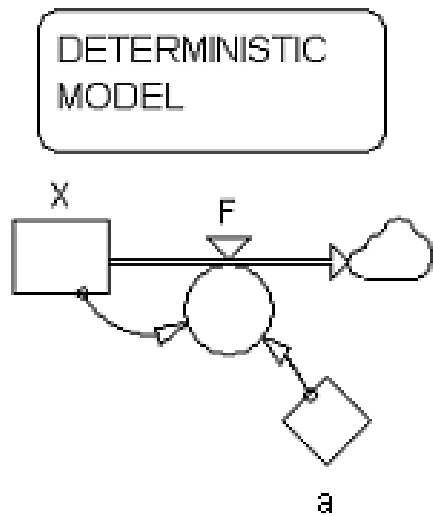
$$x(t+\Delta t) = x(t) - \Delta t \cdot F$$
$$F = a \cdot x$$

$$x(t+\Delta t) = x(t) - \Delta t \cdot F$$
$$F = Po[\Delta t \cdot a \cdot x]/\Delta t$$

$$a = 0.1, \quad x(0) = 100$$

Example 2: *Lotka-Volterra model*
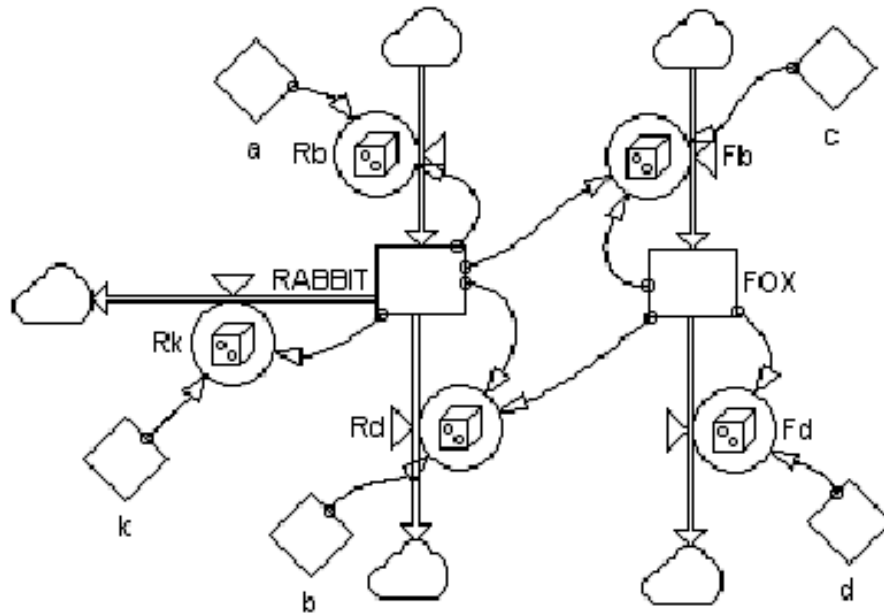
Volterra equations:
a=0.2, b=0.005, c=0.005, d=0.3, k=0.001
gives stat. values: X(0)=60, Y(0)=28

STATIONARY VALUES:
X=0, Y=0

X=a/k, Y=0

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

Here Foxes became extinct!
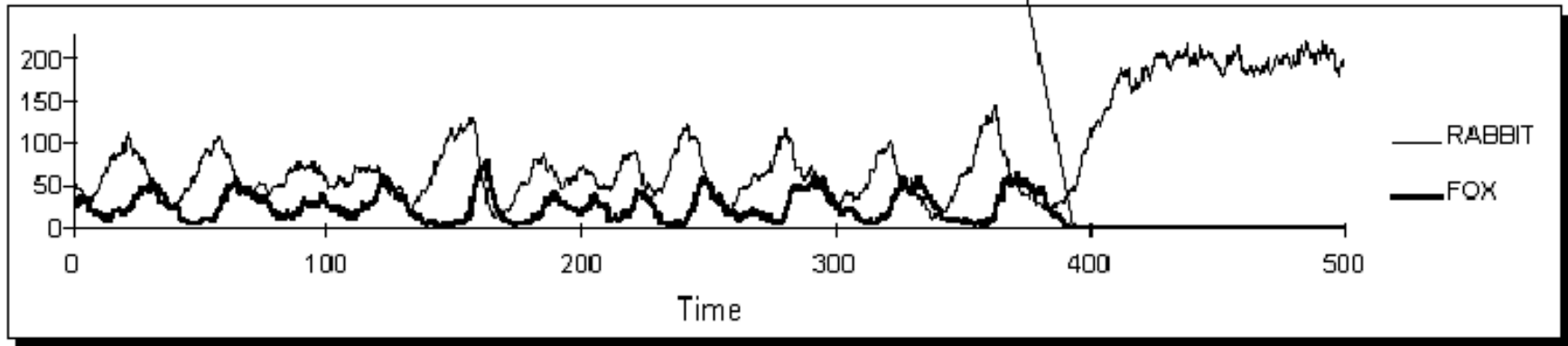
Conclusions: 1) Stochastics excites dynamics! (Even when started in equilibrium!)
2) A stochastic model may switch to another dynamic ***mode***!

Why Poisson simulation?

- Stochastics excite dynamics and dynamics change the stochastic conditions.

- The model may switch between modes (cfr. Volterra).

- Average from a stochastic model may differ largely from that of a deterministic one.

- Two deterministic models that behave exactly the same may behave quite differently when stochastics are added.

- If dynamics and stochastics are modelled separately, both the statistic and the dynamic estimates become wrong!

- Even when they give the same results in average – PoS provides statistical estimates to the results (variations, C.I., correlations etc.)

Dynamics and stochastics must be treated together in a correct way when both aspects are important !!!

Part III.   Statistical tools

Contents:
A. Statistical estimates of outcomes.
B. Parameter estimation & resampling (for linear models)
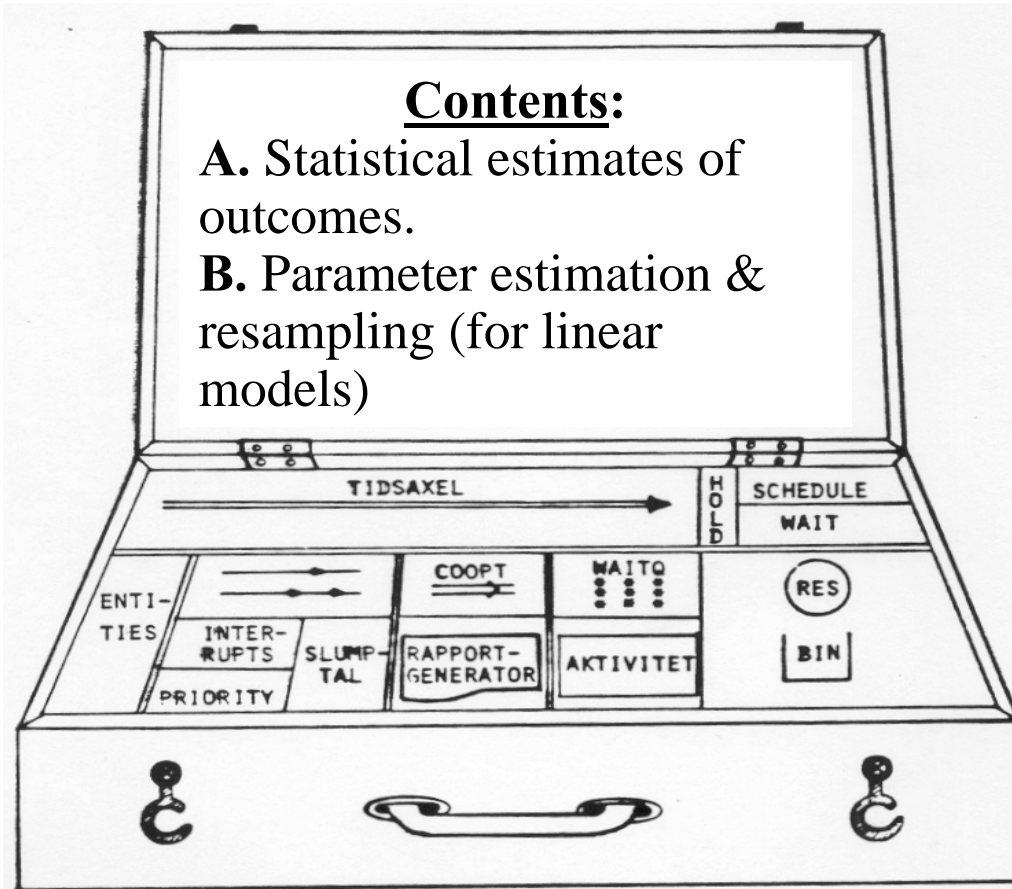
- A deterministic model is run once.
- A stochastic model must be executed many times and the results should be collected and transferred into statistical estimates.

A. Statistical estimates of <u>outputs</u>
- Mean
- Variance
- Confidence interval
- Min, Max
- Percentiles
- correlations

B. Statistical estimates of <u>model parameters</u> (by resampling) – Unbiased method for linear models.
- Mean
- Variance
- Confidence interval
- Min, Max
- Percentiles

Statistics is so much <u>easier</u>, <u>direct</u> and more <u>flexible</u> in simulation than in statistical theory!

For example just make, say,10 000 runs and take the range of e.g. 9 500 of the them
    ⇒ 95% C.I.
The model is usually so complicated that a wise statistician wouldn't even try – and an unwise one would start to simplify it beyond any sense!
It is a fact that the mathematically and statistically most studied disease: Morbus Simplicitus is not yet discovered by the medical profession! (See Frame 50.)[28]

A. Statistic estimates of outcome

Ex: *Lanchester's model of warfare.* Force 1 (x1 entities) with hitting power *a* and Force 2 (x2 entities) with hitting power *b* in combat, fighting to one force is wiped out. • Will the stronger force always win? • How many will survive in the winning force? • How long does the fight take?

$$\begin{cases} dx1/dt = -b{\cdot}x2 \\ dx2/dt = -a{\cdot}x1 \end{cases}$$

x1(0)=5
x2(0)=3
a=b=0.1

Mathematical solution for a=b and x1(0)=5 & x2(0)=3:

dx1/dx2=x2/x1

∫x1dx1 = ∫x2 dx2

$x1^2|0..T = x2^2|0..T$

x1(0)=5 and x2(0)=3 gives because of x2(T)=0:

$x1^2(T)-x1^2(0)=-x2^2(0)$

Why x1(T)=4 when all x2 are killed.
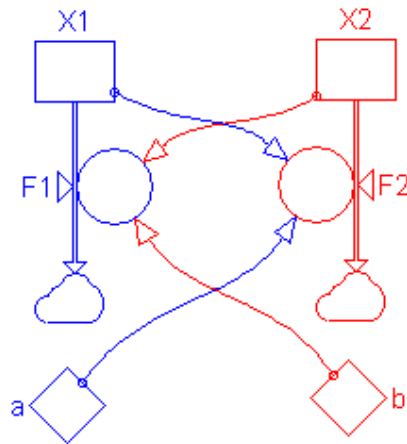
Results:
$\begin{cases} x1(T)=4 \\ x2(T)=0 \end{cases}$
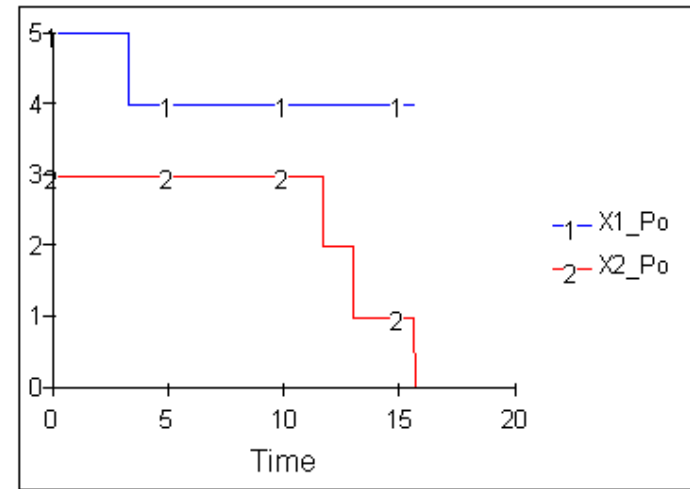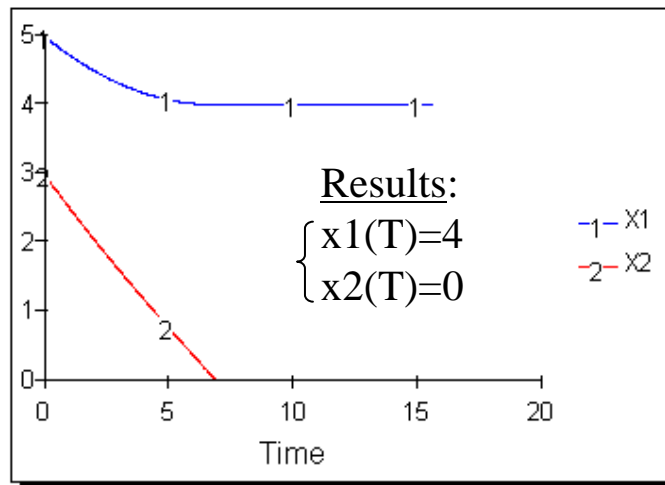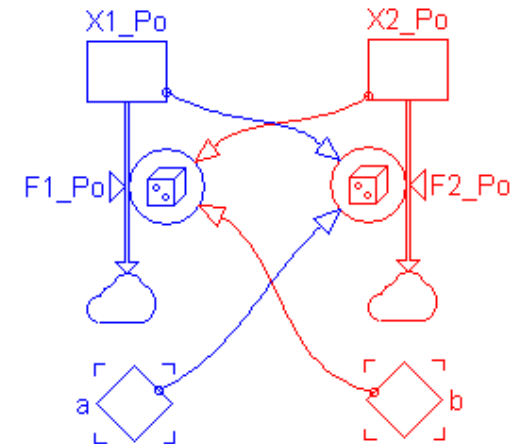
Wanda – StocRes: Calculating statistical estimates.

WANDA StocRes 1.1

Select File | C:\Program\Microsoft Office\Word_Filer\Methods&Tools\Lanchest.sim

Result Variable [ ] Add | Del * | Back *

Prob.-Level
○ 1-sided
● 2-sided
$1-\alpha$ 95
Help C.I. λ 1.9604

Percentile 50 %

Stocastic Results:

Name	Average	Std.Dev.	Conf. Int.	Min	Max	Percentile	*
x1 _Po	3.463000	1.653289	3.360507 - 3.565493	-3.47E-18	5.000000	4.000000	
x2 _Po	0.232000	0.694736	0.188931 - 0.275069	-3.47E-18	3.000000	-3.47E-18	
Win_x1	0.884000	0.320385	0.864138 - 0.903862	0	1.000000	1.000000	
Battle_Time	7.751216	5.438783	7.414048 - 8.088384	0.422000	39.03000	6.415000	

- $E[x1[T] \neq 4!$

- Sometimes x2 will win!

- The battle time also differs from a deterministic solution!

No. Runs 1000
Runs done 1000

Send Seed ☐
Skip On Zero ☐

Cont. | Reset | Print
Halt | Break | Help

☐ E-format
Status:
DONE

Mean & C.I. Difference (requires exactly two *-marked variables.)
Calc CI-Dif | [ ] | [ ] | [ ] | Help C.I. Dif

Display by marking exactly one variable by a *. ——— Send to Notepad or a Text-file
● Unsorted
○ Sorted (on *)
Histogram (*) | Scatter Plot (**) | To Notepad | Select File

Read more about Lanchester's model at Frame 44.

2004,08,19 - 14,14,41   Exec. Time: 51.76 min.   DT=0.002,   END when (x1=0 or x2=0).

Histograms and scatter plot (pdf)
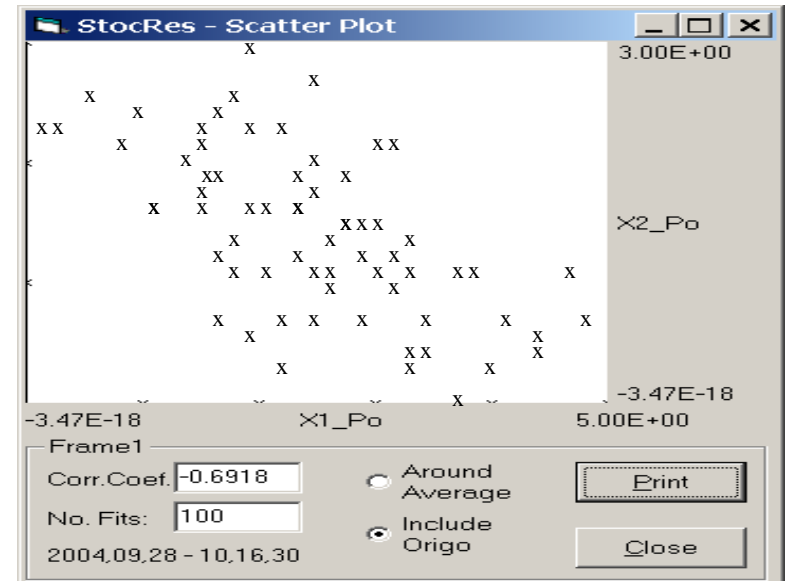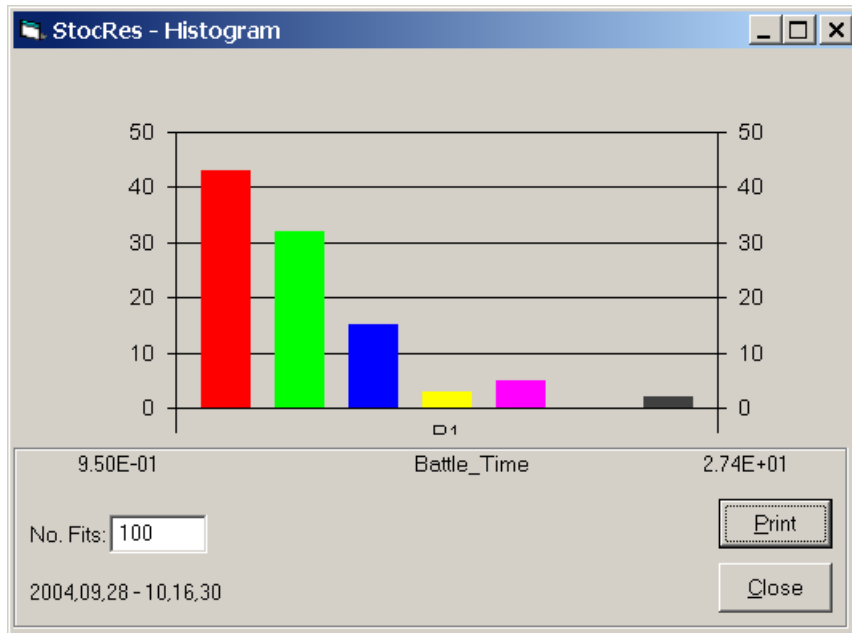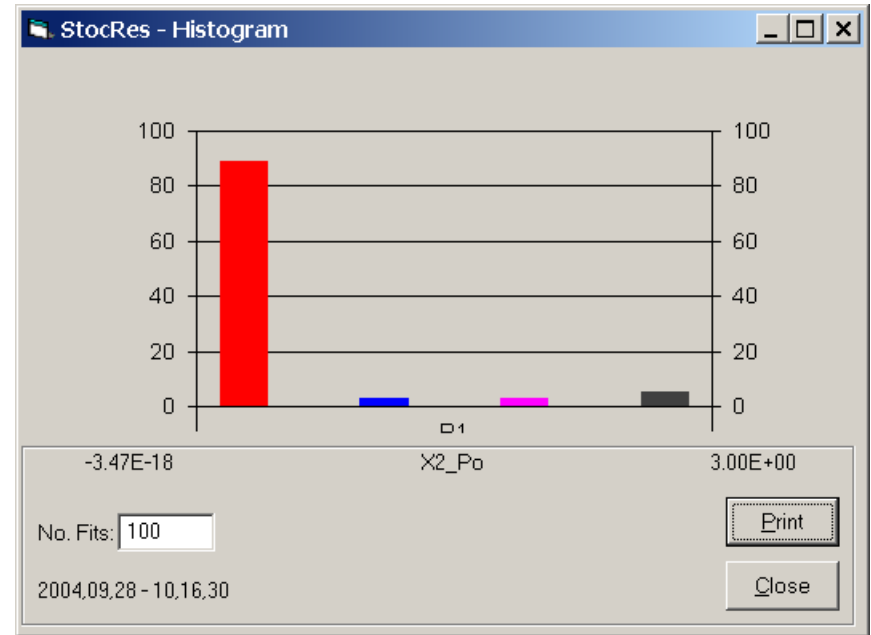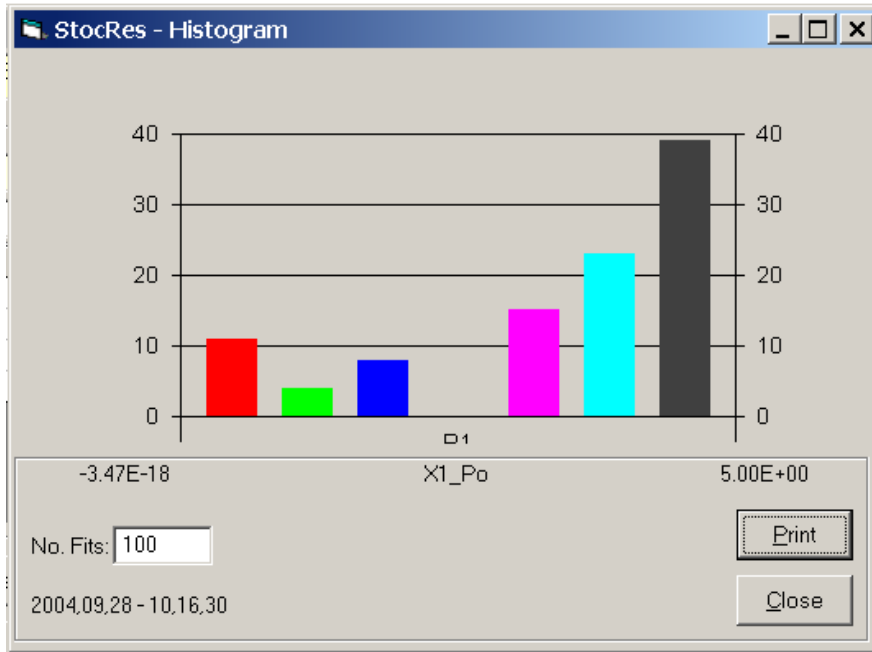
B. Parameter estimation
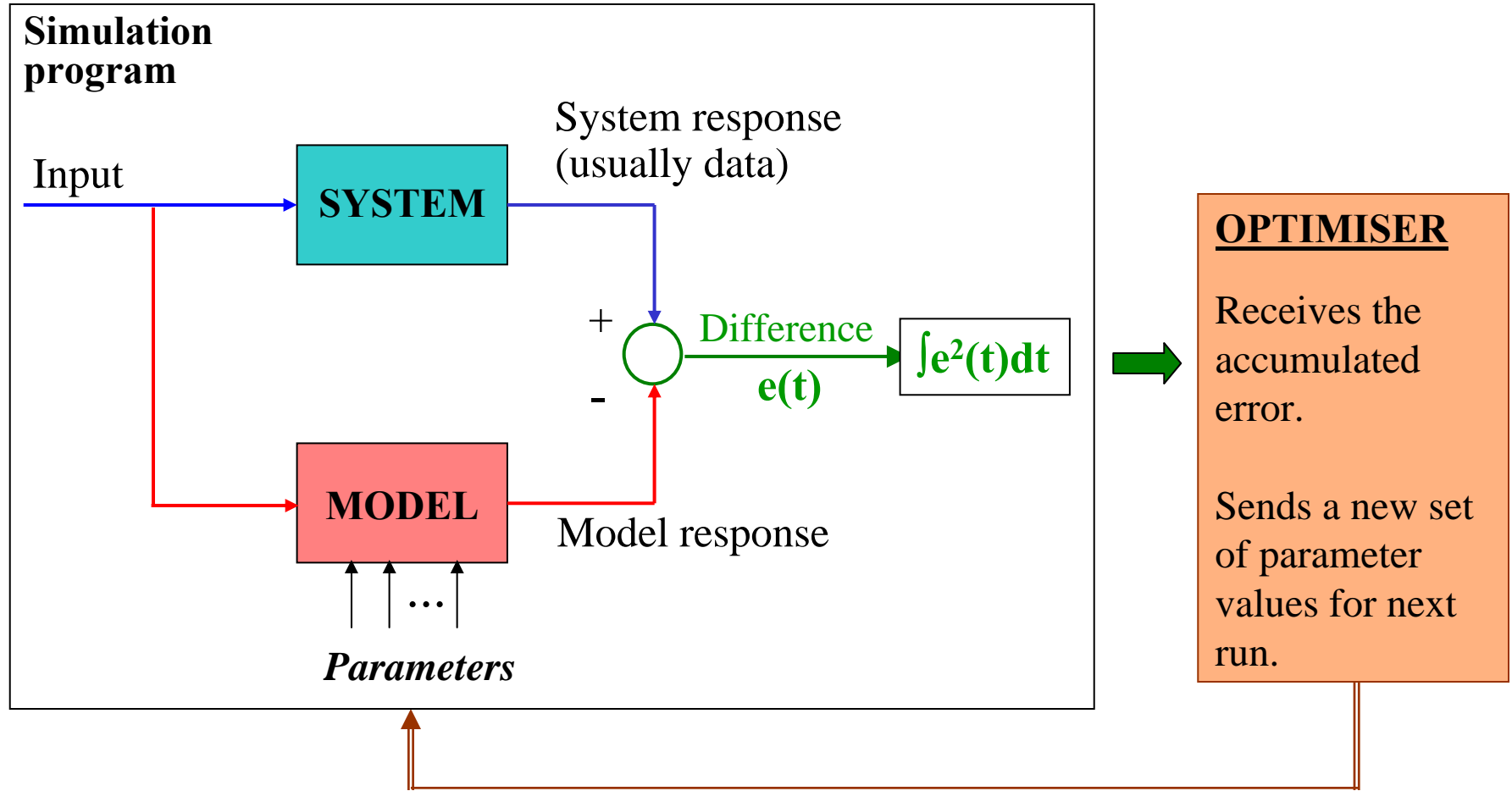
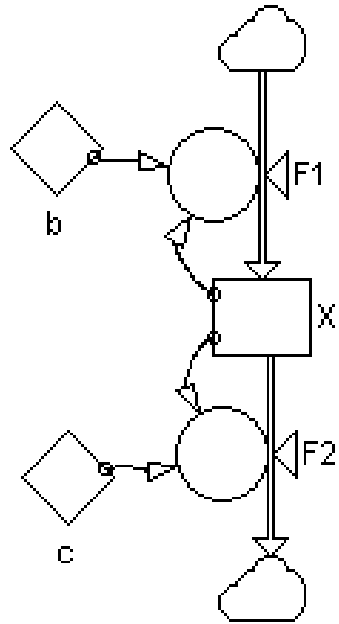Model fitting is about finding that model whose behavior maximally fits the system's.

Ex: *Logistic growth model.* The number of animals, x(t), will breed (+b·x) and be reduced by inter-species competition (-c·x^2). ● How will the number of animals develop?

$$\Delta x(t+ \Delta t) = \Delta x(t) + \Delta t·(F1-F2)$$
$$F1 = b·x$$
$$F2 = c·x^2$$

$\Longrightarrow$
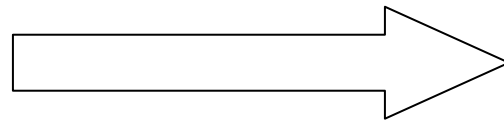
$$F1 = Po[b·x·\Delta t]/\Delta t$$
$$F2 = Po[c·x^2·\Delta t]/\Delta t$$
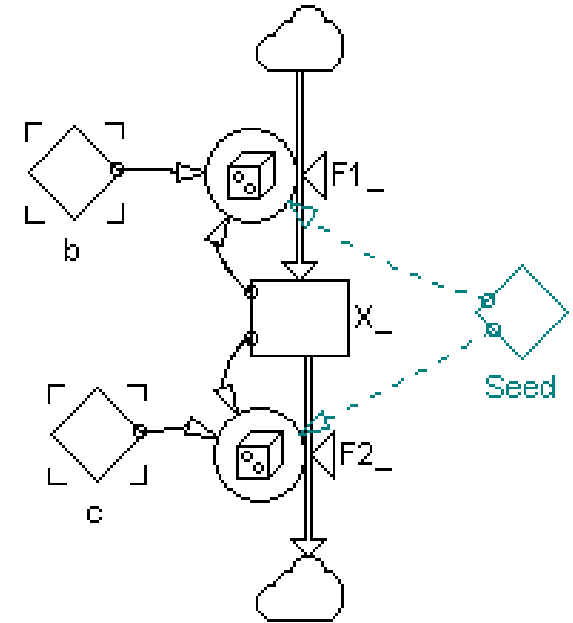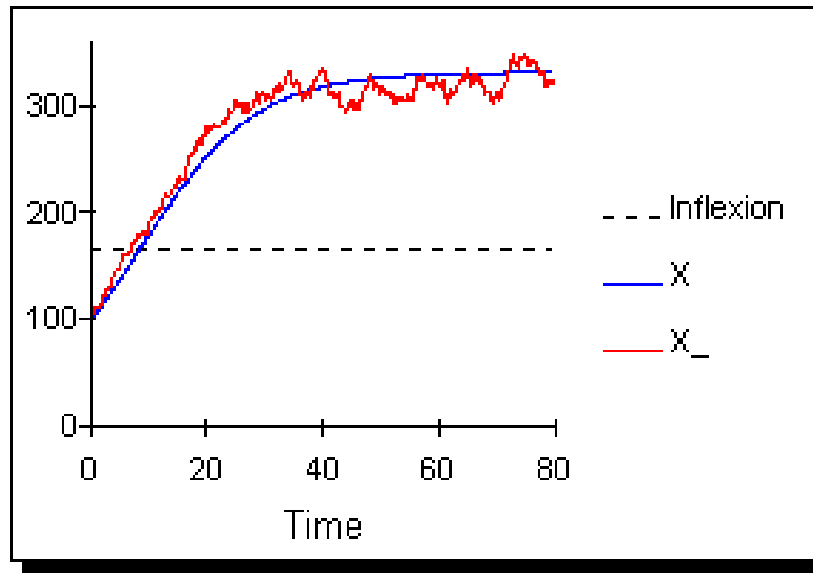
b=0.1, c=0.0003 , x(0) = 100

This model is not linear. But the relatively large number of entities make the deterministic and stochastic solutions close to equal in average!

Fitting the *deterministic* logistic model to system data

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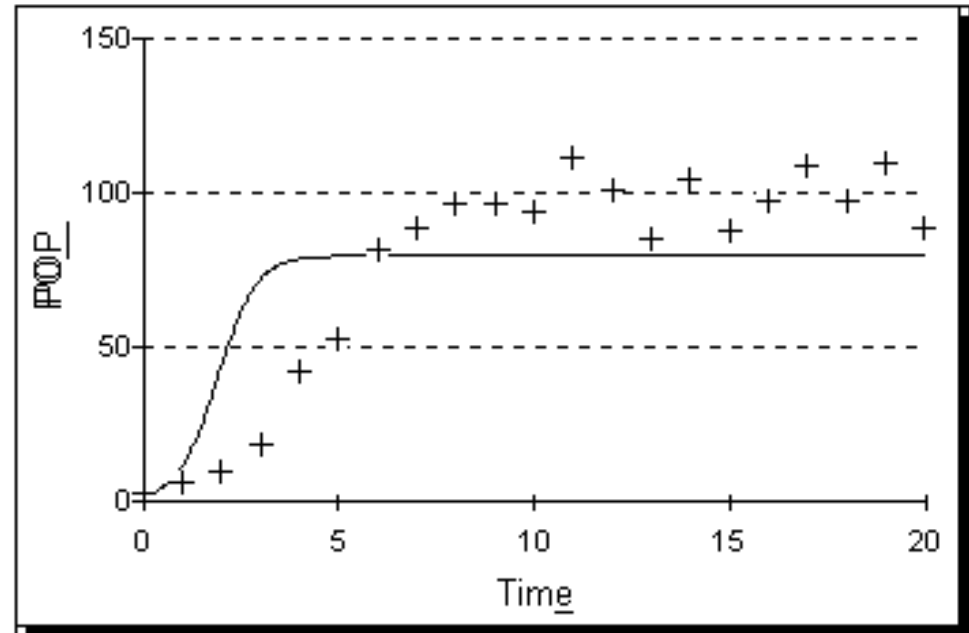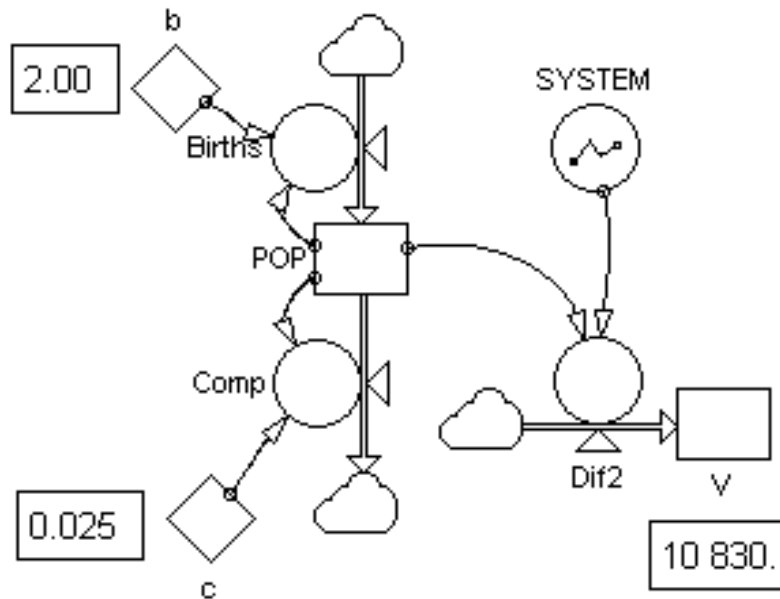
The difference between SYSTEM and model (POP) is calculated and squared. At each time step this squared error (DIF2) is cumulated into V (which was empty at start of simulation). V therefore gives a measure of how well SYSTEM and MODEL fit. We want to find that combination of the parameters b and c that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

PowerOpt: An optimiser used for model fitting
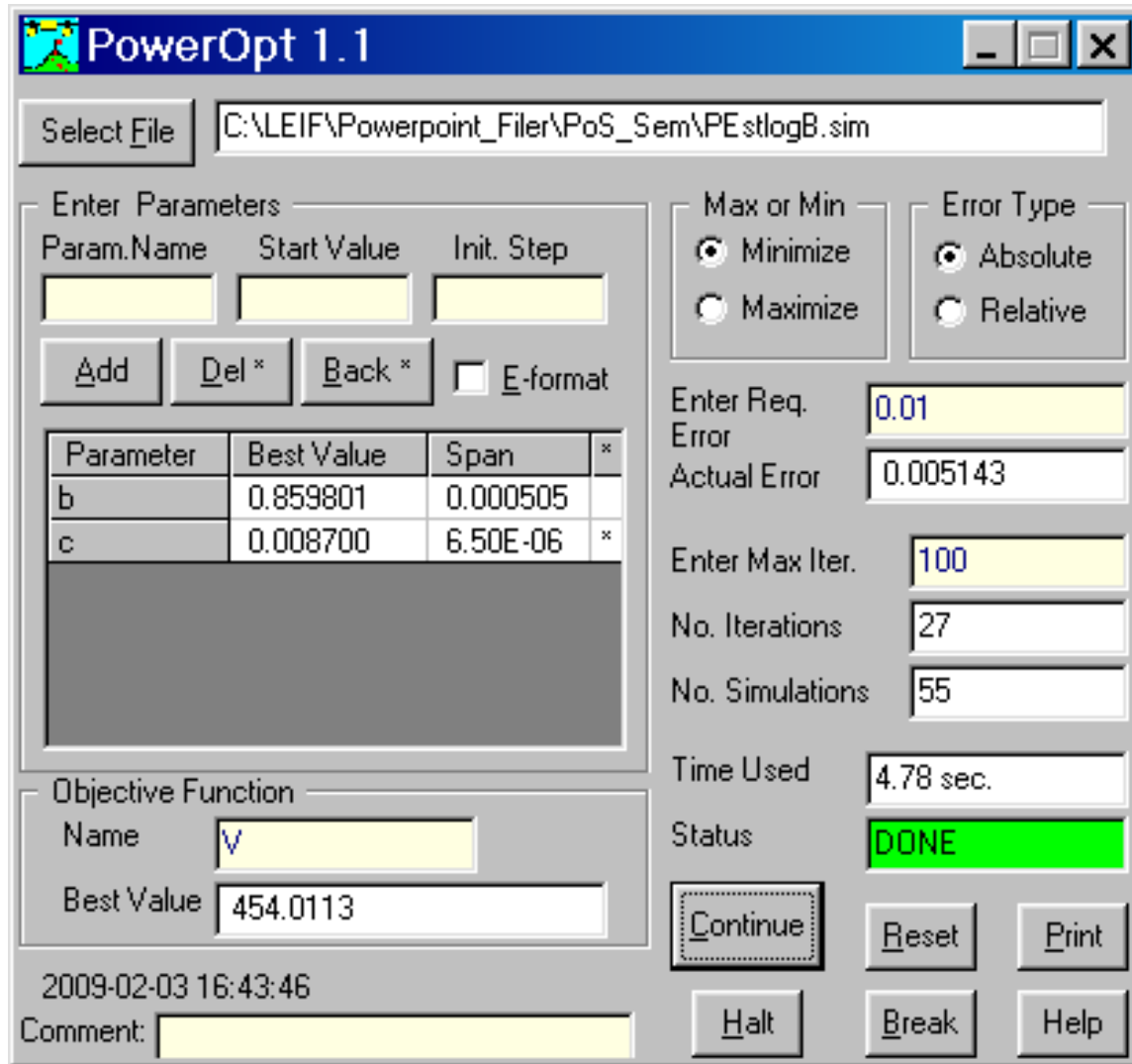
It took 4.8 seconds to find the optimal set: a=0.86 & b=0.0087

Best fit of parameters b & c - V is minimised!

File: PEstLogB.sim

b
0.86
Births

SYSTEM

POP

Comp

0.0087

c

Dif2

v

454
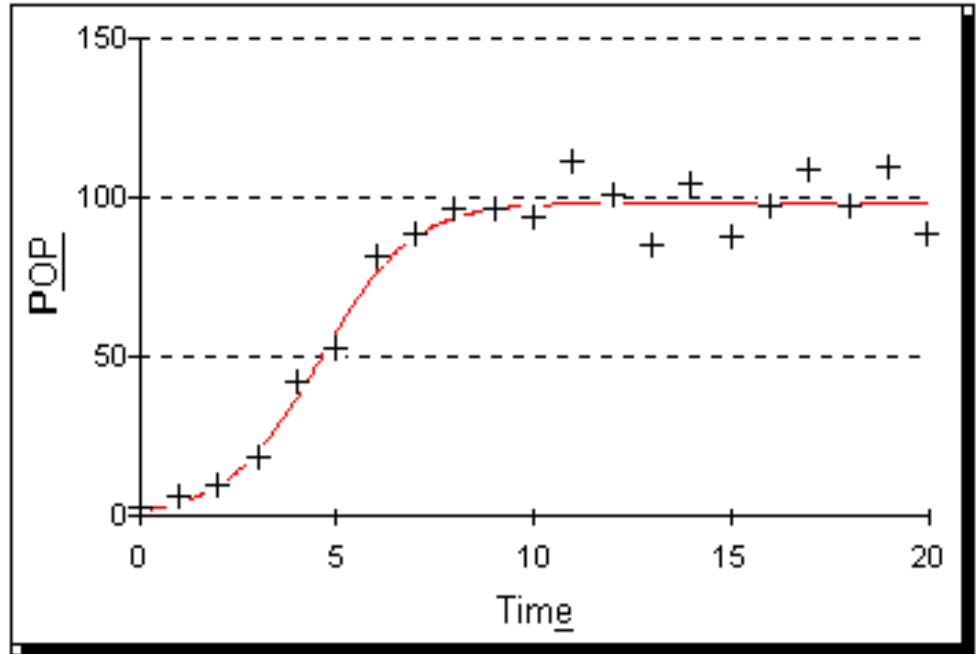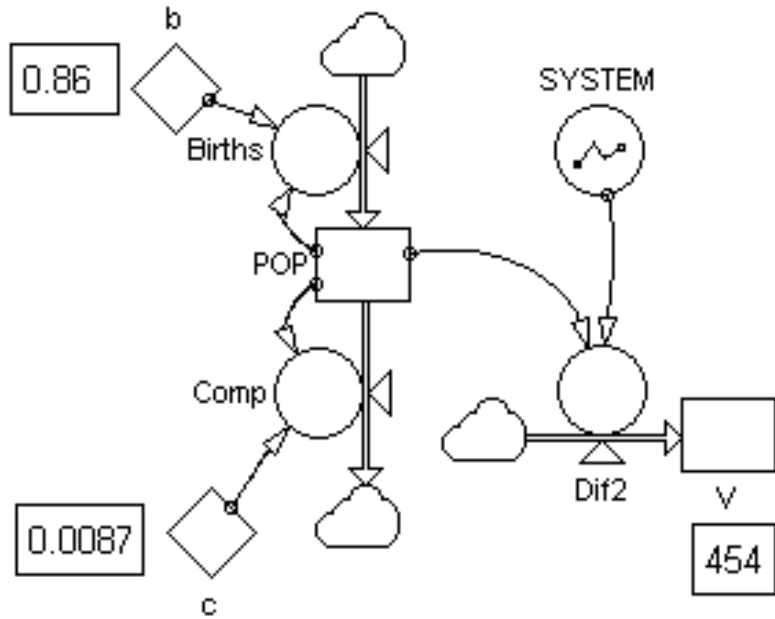
Variations around the best fit model!

File: PEstLogC.sim.
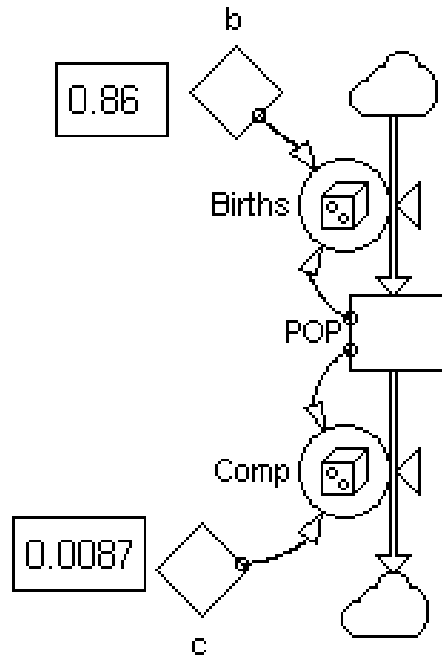
(Note: The logistic model is not linear! We have here fitted the deterministic model to data.)
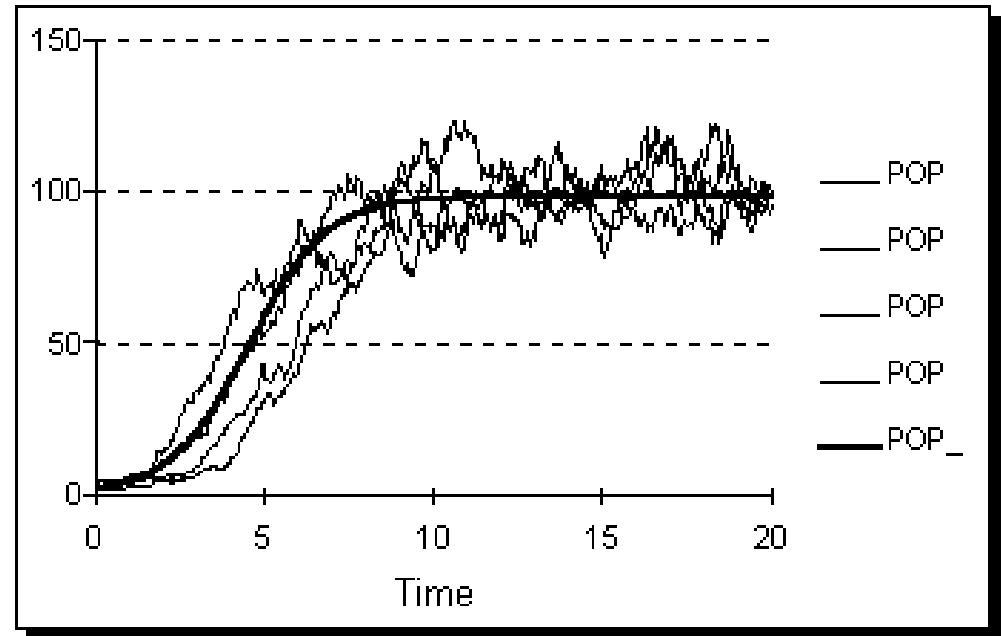
- **By running the Poisson model N times we get the variations around an average ("*Resampling*").**

- **If we want to estimate the variations in the parameters *b* and *c* we fit *b* and *c* to each of the N stochastic curves.**

- **From these N estimates of *b* & *c* we get their variations.**

Wanda – ParmEst   (Linear models only)
Fitting, Resampling and Statistics of the logistic model.

Parameterisation

We have studied dynamic models of the form: dx/dt=f(x,t).

To make the example simple we now study a model of type dx/dt=f(t) and chose among the polynomial models:

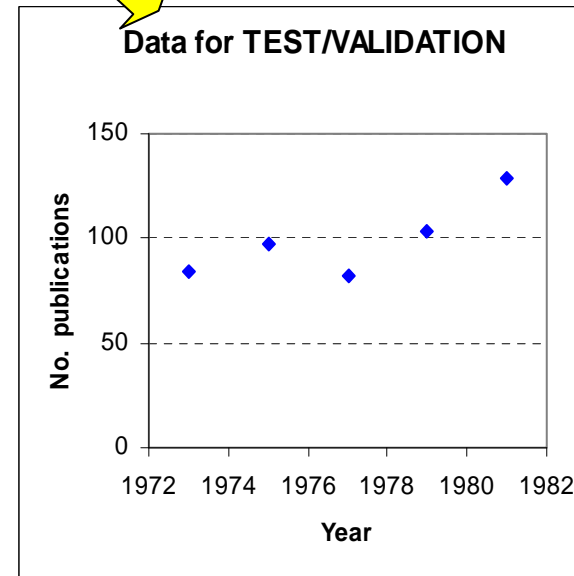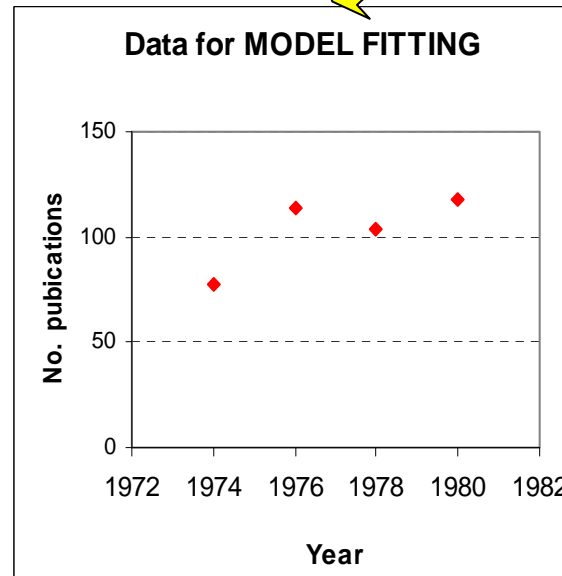0) $F=a$

1) $F=a+b \cdot t$

2) $F= a+b \cdot t+c \cdot t^2$

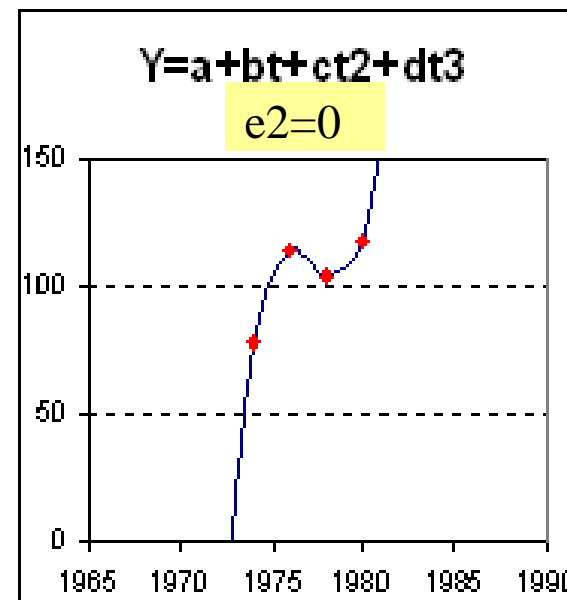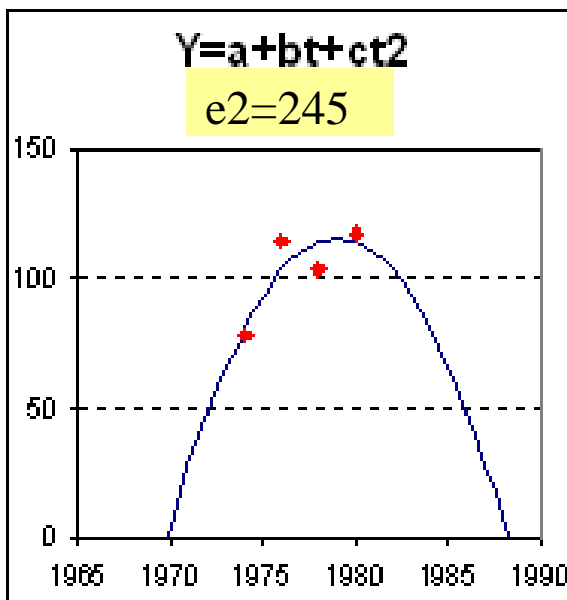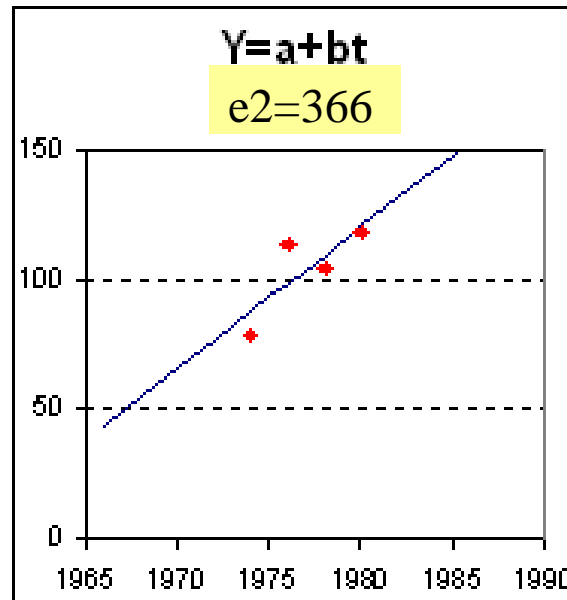3) $F=a+b \cdot t+c \cdot t^2+d \cdot t^3$
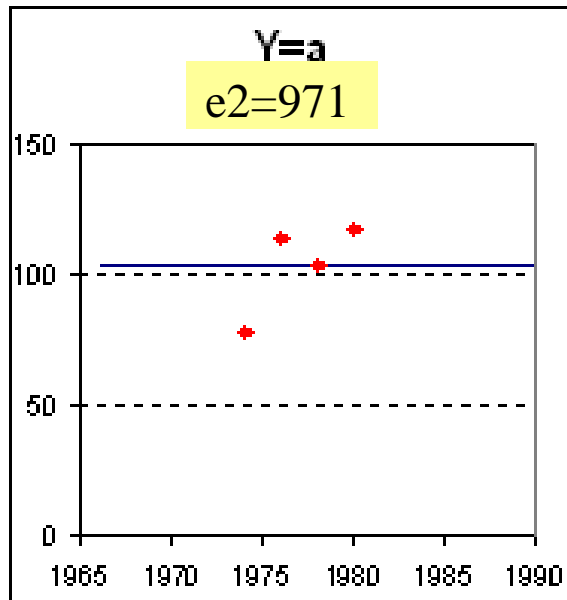
for best fit to our observed data.

The number of publications at the technical faculty in Uppsala in 1973-81.

All data
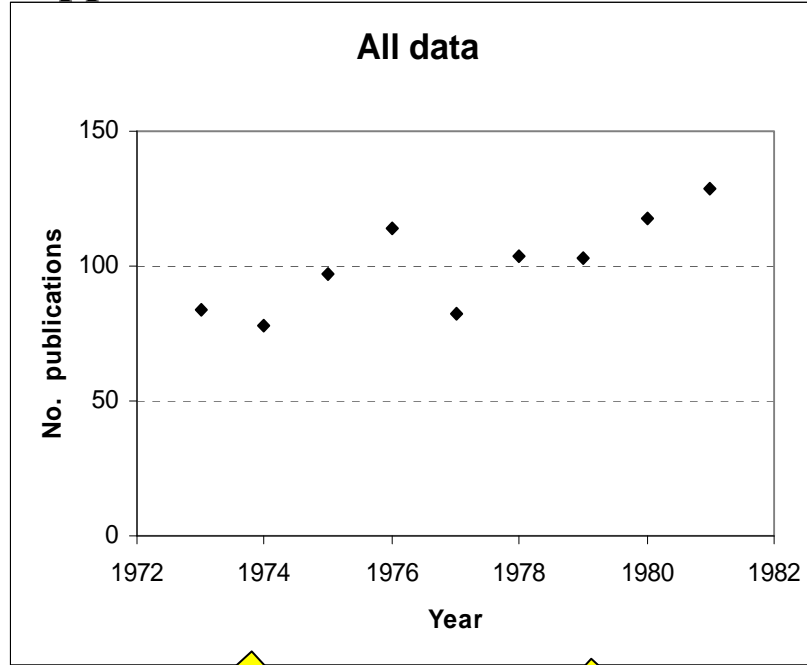
Note: *This separation of data is not optimal – but is used to make some ideas very clear!*

Data for MODEL FITTING
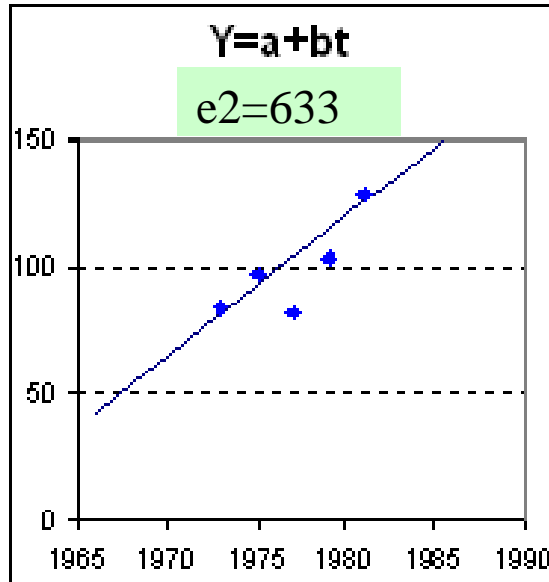

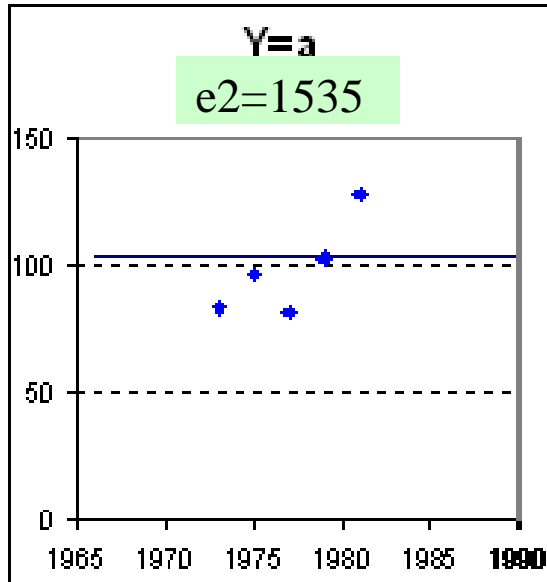**Data for TEST/VALIDATION**

39

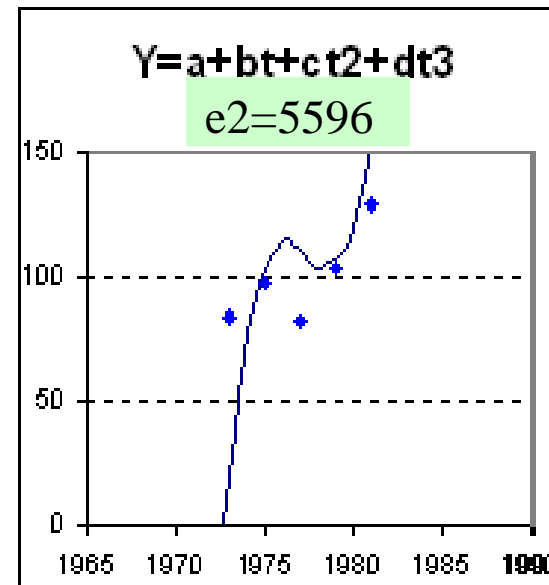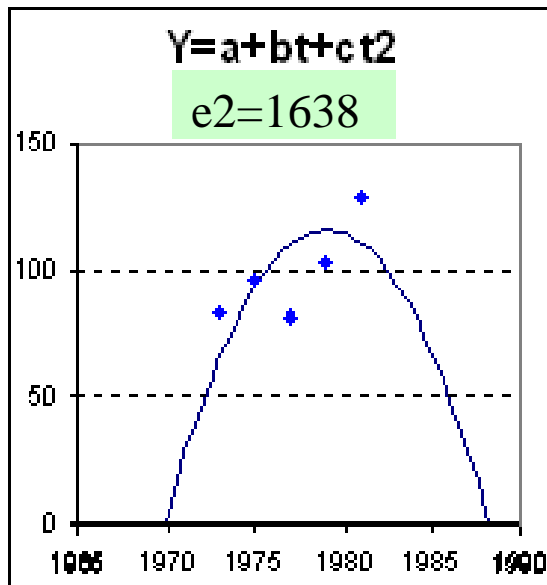# The best least square fit to data using the four polynomial models:

1. Which model is the best?

2. Interpret the models! What would they imply if they are correct?

Test the fitted models to independent data:

Well! What model do you now prefer??

The model should capture the nature of the system!

Do not model the random errors!

You can always make a best fit to whatever model!

The Significant Correlation Between the Average Annual Age-Specific Incidence Rates of Malignant Melanoma and the Topography of the Norwegian West Coast.   (O.H. Iversen, J.I.R. 1978.)
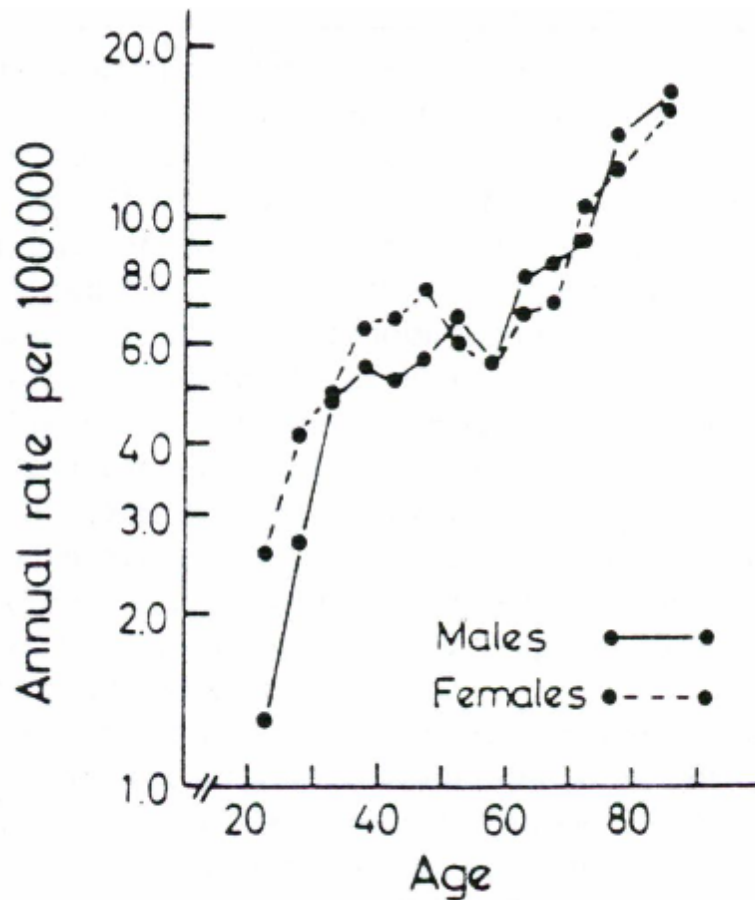

FIGURE 4. Incidence of malignant melanoma in Norway. Stolen from Knut Magnus (1975).
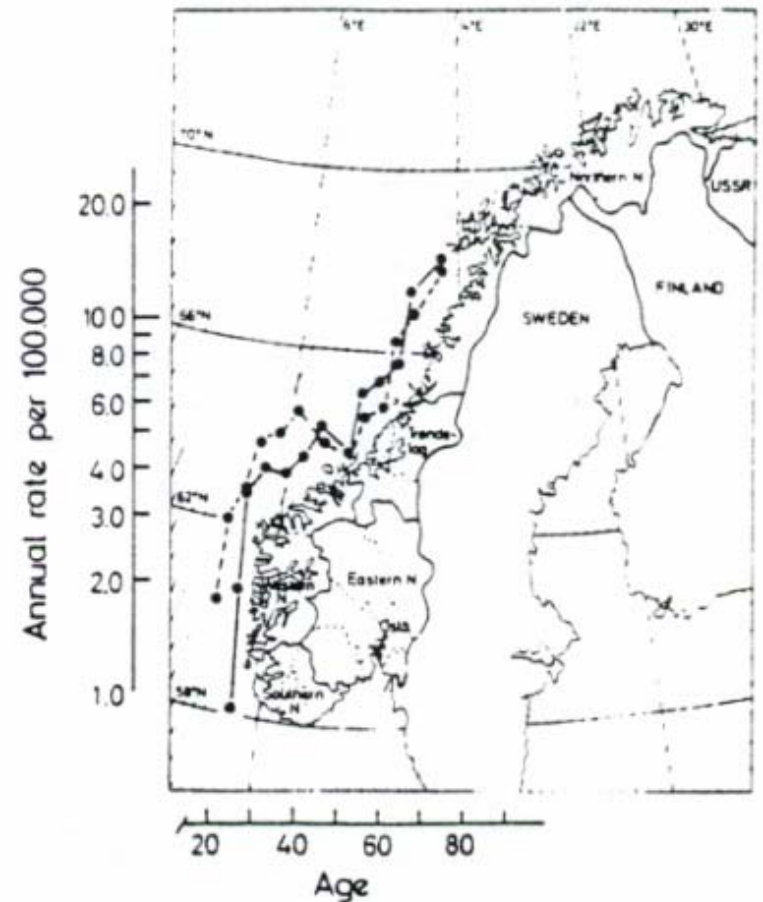
FIGURE 5. This figure shows the correlation between the Norwegian coast and the incidence of malignant melanoma.

Part IV. The danger of routine-like modelling
- bad, worse and catastrophic examples

1. Noise must not just be added or multiplied to a deterministic model. – See artefacts at Frame 8.

2. Deterministic and stochastic models may behave very different – also in average! – See Volterra at Frame 25, Lanchester's model of warfare at Frame 29+ & 44, and the SIR model at Frame 45-49.

3. Wrong sojourn-time distribution for a stage.

   (A stage is something different from a state or state variable (compartment)) – See Green's hypothesis at Frame 51.

2) Lanchester's model revisited (See Frame 29)

The model with Force 1 (x_1 entities) and Force 2 (x_2 entities) is linear and of the form: $dx/dt = \mathbf{A}x$. Why don't we get the same average for the deterministic and stochastic cases?
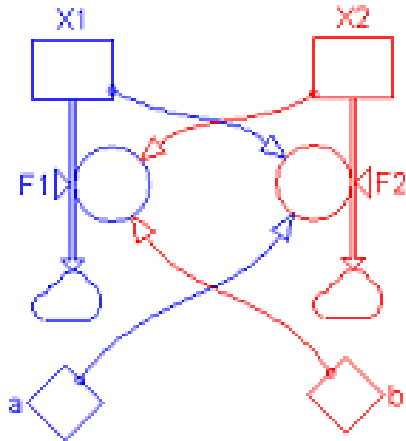
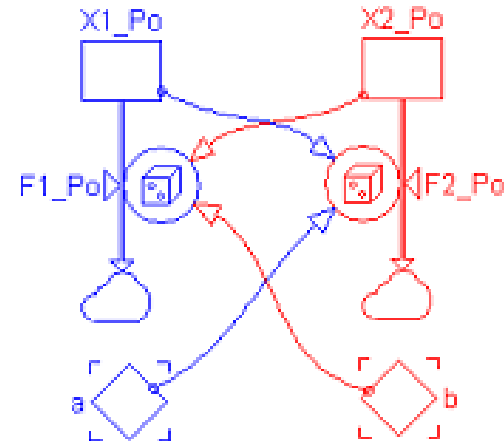$$\Delta x_1(t+\Delta t) = x_1(t) - \Delta t \cdot b \cdot x_2$$
$$\Delta x_2(t+\Delta t) = x_2(t) - \Delta t \cdot a \cdot x_1$$

$$\Delta x_1(t+\Delta t) = x_1(t) - Po[\Delta t \cdot b \cdot x_2]$$
$$\Delta x_2(t+\Delta t) = x_2(t) - Po[\Delta t \cdot a \cdot x_1]$$

• For the deterministic model x_1 and x_2 can take any values: $-\infty < x_i < \infty$. Then negative soldiers will start to generate more enemies!

• The stochastic population model can only (correctly) permit: $0 \leq x_i < \infty$.

Lanchester's model is an example of a simple *linear* model.

Still, *there exists no simulation time for which: E[Stoch. model results] = Deterministic model result.*

44

__Example__: *A Conceptual SIR model*

The classical SIR model (~ Kermack & McKendrick, 1927):

A <u>population</u> consisting of individuals is affected by an infectious disease. The population is *homogeneous*, i.e. there are no differences between individuals due to age, sex, behaviour etc.

<u>The disease</u> has the three consecutive stages **S→I→R**. [**S**usceptible, **I**nfectious & **R**ecovered (**R**emoved)]. The assumption is that every individual meets every other under equal conditions over time. Furthermore it is assumed that:

- An event is *statistically independent* of all other events at that point in time.
- The probability of a **S**usceptible individual being infected by an **I**nfectious individual is *p* per time unit.
- The mean time an individual resides in stage **I** is *T* time units.

__The conceptual model__ is thus a *stochastic* model of *discrete individuals* operating in c*ontinuous time*.

The macro SIR model for simulation

$$S(t+\Delta t) = S(t) - \Delta t \cdot F_1(t)$$

$$I(t+\Delta t) = I(t) + \Delta t \cdot F_1(t) - \Delta t \cdot F_2(t)$$

$$R(t+\Delta t) = R(t) + \Delta t \cdot F_2(t)$$
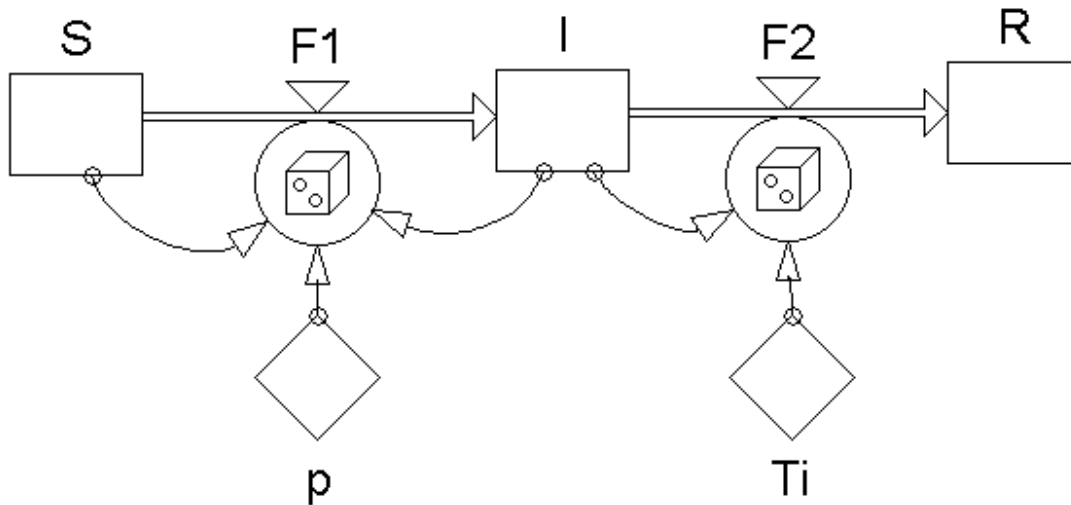
$$F_1(t) \approx Po[\Delta t \cdot S(t) \cdot I(t)]/\Delta t \quad *$$

$$F_2(t) = Po[\Delta t \cdot I(t)/T]/\Delta t$$

$$S(0)=1000$$

$$I(0)=1$$

$$R(0)=0$$

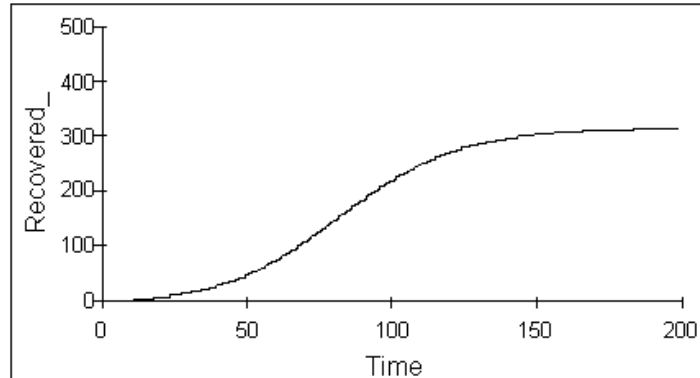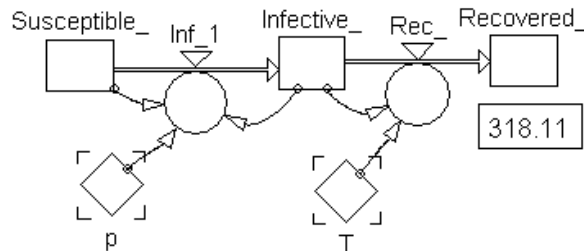Note that S+I+R =1001 is rather large! But I=1 will still cause problems in a deterministic approach!

*) or $F_1(t) = Po[\Delta t \cdot S(t) \cdot (1-(1-p)^{\Delta t \cdot I(t)})]/\Delta t$ to be exact.

Stochastic and deterministic SIR models

$$dS/dt = -p*S*I$$
$$dI/dt = p*S*I - I/T$$
$$dR/dt = I/T$$
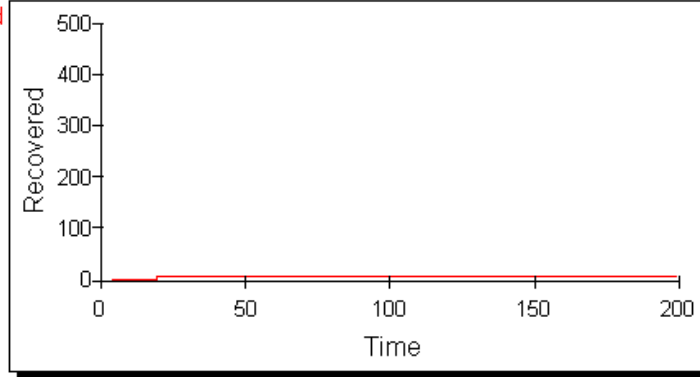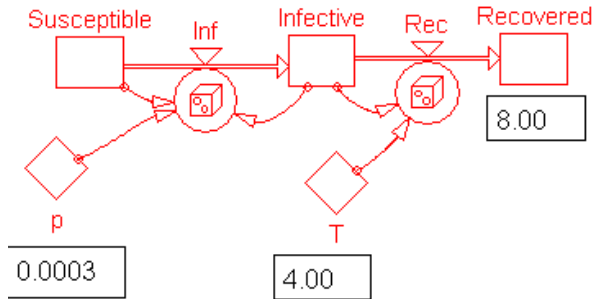
The stochastic and deterministic SIR model create very different results and conclusions!!!

Results

Probability distribution functions (pdf) over the number of infected individuals in different replications
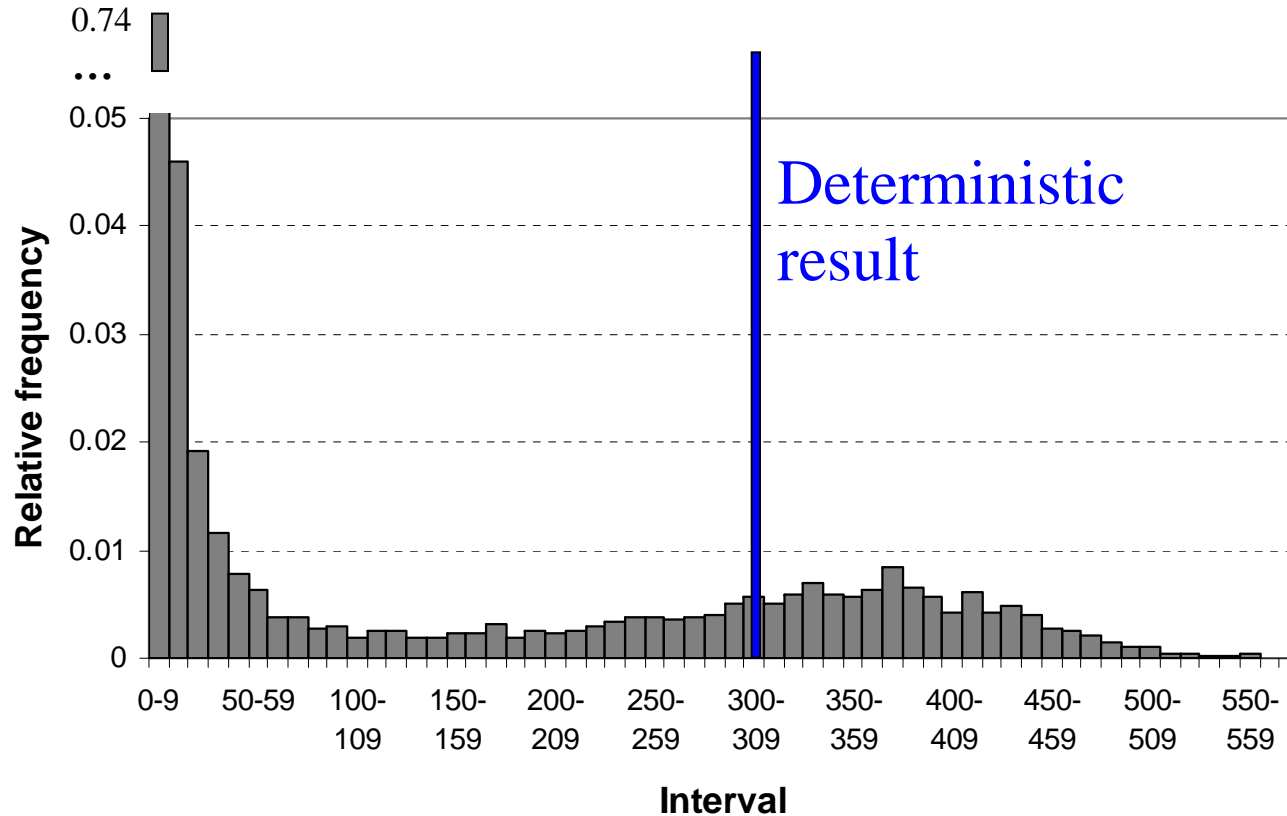
Comparison of the pdf of the number of susceptible individuals becoming infected during the epidemic from 10 000 replications of the stochastic model and the result of a deterministic model.

Results

Results obtained using the *stochastic* model from 10 000 replications and one from a *deterministic model.*

	Av. # infected (95% C.I.)	Min & Max	Av. length of epidemic (95% C.I.)	Min & Max
Stoch. Macro model	54.5 (52.1– 56.9)	0 & 560	21.9 (21.1– 22.6)	0.1 & 223.4
Determ. Macro model	318.3	318.5	∞ (207.4 for 1 case left)	∞

3) Feasible model structures
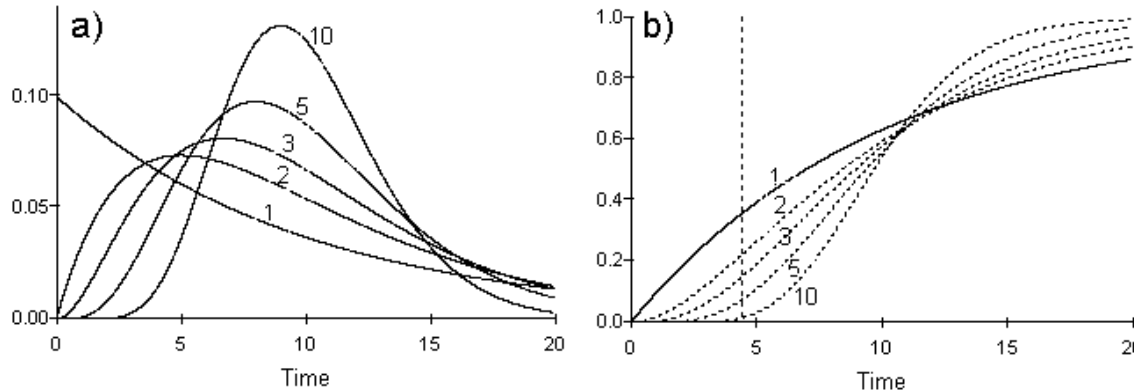

Figure a) $\Gamma(\alpha,\beta)$ distribution of sojourn times for $\alpha=1$, 2, 3, 5 and 10 with the same overall time constant $\beta=10$ time units. b) The cumulated output from these distributions.

Feasible and unfeasible representations of sojourn time in a stage:

Most processes in real life require a finite time. For e.g. biological/medical processes a finite time is needed for digesting, transportation of nutrients, defeating a disease, developing a cancer, growing an embryo, etc. For such a process a first order process is biologically *unfeasible* because it means that the probability of performing the process in almost zero time, i.e. finishing it within $(0,\varepsilon)$, is larger than finishing it in any other time interval $(t,t+\varepsilon)$ of length ε.
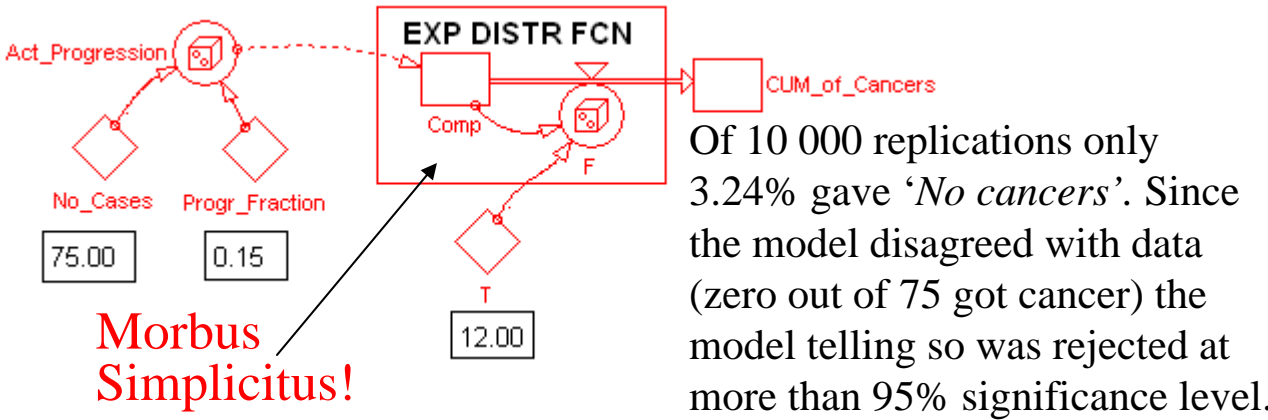

"Morbus Simplicitus"
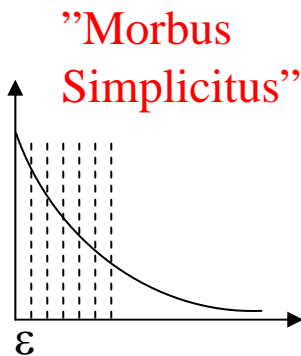
Nor is a second order delay feasible because there exists a finite probability to accomplish the task in zero time $(0,\varepsilon)$. A biologically feasible model, therefore, must be of order three or more.

Only a few non-biological processes, like radioactive decay, may be accurately modelled with a single compartment.
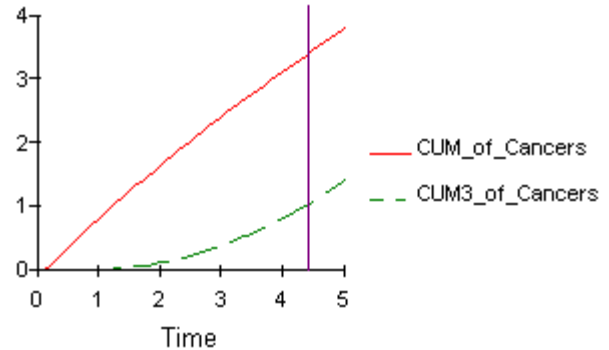
50

Example: G. H. Green's hypothesis: C*ervical cancer doesn't develop from cancer in situ?*

A medical professor in Auckland, New Zealand, followed during the 1960s a large number of women with *cancer in situ* of the cervix uteri by frequent smear tests to see if they progressed into *invasive* cancer.

75 patients, who showed persistent cancer in situ, were followed up for in average 53 months. *None of them developed invasive cancer*. To test the *Null hypothesis* that '*Cancer in situ causes cancer*' a Markov model, equivalent with the PoS model below was built:

Of 10 000 replications only 3.24% gave '*No cancers'*. Since the model disagreed with data (zero out of 75 got cancer) the model telling so was rejected at more than 95% significance level.

Deterministic results of the two models (which both are linear!).

… instead of the feasible model:

Of 10 000 replications 34.14% gave '*No cancers'*. No conclusions could then be drawn even at 70% significance level.
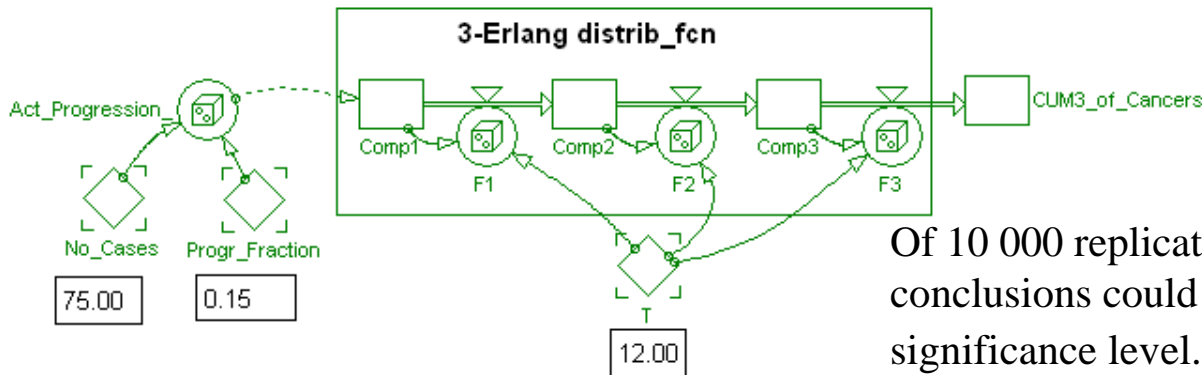
The professor, believing his exponential model, questioned the need for any treatment at all of cancer in situ. [Green, 1970]. This resulted later on in a large number of cancers and deaths!

<u>References</u>

• Gustafsson L. Poisson Simulation – A Method for Generating Stochastic Variations in Continuous System Simulation. Simulation 74:5, 264-274, May 2000.

• Gillespie D. Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. J. Chem. Phys. 115:1716-1733, 2001. (Gillespie calls PoS for "Tau-leap simulation")

• Gustafsson L. Studying Dynamic and Stochastic Systems Using Poisson Simulation. In Liljenström H. and Svedin U. Eds. *Micro – Meso – Macro: Addressing Complex Systems Couplings.* World Scientific Publishing Company. Singapore, 2004.

• Gustafsson L. Poisson Simulation as an Extension of CSS for the Modeling of Queuing Systems. Simulation 79:9, 528-541, September 2003.

• Gustafsson L. Tools for Statistical Handling of Poisson Simulation – Documentation of StocRes and ParmEst. Dept. of Biometry and Engineering. The Swedish University of Agricultural Sciences. 2004.

• Hedqvist T. Wanda for MATLAB – Methods and Tools for statistical handling of Poisson Simulation. Wanda StocRes for MATLAB – User Documentation, and Wanda ParmEst for MATLAB – User Documentation. Thesis Report, Master of Science in Engineering, Uppsala University, 2004.

• Gustafsson L and Sternad M. Bringing Consistency to Simulation of Population Models - Poisson Simulation as a Bridge between Micro and Macro Simulation, Math. Biosciences 209, pp 361-385, 2007.

• Gustafsson L. A Uniform Theoretical and Practical Base to Consistent Micro, Macro and State-Based Modelling. Manuscript.

• Gustafsson L. et al. System och Modell – en introduktion till systemanalysen, Studentlitteratur, 1982.

• Blom G. Sannolikhetsteori med tillämpningar (Bok A). Studentlitteratur 1984.

• Blom G. Statistikteori med tillämpningar (Bok B). Studentlitteratur 1984.

• Iversen OH. The Significant Correlation Between the Average Annual Age-Specific Incidence Rates of Malignant Melanoma and the Topography of the Norwegian West Coast. The Journal of Irreproducible Results. Vol 24. No. 3:17-19. 1978.

The steps in a PROJECT