ODE-based biomodeling

Ion Petre

Computational Biomodeling Laboratory
Turku Centre for Computer Science
Åbo Akademi University
Turku, Finland

http://combio.abo.fi/

1. GENERALITIES

The blind men and the elephant

John Godfrey Saxe's (1816-1887) version of the legend:

- First man (feeling the side): like a wall
- Second (the tusk): like a spear
- Third (the trunk): like a snake
- Fourth (the knee): like a tree
- Fifth (the ear): like a fan
- Sixth (the tail): like a rope

Source: Phra That Phanom chedi, Amphoe That Phanom, Nakhon Phanom Province, northeastern Thailand.

Picture downloaded from Wikipedia
Author: Pawyi Lee

Modeling

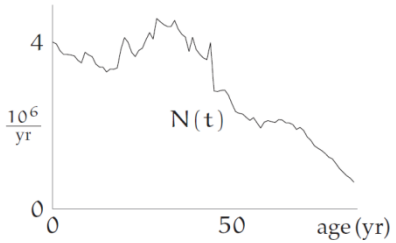
- What is a model?
  - A (partial) view of the reality
  - An abstraction of the reality
  - A representation of the (supposedly) main features of the reality, including the connections among them
  - For a given object of study, many models may be given, possibly focusing on different features of the object
- What a model is not
  - A model is not the reality
  - A model is not certain!
- Many types of models exist!

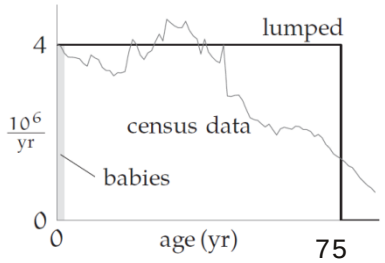
"All models are wrong, some are useful"

Box, G.E.P., Robustness in the strategy of scientific model building, in Robustness in Statistics, R.L. Launer and G.N. Wilkinson, Editors. 1979, Academic Press: New York.

An example: choose your hypothesis

- From S.Mahajan: *Street-fighting mathematics*, MIT Press, 2010
- Problem: how many babies (0-2 year olds) are in the US?
  - Exact solution: look at the plot with the birth dates of every person in the US
    - Huge effort; collected every 10 year by the US Census Bureau

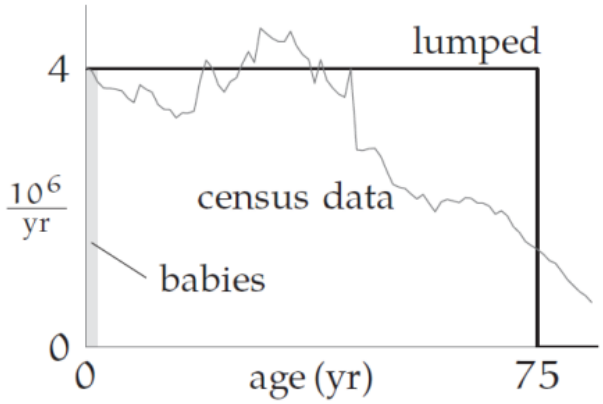




  - Approximation
    - US population: 300 million in 2008
    - Assume a life expectancy of 75 (a model where everybody still alive at 75 dies abruptly on their 75th birthday)
    - Lump the curve into a rectangle: width of 75, height to be calculated

Choose your hypothesis (continued)

- Height of the rectangle:
  - Total population of US: 300 million (2008)
  - Height: 300.000.000/75=4.000.000
- Result: calculate the area of a rectangle with height 4.000.000 and width 2
  - Result: 8.000.000 babies 0-2 year of age
  - Compare with the Census Bureau's figure: 7.980.000 !!



Model: Life inside a cell

- Simplifications often made by biomodelers
  - Cell is “like a bag of chemicals floating in water”
  - Metabolites flow around chaotically
  - In a DNA sequence, A is always matched with T, C always with G
  - Processes are isolated from each other and from the environment
  - …
- The reality is surprisingly complex
  - The cell has a skeleton, gives it flexibility
  - Many intracellular boundaries, many specialized organelles
  - … synchronization, signal propagation, cooperation
  - Some particles do move chaotically, but some others are transported
  - Some aspects are discrete (on/off), some others are continuous-like (always on, variable speed)
  - Huge pressure, crowded

A view on “The Inner Life of a Cell” (Harvard University, 2006):

Artistic representation of metabolite transportation, protein-protein binding, DNA replication, DNA ligase, microtubule formation/dissipation, protein synthesis, …

Mathematical modeling

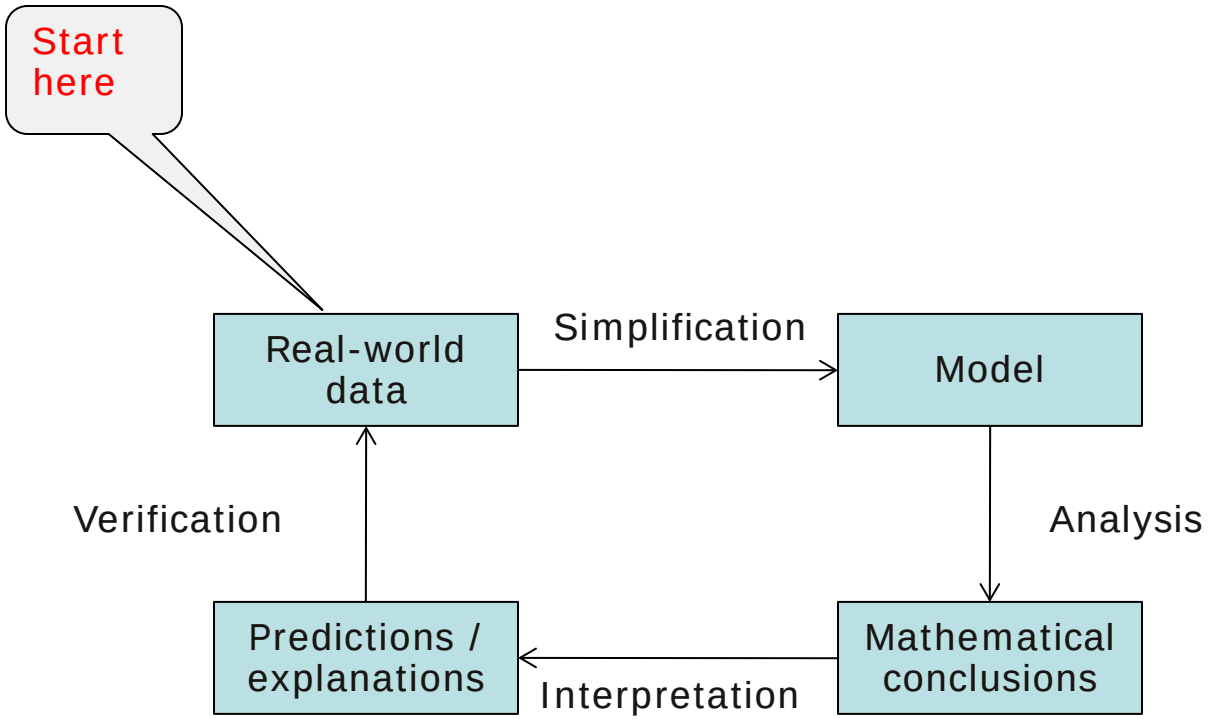
- We focus in this lecture on mathematical models
  - As we saw, (many) other types of models exist
  - “Model” is indeed a very overloaded word
  - In this lecture a model is a *mathematical* representation of the reality
  - Models that mimic the reality by using the language of mathematics

- Goal of the lecture
  - An introduction to the process of mathematical modeling
  - Give a number of techniques used for:
    - Building a model
    - Analyzing a model
  - Main tools: (systems of) ODEs

Mathematical models

- Starting point for modeling: divide the world into 3 parts
  - Things whose effects are neglected
    - Ignore them in the model
  - Things that affect the model but whose behavior the models is not designed to study
    - External variables, considered as parameters, input, or independent variables
  - Things the model is designed to study the behavior of
    - Internal (or dependent) variables of the model
- Deciding what to model and what not is difficult
  - Wrong things neglected: the model is no good
  - Too much included: hopelessly complex model
  - Choose the internal variables wrongly: the model will not capture its target
  - How general should the model be: model a table (any table?) or the specific table in front of the modeler

Modeling cycle



Model validation

- Any model must always be subjected to experimental validation against the reality
- A model may be invalidated by experimental data
- No set of experimental data can confirm the “truthfulness” of a model

2. FORMULATING AN ODE MODEL

Modeling with differential equations

- Modeling strategy
  - We model the change in the values of all variables:
  - Future value = present value + change
  - We describe the change as a function of the current values of all variables
  - If the process takes place continuously in time, it leads to differential equations
- Each species s modeled as a function $s: R_+ \rightarrow R_+$
  - Concentrations
- Dependencies expressed as systems of ODEs

- Bad news: the equations are often non-linear and in general they cannot be solved analytically

Example: population growth

- Example: population growth (the Malthus model, 18th century)
  - Problem: Given a population's size P_0 at time $t=t_0$, predict the population level at some later time t_1
  - We consider two factors: birthrate and death rate. We ignore immigration and emigration, living space restrictions, food avail, etc.
  - Birthrate: influences by many factors, including infant mortality rate, availability of contraceptives, abortion, health care, etc.
  - Death rate: influences by sanitation, public health, wars, pollution, medicine, etc.
  - Assume that in a small interval of time, a percentage b of the population is newly born and a percentage c of the population dies
  - We write an equation for the change in the population: $dP/dt = bP(t) - cP(t)$, i.e., $dP/dt=(b-c)P(t)$
  - The solution is: $P(t)=P_0\exp((b-c)(t-t_0))$

Example: population growth

- Verifying the model: numerical fit and validation
  - Population of US in 1990: 248.710.000 and in 1970: 203.211.926
  - Plugging in these numbers, we obtain that b-c=0.01
  - Predict the population in 2000: 303.775.080
  - The real population level in 2000: 281.400.000.
  - The model prediction is about 8% off the mark. Not too bad!
  - Predict the population level in 2300: 55.209.000.000.000!!!
  - **Conclusion:** the model is unreasonable over long periods of time

A refined model for population growth

- In the basic model we have assumed that the change in the population is proportional to the current population level: $dP/dt = kP(t)$

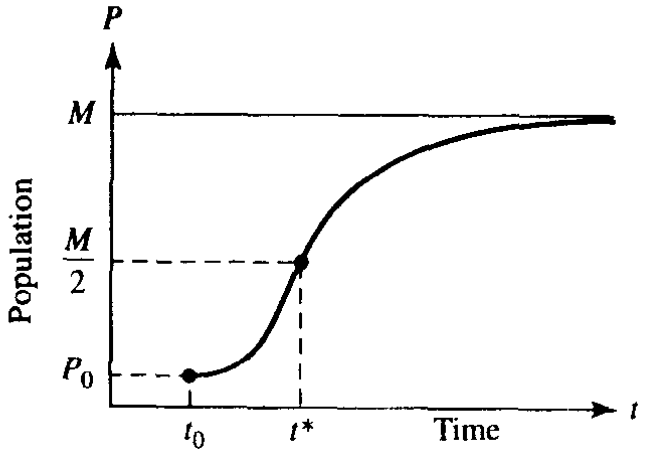
- Assume that k is not constant
  - Assume that it depends on the population level
  - For example: as the population increases and gets closer to a maximum level M, k decreases
  - One possible (simple, linear) model for this: $k=r(M-P(t))$
  - Our equation: $dP/dt=r(M-P(t))P(t)$

  - Such a population model for US was proposed in 1920, with M=197.273.522, determined based on census figures for 1790, 1850, 1910

  - Verifying the model: very good predictions up to 1950, too small predictions for 1970, 1980, 1990, 2000
  - Not surprising: immigration, wars, advances in medicine not considered

  - **Note: Verifying the model on the growth of yeast in culture gives excellent predictions**

Figure 10.4
Graph of the limited growth model

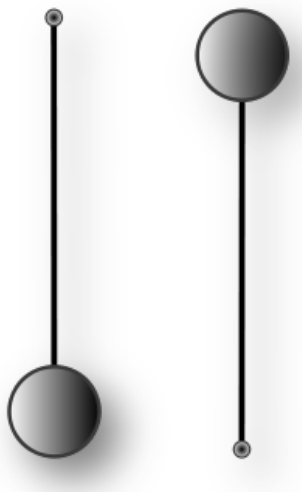


Giordano et al. A first course in mathematical modeling. (3rd edition), Page 375

Stable and unstable equilibria / steady states

- Equilibrium point / steady state: one where all ODEs in the model are zero

- Types of equilibrium points (informal definitions)
  - Stable: starting from a nearby initial point will give an orbit that remains nearby the original orbit
    - Asymptotically stable (attractor): starting from a nearby initial point will give an orbit that converges towards the original orbit
    - Example: a pendulum in the lowest position
  - Unstable: starting from a nearby initial point *may* give an orbit that goes away from the original orbit
    - Example: a pendulum in the highest position



Stable-unstable equilibrium

Source for picture: Wikipedia

Graphical solutions

- Consider autonomous systems of first-order ODEs $dx_i/dt = f_i(x_1, x_2, \ldots x_n)$
  - not time dependent
  - consider its solution as describing a trajectory in the n-dimensional plane, with coordinates $(x_1(t), x_2(t), \ldots, x_n(t))$
    - convenient to think about it as the movement of a particle
  - Having an autonomous system implies that the direction of movement from a given point on the trajectory only depends on that point, not on the time when the particle arrived in that point
    - Consequence: only one trajectory going through any given point
    - Equivalently: two different trajectories cannot intersect
    - Consequence: no trajectory can cross itself unless it is a closed curve (periodic)
  - the n-dimensional plane $(x_1, x_2, \ldots x_n)$ is called a phase plane

Graphical solutions

- Consider autonomous systems of first-order ODEs $dx_i/dt=f_i(x_1,x_2,...x_n)$
  - if $(e_1,e_2,...e_n)$ is an equilibrium point, then the only trajectory going through that point is the constant one
    - Consequence: a trajectory that starts outside an equilibrium point can only reach the equilibrium asymptotically, not in a finite amount of time

- The resulting motion of a particle can have one of the following 3 behaviors:
  - approaches an equilibrium point
  - moves along or approaches asymptotically a closed path
  - at least one of the trajectory components becomes arbitrarily large as t tends to infinity

Example: a competitive hunter model

- Assume we have a small pond that we desire to stock with game fish, say trout and bass. The problem we want to solve is whether it is possible for the two species to coexist

- Model formulation
  - The change in the level of trout X(t):
    - Assuming food is available at an infinite rate: increase of trout population at a rate proportional to its current level: aX(t)
    - Assume that the space is a limitation for the co-existance of the two species in terms of the living space. The effect of the bass population is to decrease the growth rate of the trout population. The decrease is approximately proportional to the number of possible interactions between trout and bass: -bX(t)Y(t)
    - Equation: $dX/dt = aX(t) - bX(t)Y(t)$
  - Similar reasoning for the level of bass Y(t):
    - $dY/dt = mY(t) - nX(t)Y(t)$

Example: a competitive hunter model

- Model: dX/dt = aX(t) – bX(t)Y(t), dY/dt=mY(t) – nX(t)Y(t)

- Question: can the two populations reach an equilibrium where both are non-zero
  - Answer: ax-bxy=0, my-nxy=0
  - Solution: either x=y=0, or x=m/n, y=a/b

- Difficulty: impossible to start with exactly the equilibrium values (they might not even be integers)
  - so, we cannot expect to start in an equilibrium point
  - study the property of the equilibrium, hoping it is a stable one

Example: a competitive hunter model (continued)

- Equilibrium points: (0,0), (m/n, a/b)
- Additional question: what is the behavior if we start close to the equilibrium point?
  - Solution: we study the tendency of X(t), Y(t) to increase/decrease around the equilibrium point. For this, we study the sign of the derivatives of X(t), Y(t)
  - dX/dt≥0 ⇔ aX-bXY≥0 ⇔ a/b≥Y
  - dY/dt≥0 ⇔ mY-nXY≥0 ⇔ m/n ≥ X

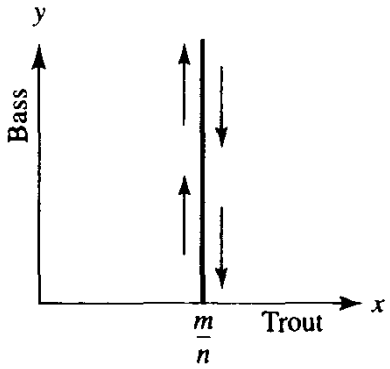


Figure 11.4
To the left of $x = m/n$ the trajectories move upward; to the right they move downward

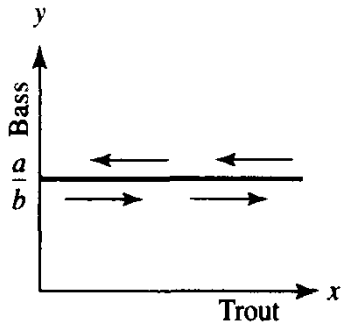
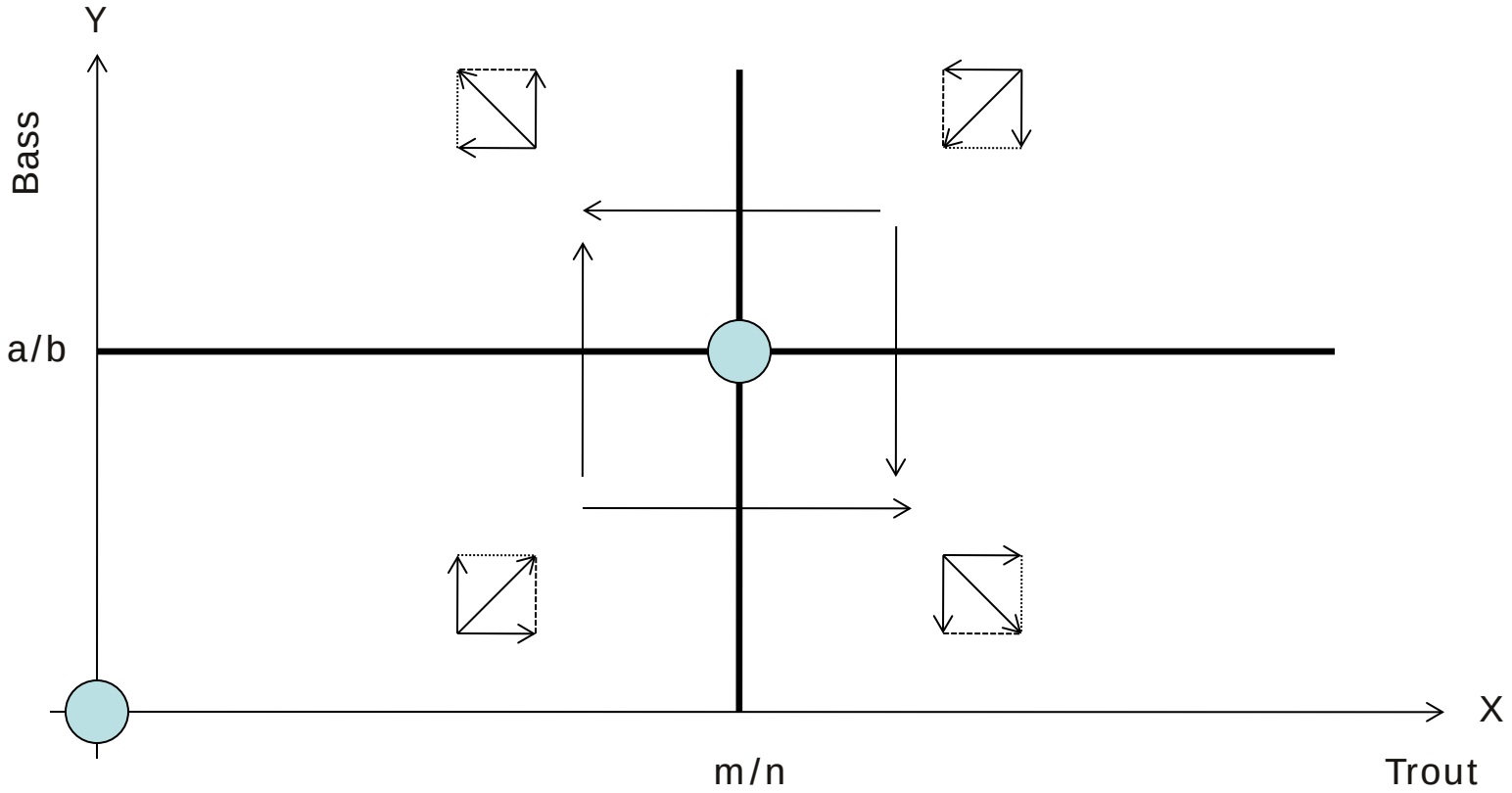


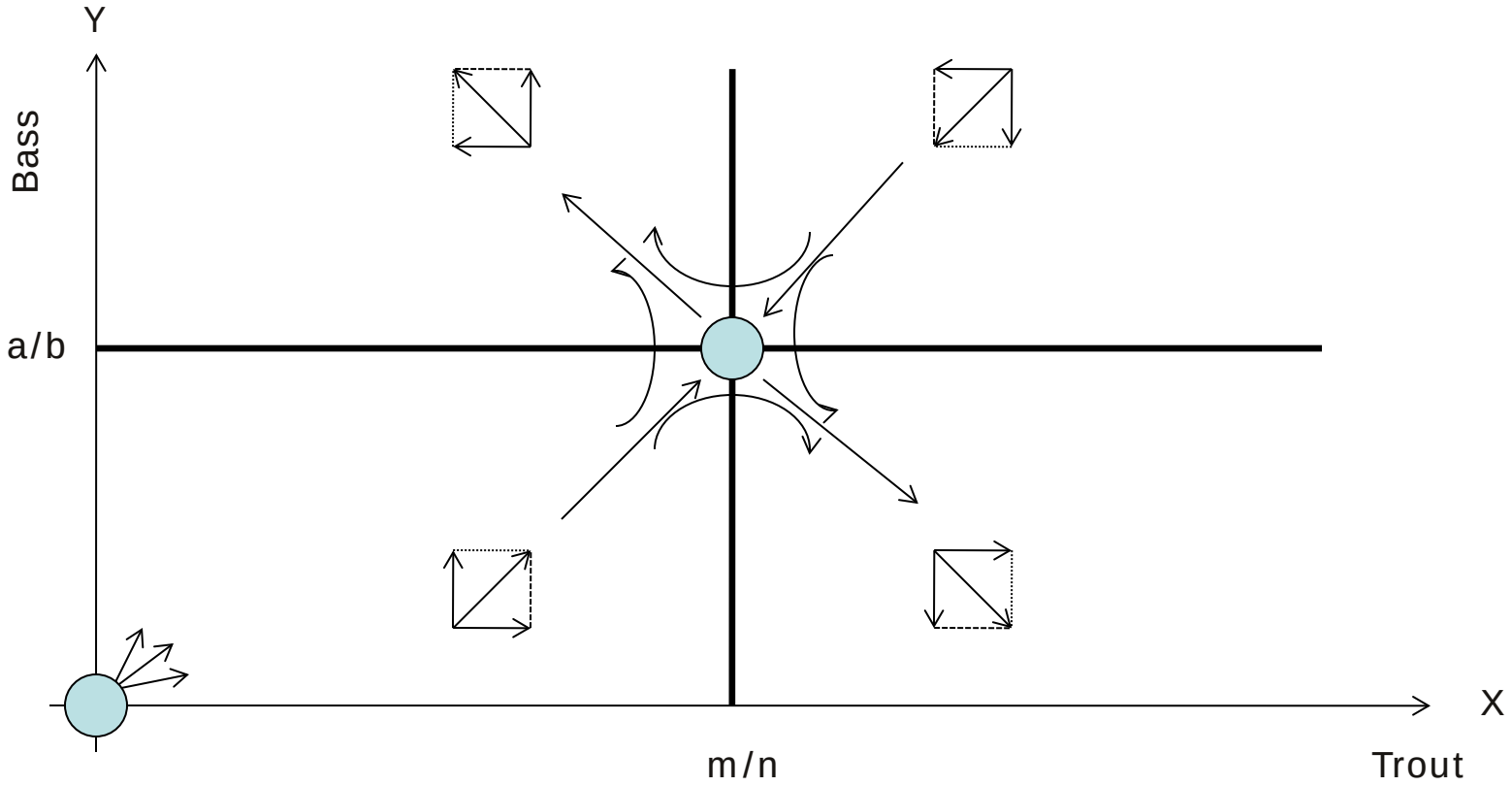
Figure 11.5
Above the line $y = a/b$ the trajectories move to the left; below the line they move to the right

Giordano et al. A first course in mathematical modeling. (3rd edition), Page 421

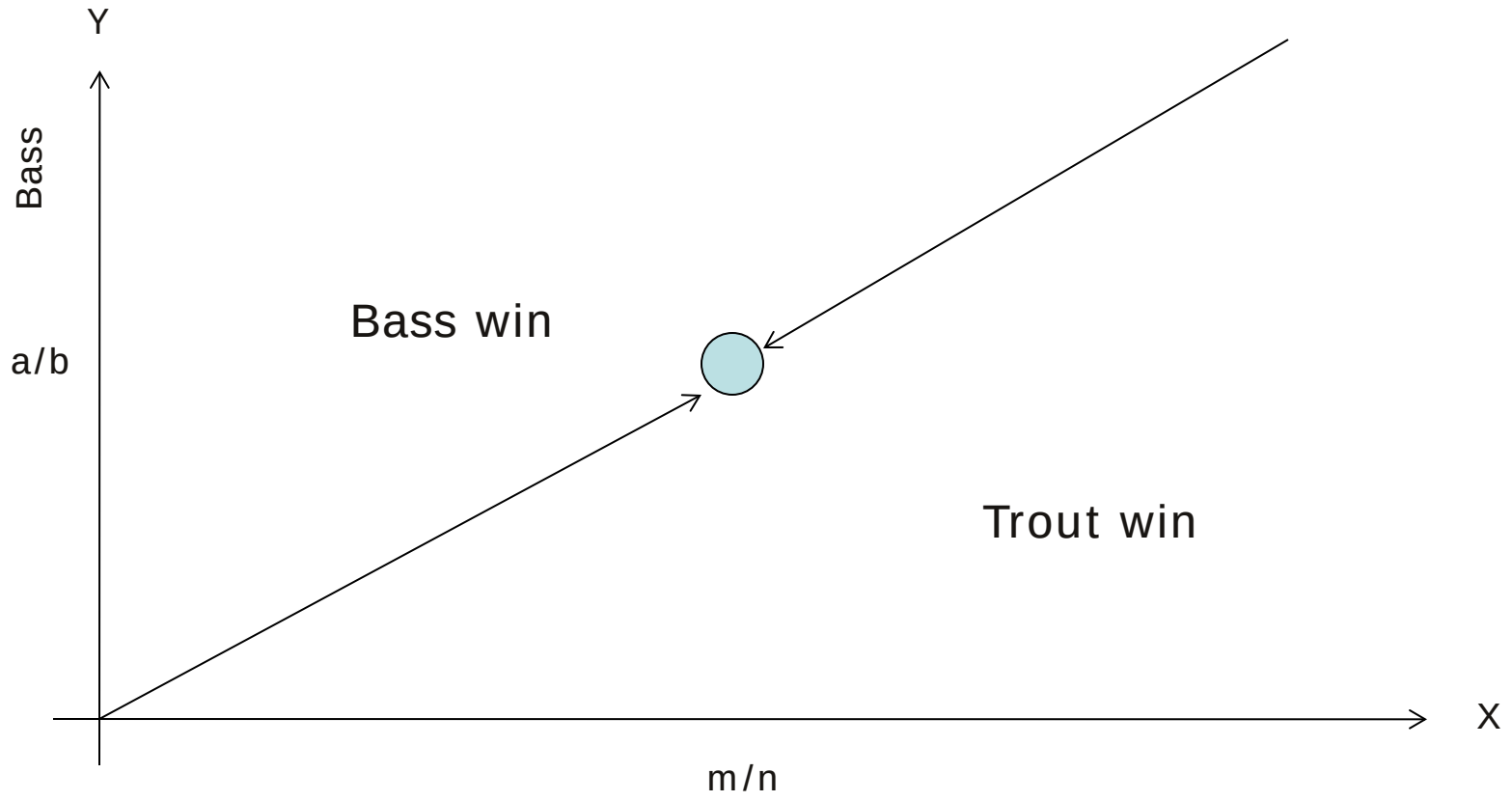
Graphical analysis of the trajectory directions



Graphical analysis of the trajectory directions around the equilibria



Graphical analysis of the trajectory directions



Conclusion: the co-existance of the two species is highly improbable

Limits of graphical analysis

- Not always possible to determine the nature of the motion near an equilibrium based on graphical analysis
  - Example: the behavior in Fig 11.9 through graphical analysis is satisfied by all 3 trajectories in Fig 11.10
  - Example: The trajectory in Fig 11.10c could be either growing unboundedly or approach a closed curve

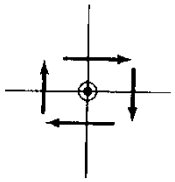


Figure 11.9
Trajectory direction near a rest point

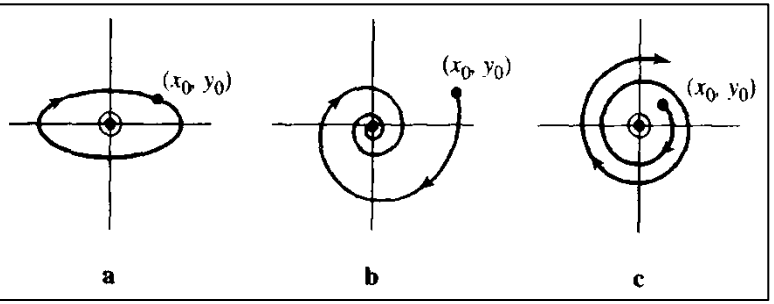


Figure 11.10
Three possible trajectory motions: **a.** periodic motion, **b.** motion toward an asymptotically stable rest point, and **c.** motion near an unstable rest point

Giordano et al. A first course in mathematical modeling. (3rd edition), Page 422-423

3. KINETIC MODELING OF REACTION NETWORKS

Modeling: from “art” to automatization

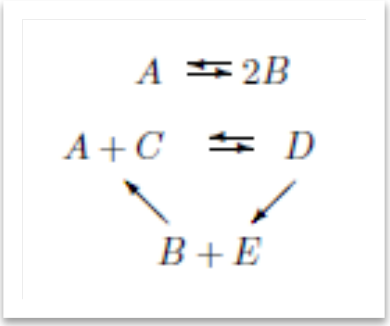
- The type of modeling shown so far required a great deal of creativity from the modeler in formulating the model

- For the remaining of this lecture:
  - models as reaction networks
  - separate the formulation of the model in two different stages
    - first identify the variables and describe their interactions using a simple syntax: chemical reaction networks (or sometimes rules)
    - second, build the associated mathematical model
      - this is uniquely determined by the first part and by the choice of a modeling principle (such as mass-action, Michaelis-Menten, etc.)

Chemical reaction networks

- Chemical reaction network:
  - finite set of species
  - finite set of reactions represented as rewriting rules
  - input on the left hand side, output on the right hand side
  - multiplicities indicated in the rewriting rule

- Example:



  - The inputs (the reactants) are consumed in the number of copies indicated by the reaction and the output (the products) are created with the indicated multiplicity

Stoichiometry

- The **stoichiometric coefficients** denote the quantitative proportion in which substrate and product molecules are involved in a reaction.
- In the case of a reversible reaction the stoichiometric coefficient values depend on the chosen direction. Usually, the direction is chosen to be ‘left-to-right’.

For a reversible reaction

$$R_\mu: m_{\mu,1}S_1 + m_{\mu,2}S_2 + \ldots + m_{\mu,N}S_N \leftrightarrow m'_{\mu,1}S_1 + m'_{\mu,2}S_2 + \ldots + m'_{\mu,N}S_N$$

the stoichiometric coefficients are:

$$m'_{\mu,1} - m_{\mu,1}, m'_{\mu,2} - m_{\mu,2}, \ldots, m'_{\mu,N} - m_{\mu,N}.$$

Stoichiometric matrix

- Stoichiometric matrix $N=(n_{ij})_{sxr}$: n_{ij} denotes the stoichiometric coefficient of species S_i in reaction R_j.

- Example:

Reaction network

r_1: A ⇄ 2B
r_2: A+C ⇄ D
r_3: D → B+E
r_4: A+B ⇄ D+B

Stoichiometric matrix

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{bmatrix} -1 & -1 & 0 & -1 \\ 2 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Stoichiometric matrix

- The stoichiometric matrix contains valuable information about the structure of the network
  - calculate the mass conservation relations
  - calculate the steady states
  - which combinations of individual fluxes are possible in steady state
  - calculate sensitivity coefficients

- Discuss some of them in the rest of this lecture

Chemical reaction networks

- A chemical reaction network gives rise to a dynamical system
  - describe how the state of the network changes over time
  - State of the system: the concentration of all species at time t
  - Question: how do we express the change in the concentrations in time?
    - General kinetics: associate to each reaction a function specifying how fast its reactants/products are consumed/produced – **reaction rate**
    - Simultaneous update (e.g., as a system of ODEs) of all species

- In this lecture we discuss a few kinetic laws
  - Law of mass-action
  - Enzyme kinetics
    - Michaelis-Menten
    - Inhibition

TUCS

Mass-action kinetics

Mass-action models for biochemical reaction networks

- The mass action kinetics model is derived based on the Boltzmann's kinetic theory of gases and is justified under the assumption of
  - **constant temperature** and
  - **fast enough diffusion in the cell**,

  which ensures that the mixture of substances is "well-stirred", i.e. homogenously distributed in a fixed volume *V*.

The law of mass action

- Waage, Guldberg 1864, Guldberg, Waage 1867, 1879
  - The reaction rate is proportional to the probability of a collision of the reactants
  - The probability of the collision is proportional to the concentration of reactants to the power of the molecularity
  - For a reaction $n_1A_1+n_2A_2+\ldots+n_mA_m \rightarrow$ products, the reaction rate is

$$v = kA_1^{n_1}(t)A_2^{n_2}(t)\ldots A_m^{n_m}(t)$$

  - For a reversible reaction $n_1A_1+n_2A_2+\ldots+n_mA_m <-> r_1B_1+r_2B_2+\ldots+r_sB_s$, the reaction rate is $v=v_1-v_2$, where v_1 is the rate of the "left-to-right" reaction and v_2 is the rate of the "right-to-left" reaction

Writing the mass-action ODE model

- The reaction rate gives the amount with which the concentration of every metabolite involved in the reaction changes per unit of time
  - For a consumed metabolite, the change will be –v(t)
  - For a produced metabolite, the change will be v(t)

- Example

  - For a reaction A->, the reaction rate is v(t)=kA(t)
    - dA/dt=-kA(T)

  - For a reaction A+B→C, the reactions rate is v(t)=kA(t)B(t), for some constant k
    - dA/dt=-kA(t)B(t), dB/dt=-kA(t)B(t), dC/dt=kA(t)B(t)

Coupled reactions

- Assume we have a set of reactions
  - $A+B->C$
  - $A+2C<->B$
  - $C->2A$

- Write the rates of all reactions
  - $v_1=k_1AB$
  - $v_2=k_2^+AC^2-k_2^-B$
  - $v_3=k_3C$

- Write the differentials: for each metabolite, consider all reactions where it participates
  - $dA/dt=-v_1-v_2+2v_3=-k_1AB-k_2^+AC^2+k_2^-B+2k_3C$
  - $dB/dt=-v_1+v_2=-k_1AB+k_2^+AC^2-k_2^-B$
  - $dC/dt=v_1-2v_2-v_3=k_1AB-2k_2^+AC^2+2k_2^-B-k_3C$

A predator-prey model

- A model where we have two species, one being the primary food source for the other

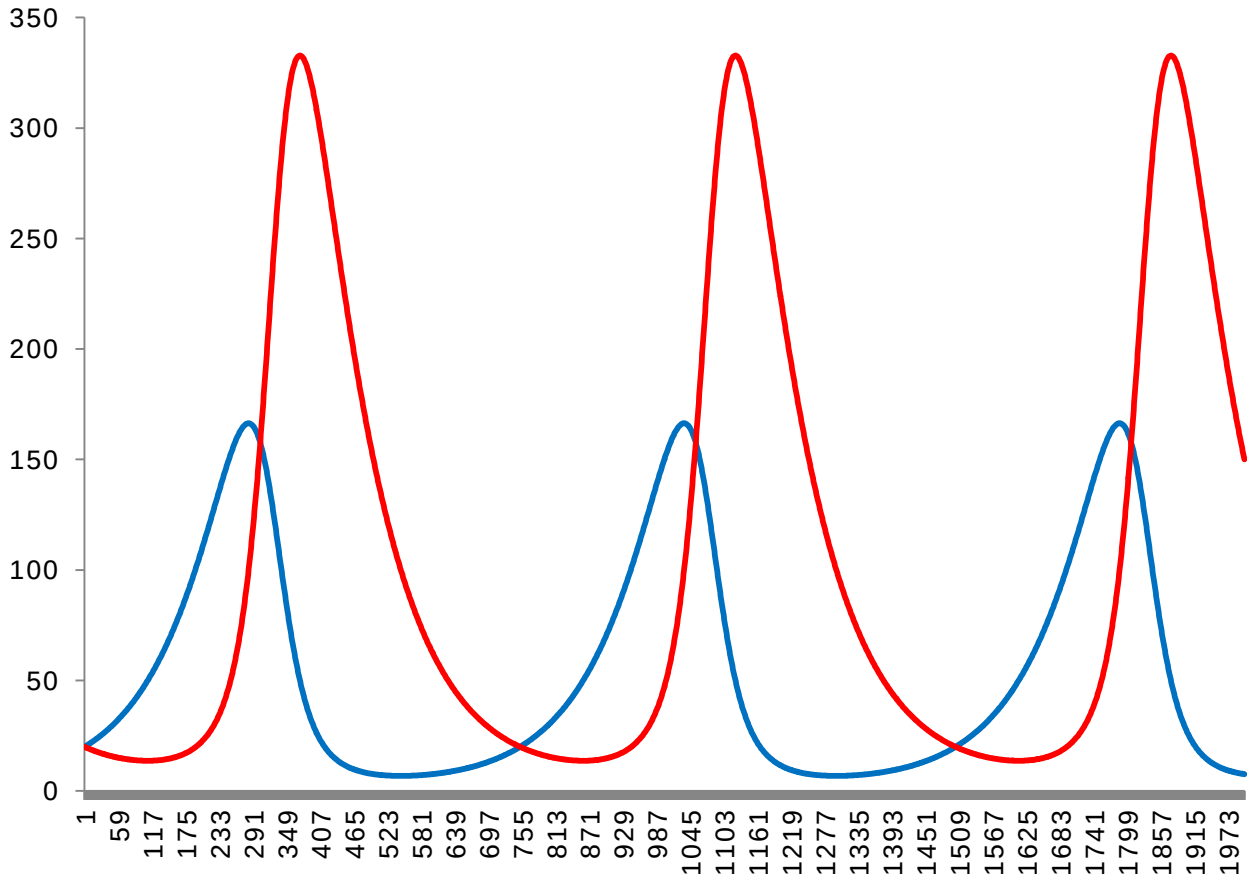
- Problem
  - Whales, krill
  - Whales eat the krill; the krill live on the plankton in the sea
  - If whales eat too much krill, then the krill ceases to be abundant, and the whales will starve or leave the area
  - As the population of whales declines, the population of krill increases
  - This makes the population of whales grow again, etc.

A predator-prey model

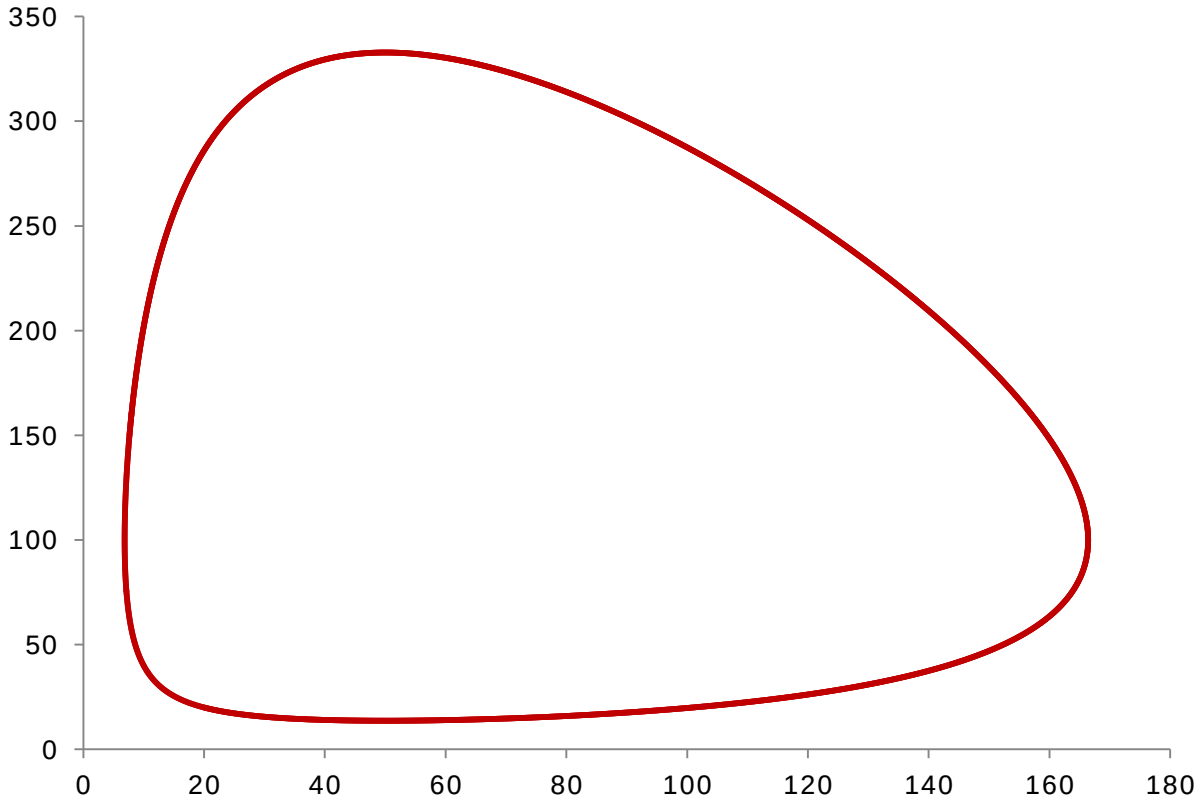
- Assumptions and model formulation (the Lotka-Volterra model)
  - The krill population x(t), the whale population y(t)
  - The model as a reaction network
    - Krill multiplies (assume infinite plankton as a food source for krill): X→2X (a)
    - Whales eat krill: X+Y→Y (b)
    - Whales die: Y→ (m)
    - Whales multiply only if there is krill: X+Y→X+2Y (n)
  - The asociated mass-action ODE model:
    - dx/dt= ax(t)-bx(t)y(t)
    - dy/dt=-my(t)+nx(t)y(t)
  - We have the model formulated as the system of the 2 ODEs

- Equilibrium points (or steady states) (x_s,y_s)
  - dx/dt=dy/dt=0
  - $(x_s,y_s)=(m/n,a/b)$ or $(x_s,y_s)=(0,0)$

A predator-prey model: numerical integration



A predator-prey model: phase portrait



Comments

- Mass-action kinetics leads to non-linear ODE models

- Even though non-linear, there is clear regularity in the structure of a mass-action ODE model
  - Reactants consumed with the same rate as products are coming
    - The model is in fact linear in terms of reaction rates
  - Well-specified form of the ODEs
    - No longer have the option of obtaining a perfect fit by changing the form of the math model
    - Can only fit the model through the kinetic rate constants
    - Fitting such models is a difficult problem
  - Very different (easier!) to analyze an ODE system coming from a mass-action reaction network than to analyze an arbitrary system of ODEs

Kinetics of enzymatic reactions

Enzymatic reactions

- Enzymes play a catalytic role in biology
  - Highly specific
  - Remain unchanged by the reaction
  - One enzyme molecule catalyzes about a thousand reactions per second
  - Rate acceleration of about 10^6 to 10^{12}-fold compared to uncatalyzed, spontaneous reactions

- Enzymatic reactions usually described using more complicated models than mass-action

- Example: A is transformed into B through the help of enzyme E
  - $A+E \rightarrow B+E$
  - Mass-action: $dB/dt=kA(t)E(t)$
  - Usual model for this reaction: $dB/dt=v_{max}A(t)/(K_M+A(t))$

- Discuss in the following the modeling of enzymatic reactions

Enzymatic reactions

- Brown (1902) proposed the following reaction model for irreversible enzymatic reactions: E+S <-> E:S -> E+P
- The associated mass-action ODE model:

1. $dS/dt=-k_1ES+k_{-1}(E:S)$
2. $d(E:S)/dt=k_1ES-(k_{-1}+k_2)(E:S)$
3. $dE/dt=-k_1ES+(k_{-1}+k_2)(E:S)$
4. $dP/dt=k_2(E:S)$

- Michaelis, Menten (1913): assume that the first part of the reaction is much faster than the second one: $k_1, k_{-1} >> k_2$
- Briggs, Haldane (1925): in some conditions, it may be assumed that E:S reaches quickly a steady state
  - This is the case if $S(0)>>E$ (the enzyme is saturated by the substrate)
- Both assumptions lead to assuming that $d(E:S)/dt=0$, i.e., E:S is constant
  - investigate what consequences this assumption has

Enzymatic reactions

- E+S <-> E:S -> E+P

  1. $dS/dt=-k_1ES+k_{-1}(E:S)$
  2. $d(E:S)/dt=k_1ES-(k_{-1}+k_2)(E:S)$
  3. $dE/dt=-k_1ES+(k_{-1}+k_2)(E:S)$
  4. $dP/dt=k_2(E:S)$

- If $d(E:S)/dt=0$ (i.e., E:S is constant):

  5. $k_1ES-(k_{-1}+k_2)(E:S)=0$

- Rewrite equation 1 into:

  1'. $dS/dt=-k_2(E:S)$

- Clearly, E+E:S is constant in the model, say $E+E:S=E_{tot}$,i.e.,

  6. $E=E_{tot}-E:S$

- Then equation 5 can be rewritten:
  $k_1(E_{tot}-E:S)S-(k_{-1}+k_2)(E:S)=0$

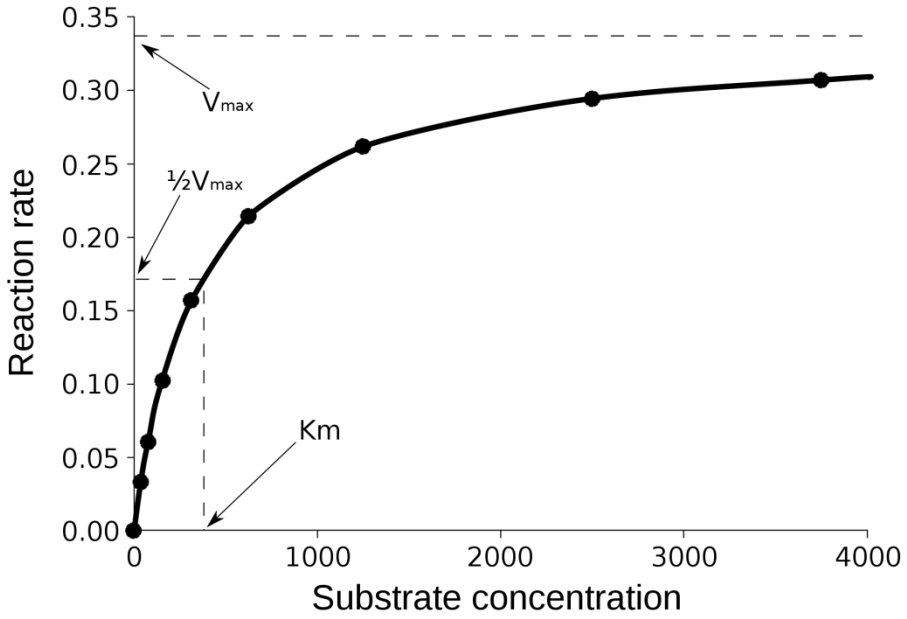
- Then $E:S=E_{tot}S/(S+(k_{-1}+k_2)/k_1)$

- Define $v_{max}=k_2E_{tot}$, $K_m=(k_{-1}+k_2)/k_1$

- In other words:
  $dS/dt=-v_{max}S/(S+K_m)$,
  $dP/dt=v_{max}S/(S+K_m)$

  - Where v_{max} is the maximal rate (velocity) that can be obtained when the enzyme is completely saturated with substrate, $v_{max}=k_2E_{tot}$
  - K_m is the Michaelis constant, $K_m=(k_{-1}+k_2)/k_1$, equal to the substrate concentration that yields the half-maximal reaction rate

Michaelis-Menten kinetics



Source of the figure: Wikipedia

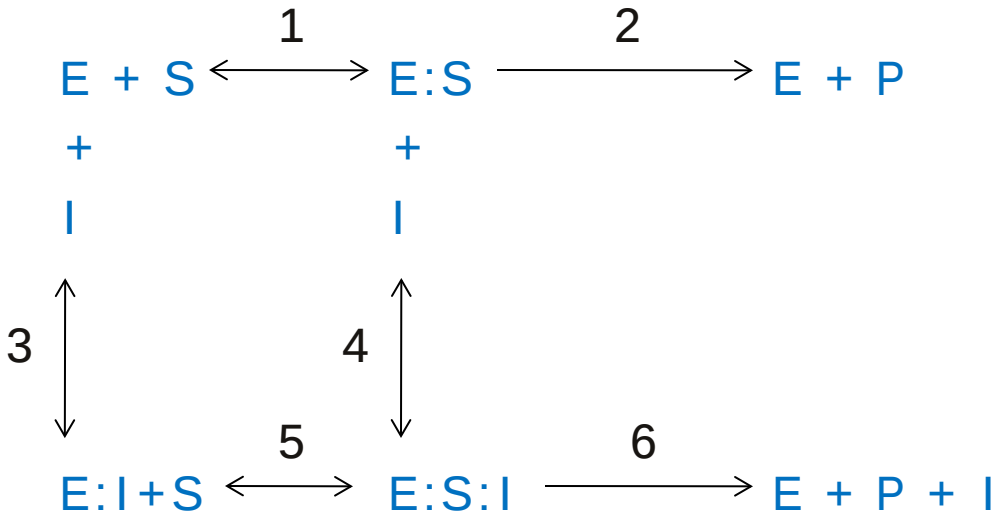
- Low substrate concentration S: reaction rate increases almost linearly with S
- High substrate concentration S: reaction rate is almost independent of S
- v_{max} is the maximal reaction rate that can be achieved for large substrate concentration
- The Michaelis-Menten constant is the substrate concentration that gives $1/2v_{max}$

Regulation of enzyme activity

- Enzymes catalyze reactions

- Other reactions may influence (regulate) the activity of the enzyme
  - Inhibitors
  - Activators
  - Widely found in metabolic pathways

General scheme of inhibition in Michaelis-Menten kinetics



- *Reactions 1,2: Michaelis-Menten*
- *Reactions 1,2,3: competititve inhibition*
- *Reactions 1,2,4: uncompetitive inhibition*

- Reactions 1,2,3,4,5: non-competitive inhibition
- Reactions 1,2,3,4,5,6: partial inhibition

4. ANALYSIS OF ODE MODELS

Mass-conservation relations

Mass conservation relations

- Frequently, the concentrations of several substances involved in biochemical reaction networks are included in so-called conservation sums.

- A characteristic feature of such substances is that they are neither produced nor degraded, however they can form complexes with other species or be part of other species.

Mass conservation relations

- Example

  reactions:

  $2A \rightleftarrows A_2$

  $A_2 + B \rightleftarrows A_2{:}B$

  $A_2{:}B \rightarrow C + A_2{:}B$

  $C \rightarrow$

$$\boldsymbol{N} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

  species:

  A, A_2, B, $A_2{:}B$, C

  - The total amounts of A and B are conserved in time. Neither of them is produced nor degraded.
    $1\times\#A + 2\times\#A_2 + 2\times\#A_2{:}B = \text{const.}$
    $1\times\#B + 1\times\#A_2{:}B = \text{const.}$

Mass conservation relations

- To identify the conservation relations we solve the following equation in matrix G:

$$GN = 0$$

  - Indeed, for such G:

$$G\frac{dS}{dt} = GNv = 0.$$

- Example (continued):

$$\mathbf{S} = \begin{bmatrix} \mathrm{A} \\ \mathrm{A}_2 \\ \mathrm{B} \\ \mathrm{A}_2 : \mathrm{B} \\ \mathrm{C} \end{bmatrix} \quad \mathbf{N} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \quad \mathbf{GN} = \mathbf{0}$$

Mass conservation relations

- The number of independent rows of G, i.e. the number of conservation relations, is equal to s-Rank(N).
  - In the example s=5 and Rank(N)=3. It follows that G contains 2 independent rows, i.e., there are two mass conservation relations.
  - Observation: if the stoichiometric matrix has full rank, it follows that the system has no conservation relations.

Mass conservation relations

- Conservation relations can be used to reduce the system of differential equations dS/dt=Nv describing the dynamics of a reaction network.

- Each conservation relation leads to one more dependent variable, that can be expressed in terms of the independent variables and eliminated from the system of ODEs

- Always check the biological meaning of each mass conservation relation

Steady states

Steady state

- **Steady state** – one of the basic concepts of dynamical systems theory, extensively utilized in modelling.
- Steady states (*stationary states*, *fixed points, equilibrium points*) are determined by the fact that the values of all state variables remain constant in time.
- In steady state it holds for a reaction network that

$$\frac{dS}{dt} = Nv = 0$$

- Solve the resulting **algebraic equation** in the unknowns $S_1,\ldots,S_s$ (the s components of the steady state)

Steady state

- Example (mass action kinetics)

$2A \rightarrow B$ (k_1)

$A+B \rightleftarrows C$ (k^+_2, k^-_2)

Steady state algebraic equations ($[A]_0$, $[B]_0$, and $[C]_0$ are unknowns)

$$\underbrace{\begin{bmatrix}0\\0\\0\end{bmatrix}}_{\mathbf{0}} = \underbrace{\begin{bmatrix}-2 & -1\\1 & -1\\0 & 1\end{bmatrix}}_{N} \cdot \underbrace{\begin{bmatrix}\mathbf{k}_1[\mathbf{A}]_0^2\\ \mathbf{k}_2^+[\mathbf{A}]_0[\mathbf{B}]_0 - \mathbf{k}_2^-[\mathbf{C}]_0\end{bmatrix}}_{\mathbf{v}_0}$$

or

$$\begin{cases} 0 = -2k_1[A]_0^2 - k_2^+[A]_0[B]_0 + k_2^-[C]_0 \\ 0 = k_1[A]_0^2 - k_2^+[A]_0[B]_0 + k_2^-[C]_0 \\ 0 = k_2^+[A]_0[B]_0 - k_2^-[C]_0 \end{cases}$$

Elementary fluxes

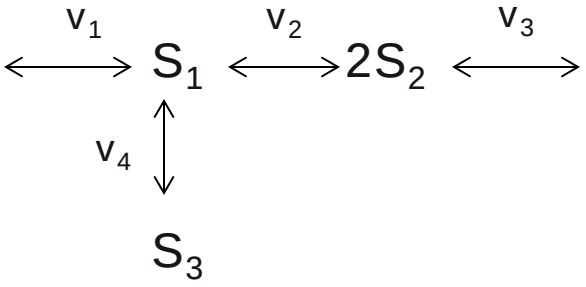
Elementary flux modes

- Concept of elementary flux mode
  - a minimal set of enzymes (or, in other words, reactions) that can operate at steady state
  - the smallest sub-networks that allow a bionetwork to function at steady state
  - a *minimal* combination of reactions whose combined effect maintains the network in steady state
    - any subset of it does not maintain the steady state
  - they offer a key insight into the objectives of the network
  - each elementary flux mode should have a clear biological interpretation in terms of the objectives of the network
  - determines whether a given set of enzymes/reactions are feasible at steady states
- Larger flux modes can be obtained by composing several flux modes: steady-state flux distributions

Calculating the elementary flux modes

- We are interested in combinations of reactions whose combined effect is to preserve the steady state
  - denote w_i the weight of reaction i in the flux mode
  - Recall: $\frac{dS}{dt} = Nv$, where N is the stoichiometric matrix and v is the vector of fluxes
  - We are interested in combinations of fluxes (w_1,…,w_r) that ensure $\frac{dS}{dt} = 0$
  - In other words, solve the equation Nw=0 in the unknown w
  - The solution is the kernel (or the null space) of matrix N

Example



- Stoichiometric matrix: $\begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
- NK=0 yields solution $K = (1 \quad 1 \quad 2 \quad 0)$
- In other words, in any steady state:
  - The rates of production and degradation of S_3 must be equal: $v_4=0$
  - $v_1+v_2+2v_3=0$

5. MODELS AND DATA

Sources of error in modeling

- Formulation errors
  - Result from errors in the model formulation
  - Significant variables were ignored
  - Interrelationships between variables were ignored or simplified
  - Relating the data to the model in the wrong way: see for example reporter systems
- Truncation errors
  - Come from the math techniques used in building the model
  - For example, an infinite series expansion may be truncated to a polynomial
- Round-off errors
  - Numerical errors coming from representing real numbers with finite precision
- Measurement errors
  - Imprecision in the collection of data
  - Physical limitations of the instruments
  - Human errors

Model fitting

- Problem: given the model and the data, is there a set of numerical values for all unknown kinetic parameters such that the numerical prediction of the model is ”close” to the data?

- Several components
  - Search for parameter values – an optimization / machine learning problem
  - Compare two sets of parameter values – introduce a suitable score function
  - Judge quality of the final model fit – introduce a measure of fit quality

Comparing two sets of parameter values

- Methods for judging the fitness of a model / comparing two sets of parameter values
  - Chebyshev criterion: minimize the largest absolute deviation
    - Intuition: more weight given to the worst point
  - Minimize the sum of absolute deviations
    - Intuition: tends to treat each data point equally and to average the deviations
  - Least-squares
    - Intuition: somewhat in-between
    - Widely used in practice
  - …

Fit quality

- Various methods for defining the a quantitative measure for the quality of a model fit
  - Here present just one, from Kuhnel et al, BMC Systems Biology (2008)
  - Only one data set at a time
  - Gives a measure of the average deviation of the model prediction from the experimental data, normalized by (the average of) the absolute values of the model prediction
  - This measure of fit quality does not discriminate against models aiming to explain experimental data with large absolute values
  - Let exp be the experimental data; m the number of experimental points

$$qual(\exp) = \frac{\sqrt{\frac{sum\_of\_squared\_deviations}{m}}}{mean\_of\_predicted\_values} \cdot 100\%$$

  - Rule of thumb (Kuhnel et al): lower than 20% value for qual(exp) can be considered as a good fit

6. HEAT SHOCK RESPONSE

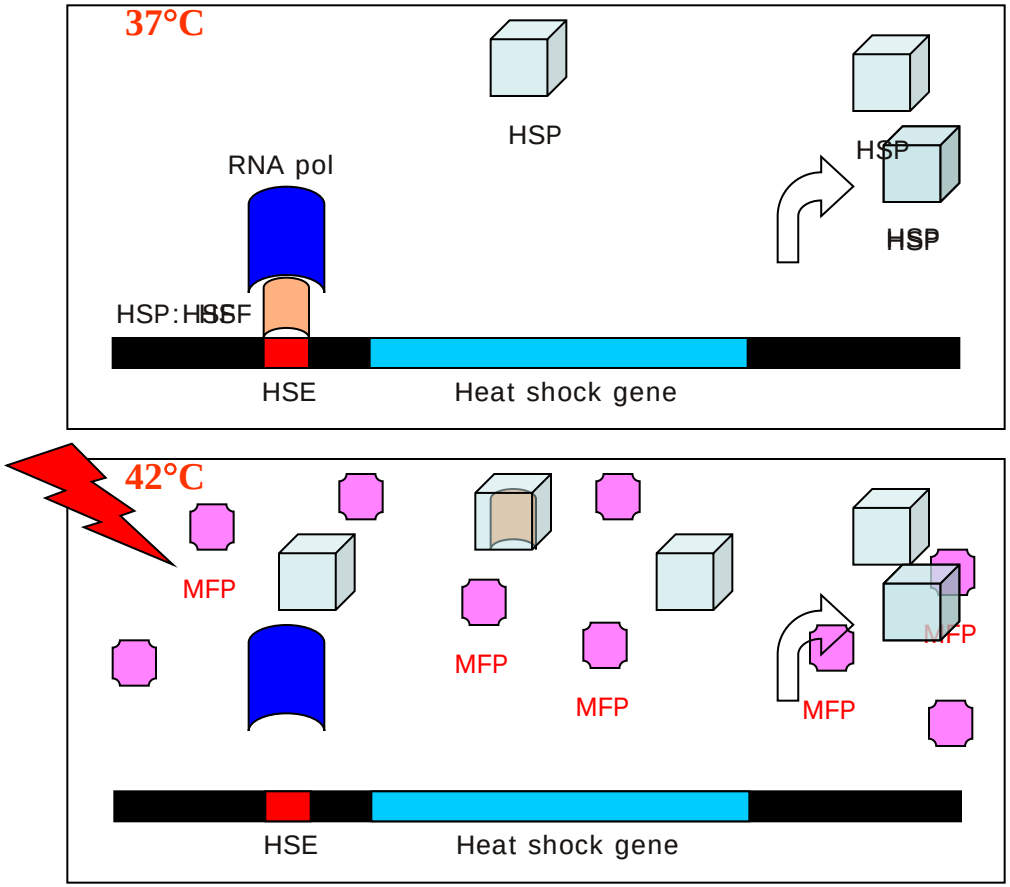
The modeling of the heat shock response

- Intense research on modeling the HSR in the last years
  - HSR is an ancient, very well-conserved regulatory network across all eukaryotes; bacteria have a similar mechanism
    - Good candidate for deciphering the engineering principles of regulatory networks
  - Heat shock proteins are very potent chaperones (sometimes called the "**master proteins**" of the cell)
    - Involved in a large number of regulatory processes
  - Tempting for a biomodeling, SysBio project because it involves relatively few main actors (at least in a first, simplified presentation)
- A number of models have been proposed
  - Some of them do not model the 3 components above
  - Some of them include modeling artifacts
- Discuss here a new, simple molecular model and its mathematical analysis
  - Standard, text-book-like molecular reactions only

Heat shock response: main actors

- Heat shock proteins (HSP)
  - Very potent chaperones
  - Main task: assist the refolding of misfolded proteins
  - Several types of them, we treat them all uniformly in our model with hsp70 as base denominator
- Heat shock elements (HSE)
  - Several copies found upstream of the HSP-encoding gene, used for the transactivation of the HSP-encoding genes
  - Treat uniformly all HSEs of all HSP-encoding genes
- Heat shock factors (HSF)
  - Proteins acting as transcription factors for the HSP-encoding gene
  - Trimerize, then bind to HSE to promote gene transcription
  - Treat uniformly all HSFs with HSF1 as base denominator
- Generic proteins
  - Consider them in two states: correctly folded and misfolded
  - Under elevated temperatures, proteins tend to misfold, exhibit their hydrophobic cores, form aggregates, lead eventually to cell death (see Alzheimer, vCJ, and other diseases)
- Various bonds between these species

A new molecular model for HSR



Our new molecular model

- **Transcription**

1. $HSF + HSF <-> HSF_2$
2. $HSF + HSF_2 <-> HSF_3$
3. $HSF_3 + HSE <-> HSF_3:HSE$
4. $HSF_3:HSE -> HSF_3:HSE + HSP$

- **Backregulation**

5. $HSP + HSF <-> HSP:HSF$
6. $HSP + HSF_2 -> HSP:HSF + HSF$
7. $HSP + HSF_3 -> HSP:HSF + 2HSF$
8. $HSP + HSF_3:HSE -> HSP:HSF + 2HSF + HSE$

- **Response to stress**

9. $PROT -> MFP$
10. $HSP + MFP <-> HSP:MFP$
11. $HSP:MFP -> HSP + PROT$

- **Protein degradation**

12. $HSP \rightarrow 0$

I. Petre et al. A simple mass-action model for the eukaryotic heat shock response, and its mathematical validation. Natural Computing (2011) 10:595-612

The mass-action ODE model

Table 1. A mathematical model for the heat shock response

$$d[\mathrm{hsf}]/dt = -2k_1^+[\mathrm{hsf}]^2 + 2k_1^-[\mathrm{hsf_2}] - k_2^+[\mathrm{hsf}][\mathrm{hsf_2}] + k_2^-[\mathrm{hsf_3}] - k_5^+[\mathrm{hsf}][\mathrm{hsp}] + k_5^-[\mathrm{hsp:hsf}] + k_6[\mathrm{hsf_2}][\mathrm{hsp}] \quad (1)$$
$$+ 2k_7[\mathrm{hsf_3}][\mathrm{hsp}] + 2k_8(\mathrm{hsf_3:hse})\,\mathrm{hsp}$$

$$d[\mathrm{hsf_2}]/dt = k_1^+[\mathrm{hsf}]^2 - k_1^-[\mathrm{hsf_2}] - k_2^+[\mathrm{hsf}][\mathrm{hsf_2}] + k_2^-[\mathrm{hsf_3}] - k_6[\mathrm{hsf_2}][\mathrm{hsp}] \quad (2)$$

$$d[\mathrm{hsf_3}]/dt = k_2^+[\mathrm{hsf}][\mathrm{hsf_2}] - k_2^-[\mathrm{hsf_3}] - k_3^+[\mathrm{hsf_3}][\mathrm{hse}] + k_3^-[\mathrm{hsf_3:hse}] - k_7[\mathrm{hsf_3}][\mathrm{hsp}] \quad (3)$$

$$d[\mathrm{hse}]/dt = -k_3^+[\mathrm{hsf_3}][\mathrm{hse}] + k_3^-[\mathrm{hsf_3:hse}] + k_8[\mathrm{hsf_3:hse}][\mathrm{hsp}] \quad (4)$$

$$d[\mathrm{hsf_3:hse}]/dt = k_3^+[\mathrm{hsf_3}][\mathrm{hse}] - k_3^-[\mathrm{hsf_3:hse}] - k_8[\mathrm{hsf_3:hse}][\mathrm{hsp}] \quad (5)$$

$$d[\mathrm{hsp}]/dt = k_4[\mathrm{hsf_3:hse}] - k_5^+[\mathrm{hsf}][\mathrm{hsp}] + k_5^-[\mathrm{hsp:hsf}] - k_6[\mathrm{hsf_2}][\mathrm{hsp}] - k_7[\mathrm{hsf_3}][\mathrm{hsp}] - k_8[\mathrm{hsf_3:hse}][\mathrm{hsp}] \quad (6)$$
$$- k_{11}^+[\mathrm{hsp}][\mathrm{mfp}] + (k_{11}^- + k_{12})[\mathrm{hsp:mfp}] - k_9[\mathrm{hsp}]$$

$$d[\mathrm{hsp:hsf}]/dt = k_5^+[\mathrm{hsf}][\mathrm{hsp}] - k_5^-[\mathrm{hsp:hsf}] + k_6[\mathrm{hsf_2}][\mathrm{hsp}] + k_7[\mathrm{hsf_3}][\mathrm{hsp}] + k_8[\mathrm{hsf_3:hse}][\mathrm{hsp}] \quad (7)$$

$$d[\mathrm{mfp}]/dt = \phi_T[\mathrm{prot}] - k_{11}^+[\mathrm{hsp}][\mathrm{mfp}] + k_{11}^-[\mathrm{hsp:mfp}] \quad (8)$$

$$d[\mathrm{hsp:mfp}]/dt = k_{11}^+[\mathrm{hsp}][\mathrm{mfp}] - (k_{11}^- + k_{12})[\mathrm{hsp:mfp}] \quad (9)$$

$$d[\mathrm{prot}]/dt = -\phi_T[\mathrm{prot}] + k_{12}[\mathrm{hsp:mfp}] \quad (10)$$

Modeling of the heat-induced misfolding

- Question: how do we model the heat-induced misfolding?
  - What is the temperature-dependant protein misfolding rate per second?
- Adapted from Pepper et al (1997), based on studies of Lepock (1989, 1992) on differential calorimetry

$$\varphi_T = (1-0.4/e^{T-37}) \times 1.4^{T-37} \times 1.45 \times 10^{-5}\ s^{-1}$$

- Formula valid for temperatures between 37 and 45, gives a generic protein misfolding rate per second

Table 1. A mathematical model for the heat shock response

$$d[\mathrm{hsf}]/dt = -2k_1^+[\mathrm{hsf}]^2 + 2k_1^-[\mathrm{hsf}_2] - k_2^+[\mathrm{hsf}][\mathrm{hsf}_2] + k_2^-[\mathrm{hsf}_3] - k_5^+[\mathrm{hsf}][\mathrm{hsp}] + k_5^-[\mathrm{hsp}:\mathrm{hsf}] + k_6[\mathrm{hsf}_2][\mathrm{hsp}] + 2k_7[\mathrm{hsf}_3][\mathrm{hsp}] + 2k_8(\mathrm{hsf}_3:\mathrm{hse})\,\mathrm{hsp}$$

$$d[\mathrm{hsf}_2]/dt = k_1^+[\mathrm{hsf}]^2 - k_1^-[\mathrm{hsf}_2] - k_2^+[\mathrm{hsf}][\mathrm{hsf}_2] + k_2^-[\mathrm{hsf}_3] - k_6[\mathrm{hsf}_2][\mathrm{hsp}]$$

$$d[\mathrm{hsf}_3]/dt = k_2^+[\mathrm{hsf}][\mathrm{hsf}_2] - k_2^-[\mathrm{hsf}_3] - k_3^+[\mathrm{hsf}_3][\mathrm{hse}] + k_3^-[\mathrm{hsf}_3:\mathrm{hse}] - k_7[\mathrm{hsf}_3][\mathrm{hsp}]$$

$$d[\mathrm{hse}]/dt = -k_3^+[\mathrm{hsf}_3][\mathrm{hse}] + k_3^-[\mathrm{hsf}_3:\mathrm{hse}] + k_8[\mathrm{hsf}_3:\mathrm{hse}][\mathrm{hsp}]$$

$$d[\mathrm{hsf}_3:\mathrm{hse}]/dt = k_3^+[\mathrm{hsf}_3][\mathrm{hse}] - k_3^-[\mathrm{hsf}_3:\mathrm{hse}] - k_8[\mathrm{hsf}_3:\mathrm{hse}][\mathrm{hsp}]$$

$$d[\mathrm{hsp}]/dt = k_4[\mathrm{hsf}_3:\mathrm{hse}] - k_5^+[\mathrm{hsf}][\mathrm{hsp}] + k_5^-[\mathrm{hsp}:\mathrm{hsf}] - k_6[\mathrm{hsf}_2][\mathrm{hsp}] - k_7[\mathrm{hsf}_3][\mathrm{hsp}] - k_8[\mathrm{hsf}_3:\mathrm{hse}][\mathrm{hsp}] - k_{11}^+[\mathrm{hsp}][\mathrm{mfp}] + (k_{11}^- + k_{12})[\mathrm{hsp}:\mathrm{mfp}] - k_9[\mathrm{hsp}]$$

$$d[\mathrm{hsp}:\mathrm{hsf}]/dt = k_5^+[\mathrm{hsf}][\mathrm{hsp}] - k_5^-[\mathrm{hsp}:\mathrm{hsf}] + k_6[\mathrm{hsf}_2][\mathrm{hsp}] + k_7[\mathrm{hsf}_3][\mathrm{hsp}] + k_8[\mathrm{hsf}_3:\mathrm{hse}][\mathrm{hsp}]$$

$$d[\mathrm{mfp}]/dt = \phi_T[\mathrm{prot}] - k_{11}^+[\mathrm{hsp}][\mathrm{mfp}] + k_{11}^-[\mathrm{hsp}:\mathrm{mfp}]$$

$$d[\mathrm{hsp}:\mathrm{mfp}]/dt = k_{11}^+[\mathrm{hsp}][\mathrm{mfp}] - (k_{11}^- + k_{12})[\mathrm{hsp}:\mathrm{mfp}]$$

$$d[\mathrm{prot}]/dt = -\phi_T[\mathrm{prot}] + k_{12}[\mathrm{hsp}:\mathrm{mfp}]$$

Parameter estimation

- Data readily available for the goal: Kline, Morimoto (1997) – heat shock of HeLa cells at 42C for up to 4 hours, data on DNA binding (HSF_3:HSE)

- Requirements for the model:
  - 17 independent parameters, 10 initial values to estimate
  - 3 conservation relations available
  - The model must be in steady state at 37C, which gives 7 more algebraic equations (each of them quadratic)
  - Altogether: 17 independent values

  - Other conditions: total HSF somewhat low, refolding a fast reaction, HSPs long-lived proteins

The modeling/simulation environment

- Our choice: **COPASI** (www.copasi.org)
  - Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006). COPASI — a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067-74.
  - User-friendly
  - Stochastic and deterministic time course simulation
  - Steady state analysis
  - Metabolic control analysis
  - Mass conservation analysis
  - Optimization of arbitrary objective functions
  - SBML-based
  - **Excellent for parameter estimation**
  - FREE!

Parameter estimation

- Standard estimation procedure in COPASI (and not only)
  - Give the data and the target function
  - Give the list of parameters
  - The software scans the range of parameters and makes choices; for each choice it evaluates the target function against the experimental data (least mean squares)
    - The way it scans the space of parameter values depends on the chosen method
    - Many sophisticated methods currently available
    - All are local-optimization methods
  - It reports the best set of values

- Estimation repeated over and over again, with various methods for scanning the parameter space, to improve on the score of the fit

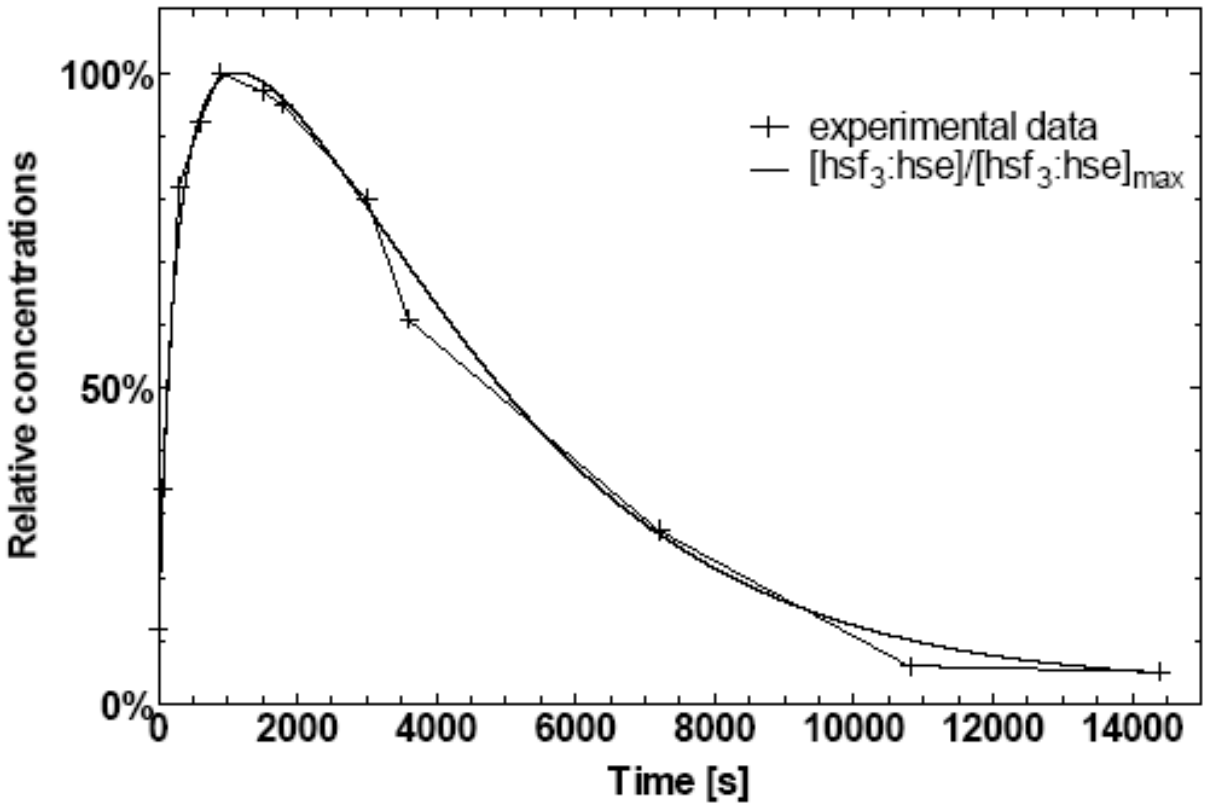
Parameter estimation

- Model fit is anecdotically easy: *"with a few free parameters, an elephant can always be fit"*!
  - Seems to come from a well known fact that for any given n points in the bi-dimensional space, a polynomial of suitable degree may be found to go through those points
  - In practice, the polynomial cannot be chosen freely
- Our problem:
  - Find suitable parameter values and suitable initial values for all variables so that the numerical prediction for [HSF3:HSE] is close to the experimental data of Kline-Morimoto (1997)
    - **Outcome:** sure enough, "relatively easy" to find!
  - **Additional requirement:** the model must be in steady state at 37C
    - This is a condition on the initial numerical values of the model
    - **Difficulty:** the values found as a good fit at 42C may not satisfy the steady state condition!
    - **Difficulty:** to give this condition as a constrain to the model fit, one has to solve analytically an algebraic system of large degree: impossible!

Parameter estimation

- **Solution:** rather than solving the algebraic system, we look for *an approximation of its solution*: translate this condition into a more extensive model fit
  - **Problem:** After obtaining the fit, the model is still not in the steady state!
  - **Solution:** replace the estimated initial values with (the numerical estimations of) the steady state at 37C. Then the resulting system remains in the steady state at 37C
  - **Problem:** The numerical fit (in absolute values) at 42C is ruined
  - **Solution:** recall that the Kline-Morimoto data is relative! In relative terms, the fit is excellent!

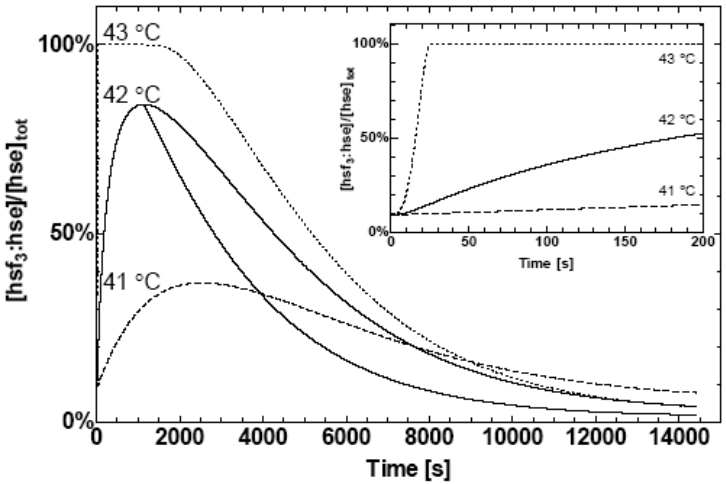
Parameter fit



I. Petre et al. A simple mass-action model for the eukaryotic heat shock response, and its mathematical validation. Natural Computing (2011) 10:595-612

Predictions and validation

1. HSF dimers are only a transient state between monomers and trimers
   - The model however does not ignore them because of kinetic considerations
   - Numerical simulations predict low levels of HSF dimers
2. Higher the temperature, higher the response
3. Prolonged transcription at 43C confirmed
   - Unlike previous models
4. Heat shock removed at the peak of the response confirms a more rapid attenuation phase

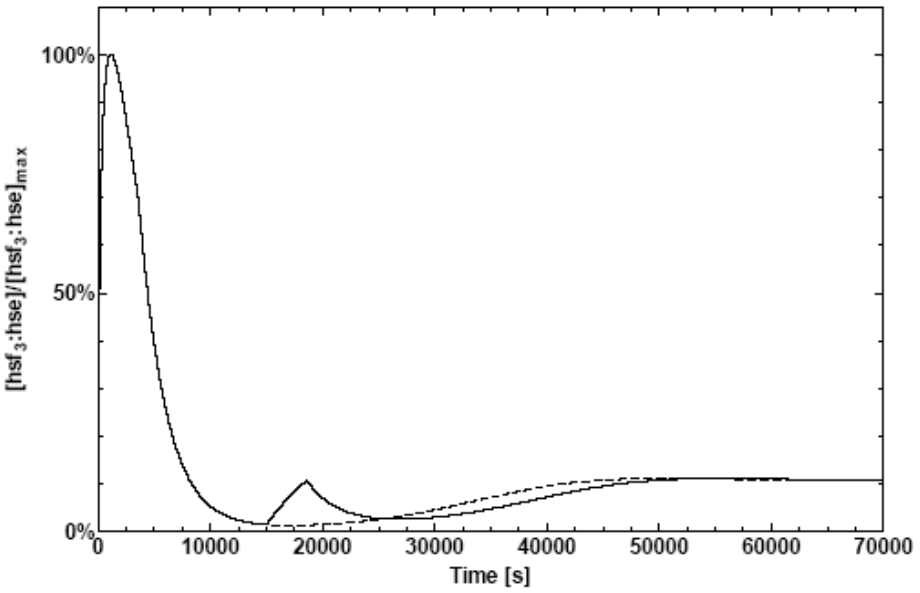


All data is in relative terms with respect to the highest value in the graph so that it can be easily compared

Predictions and validation

- Experiment: two waves of heat shock, the second applied after the level of HSP has peaked
- Observation: the second heat shock response much milder than the first
  - The reason is that the cell is better prepared to deal with the second heat shock
  - Therapeutic consequences have been suggested: “train” the cell for heat shock by an initial milder heat shock

- The model prediction is in line with the experimental observation
  - Dotted line: heat shock at 42C for two hours, behavior followed up to 20 hours
  - Continuous line: heat shock at 42C for two hours, followed by a second wave of heat shock after the level of HSP has peaked



Model identifiability

- Problem: is there a unique set of parameter values that gives a "good" fit to the experimental data and validates all the additional tests?
- Re-run the parameter estimation procedure
  - use different initial values
  - use different (types of) machine learning methods
- Results
  - We obtained 10 more sets of parameter values that fit the experimental data of Kline-Morimoto and keep the model in steady state at 37C
  - All sets failed the model validation tests

Model identifiability – systematic sampling of the parameter space

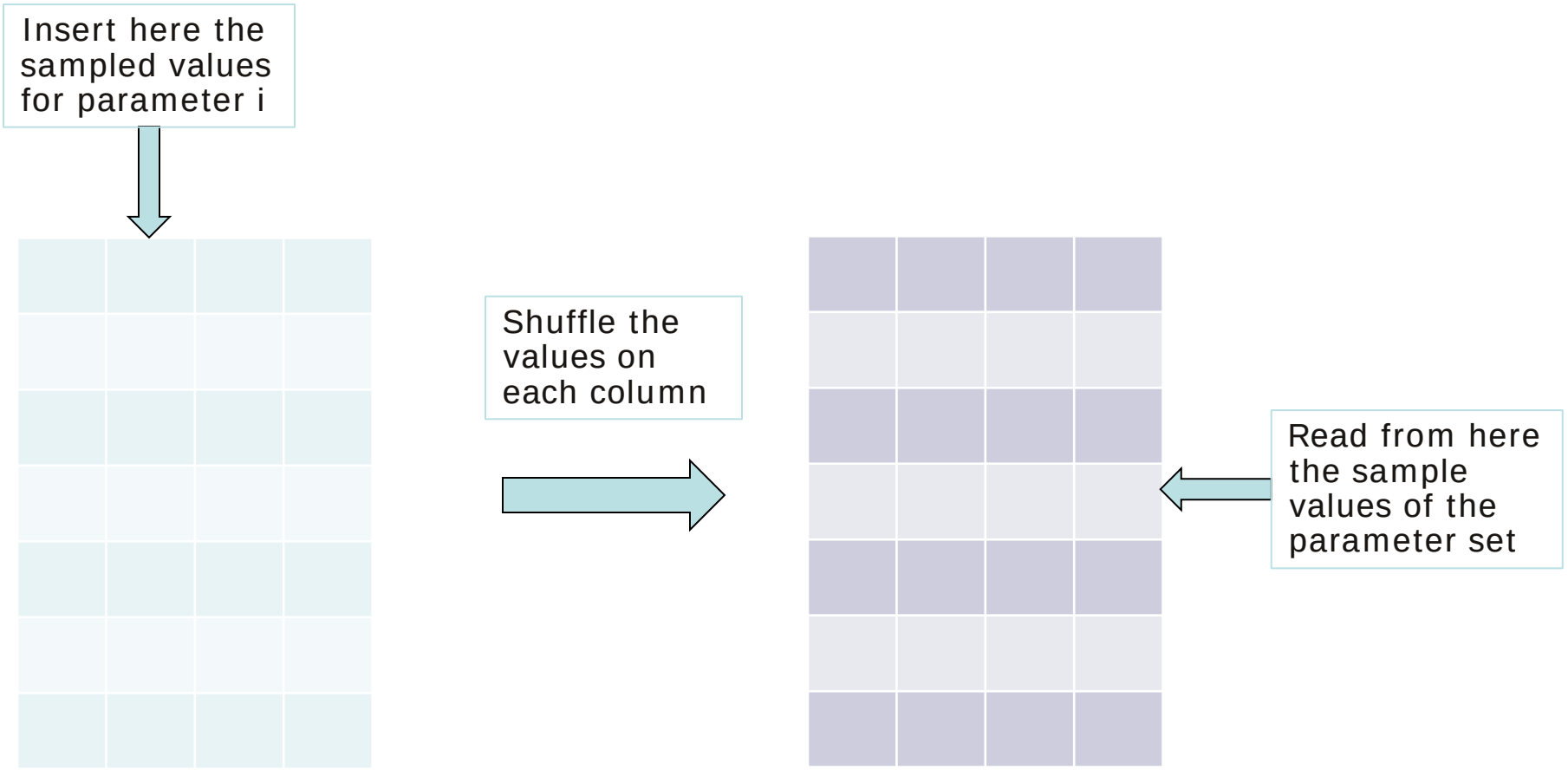
- Different approach: systematic sampling of the parameter space
  - partition the domain of each parameter into a large number of subintervals (say 100.000); sample values for that parameter from each subinterval
  - check the behavior of the model for all combinations of parameter values to get a sampling of the model behavior throughout the multi-dimensional parameter space
  - Major problem: combinatorial explosion of the number of model variants

Model identifiability – Latin Hypercube sampling

- Problem: huge number of samples to consider – $(10^5)^{17} = 10^{85}$

- Fast, practical solution: the Latin Hypercube Sampling method (McKay, 1979)
  - it provides samples which are uniformly distributed over each parameter
  - the number of samples is independent of the number of parameters
  - choose the size N of the sample; let p be the number of parameters
  - divide the domain of each parameter into N subintervals; randomly select N numerical values for each parameter i, one from each of its subintervals; place the values on column i of a matrix Nxp

Latin Hypercube sampling



Model identifiability – Latin Hypercube sampling

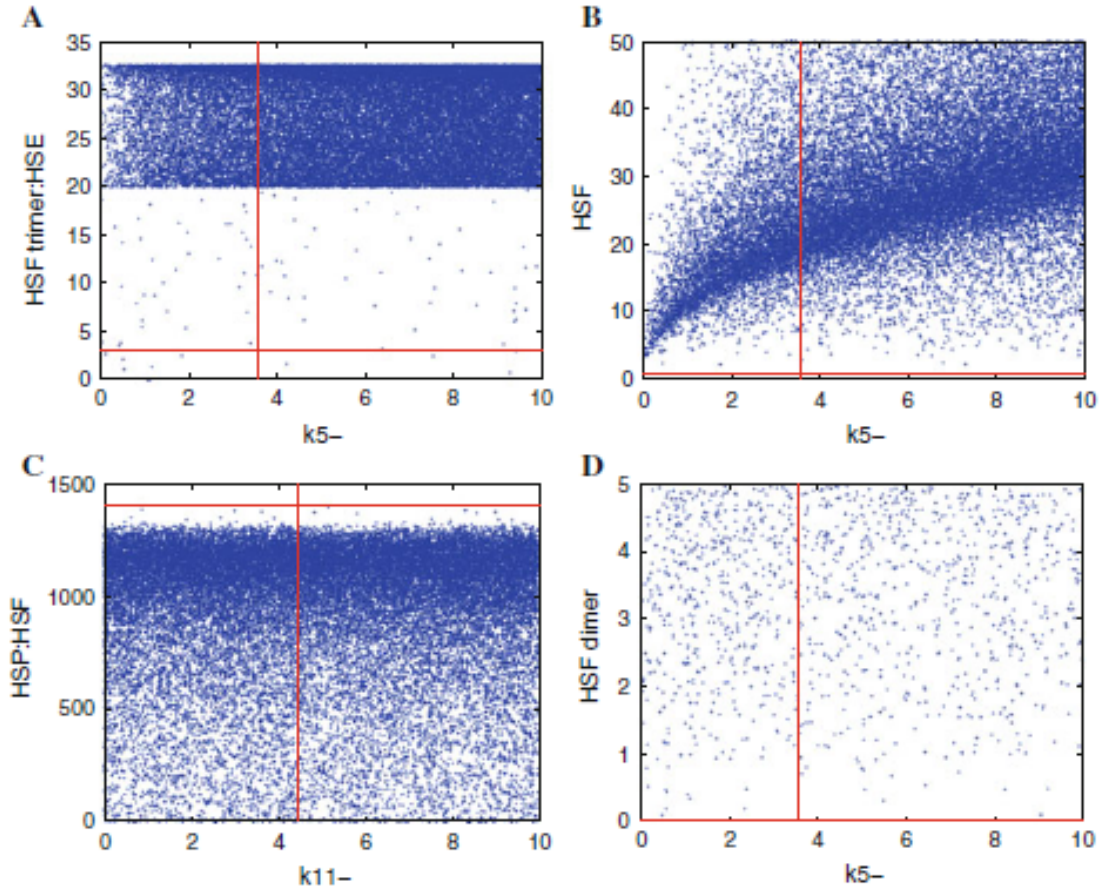


Fig. 3 Scatter plots of the steady state values at 37°C of the sampled models (*blue crosses*) and the basic model (*red horizontal line*). The *red vertical line* indicates the parameter value of the basic model. The plots of hsf (**b**) and hsf_2 (**d**) are zoomed in, hence not all points are present, i.e., the values of the remaining steady states were higher than the maximum value on the y-axes

I. Petre et al. A simple mass-action model for the eukaryotic heat shock response, and its mathematical validation. Natural Computing (2011) 10:595-612

Model identifiability – Latin Hypercube sampling

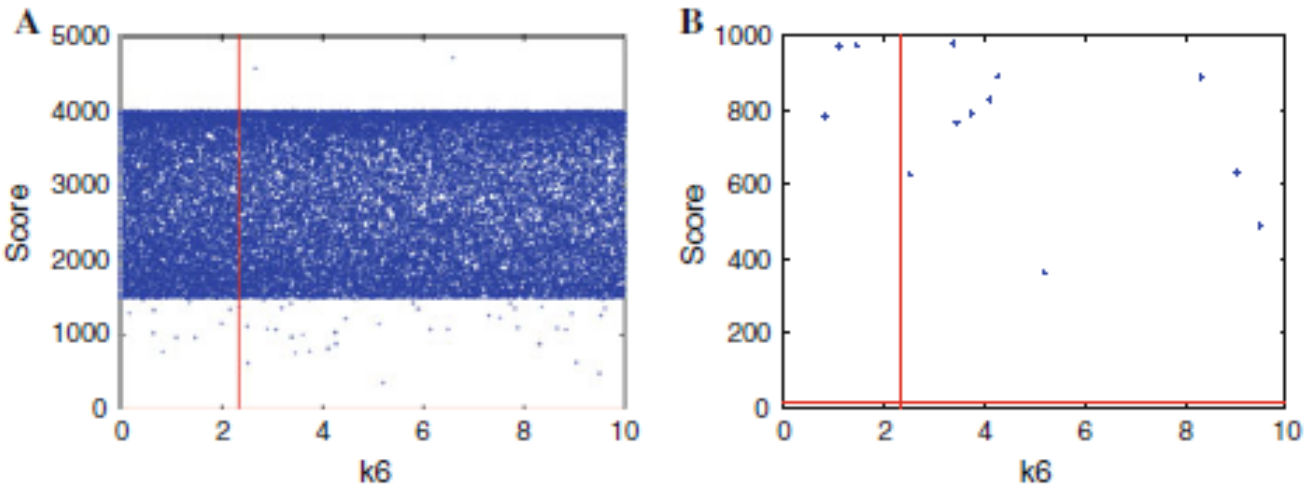


Fig. 4 Scatter plot (**a**) and its zoomed in version (**b**) of the score measuring the fit of the sampled models (*blue crosses*) and the basic model (*red horizontal line*) with respect to the experimental data. The *red vertical line* indicates the parameter value of the basic model

I. Petre et al. A simple mass-action model for the eukaryotic heat shock response, and its mathematical validation. Natural Computing (2011) 10:595-612

Model identifiability

- Conclusion
  - likely that a model of this size is not uniquely identifiable
  - finding an optimal (or at least a “good”) model setup is very difficult

7. DISCUSSION

Biomodeling with differential equations: some physical difficulties

- Assumes that the time evolution of a chemically reacting system is both continuous and deterministic
- Difficulties with this assumption
  - the time evolution is **NOT** continuous: molecular population levels increase and discrete only with discrete amounts
  - the time evolution is **NOT** deterministic (even when ignoring the quantum effects and assuming classical mechanics for the molecular kinetics)
    - it is only deterministic in the full position-momentum phase space (knowing the positions and velocities of all molecules)
    - it is not deterministic in the N-dimensional space of the species population numbers
- However:
  - in many cases the time evolution of a chemical system can be treated as continuous and deterministic
  - the difficulties come when some species populations are small, or in conditions of chemical instability
  - **Solution in these cases: stochastic models!**

Deterministic and stochastic modeling

- *Deterministic model*
  - *Given the current state of the system, all future behavior of the system is uniquely defined*
  - *Usually the model reflects the **average** behavior of the observed system*
  - *Typical methods used: differential or difference equations*
  - *Typical:*
    - *Concentrations of molecules are modeled*
    - *Reactions are taking place diffusion-like (gradient-like)*
    - *Differential equations*
- Stochastic model
  - Given the current state of the system, many possible future behavior are possible
  - Probability distributions dictate the behavior of the system
  - Well-suited to model **individual**, rather than average behavior
  - Typical
    - Number of molecules are modeled
    - Reactions are taking place following “collisions” among the reactants
    - Markov processes

Deterministic and stochastic modeling

- *ODE modeling*
  - *The objects*
    - *the concentrations of all species of interest*
    - *the rates of all reactions*
  - *Main assumptions*
    - *The system is well-stirred*
    - *The system is at thermodynamical equilibrium*
  - *Methods*
    - *Those of mathematical analysis (continuous mathematics)*

- Stochastic modeling
  - The objects
    - **the number of copies of all species of interest**
    - the rates of all reactions
  - Main assumptions
    - The system is well-stirred
    - The system is at thermodynamical equilibrium
  - Methods
    - Those of probability theory

Deterministic and stochastic modeling

- *ODE modeling*
  - *The reaction rate gives the amount with which the concentration of every metabolite involved in the reaction changes per unit of time*
    - *For a consumed metabolite, the change will be –v(t)*
    - *For a produced metabolite, the change will be v(t)*

- Stochastic model
  - It is the description of a continuous time, discrete state Markov process
  - *Grand probability function*: $P(X_1,X_2,\ldots,X_n,t)$ is the probability that at time t there are X_1 molecules of species S_1, …, X_n molecules of species S_n
  - The *grand probability function* may be obtained through a differential equation: the *chemical master equation*
    - Reason what is the probability of being in a certain state after one step

Deterministic and stochastic modeling

- Deterministic approach
  1. based on the concept of diffusion-like reactions
  2. the time evolution of the system is a continuous, entirely predictable process
  3. governed by a set of ODEs
  4. The system of ODEs is often impossible to solve
  5. it models the average behavior of the system
  6. assumes that the system is well-stirred and at thermodynamical equilibrium
  7. conceptual difficulties when small populations are involved
  8. numerical simulations are straightforward and fast
  9. impossible to reason about individual runs rather than the average

- Stochastic approach
  1. based on the concept of reactive molecular collisions
  2. the time evolution of the system is a random-walk process through the possible states
  3. governed by a single differential equation: the chemical master equation
  4. the CME is often impossible to solve
  5. it models individual runs of the system
  6. assumes that the system is well-stirred and at thermodynamical equilibrium
  7. no difficulties with small populations
  8. numerical simulations via Gillespie's SSA are slow
  9. only gives individual runs; estimate the average through many runs

Contributors

- Collaborators on the heat shock response
  - Computer Science and Math
    - Eugen Czeizler (Helsinki)
    - Vladimir Rogojin (Helsinki)
    - Elena Czeizler (Turku)
    - Andrzej Mizera (Turku)
    - Bogdan Iancu (Turku)
    - Diana-Elena Gratie (Turku)
    - Ralph Back (Turku)
  - Biology and biochemistry
    - John Eriksson (Turku)
    - Lea Sistonen (Turku)
    - Richard Morimoto (Chicago)
    - Andrey Mikhailov (Turku)
    - Claire Hyder (Turku)

- Review paper together with
  - Diana-Elena Gratie (Turku, Finland)
  - Bogdan Iancu (Turku, Finland)

- Work funded by
  - Academy of Finland
  - Turku Centre for Computer Science
  - Centre for International Mobility Finland
  - Foundation of Åbo Akademi University