Computational modeling techniques

Ion Petre  
Department of IT, Abo Akademi  
http://www.users.abo.fi/ipetre/compmod/
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

Check also http://www.simoncamilleri.com/the_truth_of_the_elephant/
Modeling

• What is a model?
  o A (partial) view of the reality
  o An abstraction of the reality
  o A representation of the (supposedly) main features of the reality, including the connections among them
  o For a given object of study, many models may be given, possibly focusing on different features of the object

• What a model is not
  o A model is not the reality
  o A model is not certain!

• Many types of models exist!

“All models are wrong, some are useful”

Huge momentum for modelling right now

• The Human Brain Project (2013-2023) is a Future and Emerging Technologies (FET) Flagship project of the European Commission
  o Aim: build and simulate complete model of the human brain to better understand its functions
  o Total budget: 1.2 billion euros

• The BRAIN Activity Map Project (2013-2023) is a US initiative
  o Aim: map the activity of every neuron in the human brain
  o Seen as the next high-impact project after the human genome project
  o Total budget: at least 3 billion dollars

• Climate change models

• Economic forecast models

• Weather models
“The sciences do not try to explain, they hardly even try to interpret, they mainly make models.

By a model is meant a mathematical construct which, with the addition of certain verbal interpretations describes observed phenomena.

The justification of such a mathematical construct is solely and precisely that it is expected to work.”

John von Neumann (1903-1957)
“Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry.

If mathematical analysis should ever hold a prominent place in chemistry - an aberration which is happily almost impossible - it would occasion a rapid and widespread degeneration of that science.”

Auguste Comte (full name: Isidore Marie Auguste Francois Xavier Comte; January 17, 1798 - September 5, 1857): Philosophie Positive, 1830
"The more progress the physical sciences make, the more they tend to enter the domain of mathematics, which is a kind of center to which they all converge. We may even judge the degree of perfection to which a science has arrived by the facility to which it may be submitted to calculation"

Adolphe Quetelet (1828)
Models: some examples

- Models of a building
  - The foundation plan
  - The water pipes plan
  - The electricity plan
  - The ventilation plan
  - The room division plan
  - Division of people into rooms

- Maps
  - Geographic map
  - Political map
  - Road map
  - ...

- Models of morality
- Family models
- Role models

- Models of society
- Political models
- Election models
- Political representation models

- Weather models
- Traffic models

- Infection models

- Company development models
- ...

September 1, 2014
A different type of model

• Model of a computer
  o Why do we need one?
    - Can we not just use a real computer?
    - The model(s) came first!
    - What can you do with a model that you cannot do with a real computer?
      ▪ **Reason** about computing!
  o Is there only one model? What is the difference?
    - Turing machine
      ▪ The factory worker model
    - Register machines (e.g., counter machines)
      ▪ The arithmetic model
    - Biocomputing
      ▪ Based on/inspired by molecular biology
    - Quantum computing
      ▪ Based on/inspired by quantum physics
    - ...
  o Classes of results
    - Computability
    - Complexity (efficiency)
    - Note: independence from the actual implementation of the computation (i.e., independent of the computer)
Another type of model: Fermi approximations

- Enrico Fermi: rough quantitative estimates from very little data
- Examples
  - How many piano tuners are in Chicago?
    - About 5 million people in Chicago
    - In average about 2 persons in each household
    - About one household in 20 has a piano that is tuned regularly
    - Pianos are tuned regularly are tuned about once per year
    - A piano tuner needs two hours to tune a piano (incl. driving)
    - A piano tuner works 8 hours per day, 5 days a week, 50 weeks per year
  - Result:
    - \((5.000.000/2)/20)*1=125.000\) piano tunings per year
    - \((50*5*8)/2=1000\) piano tunings per year for one tuner
    - \(125.000/1000=125\) piano tuners in Chicago
Approximations

- 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; numerical integration
  - Drawback: huge effort; collected every 10 year by the US Census Bureau
  - Approximation
    - US population: 300 million in 2008
    - Assume a life expectancy of 80 (a model where everybody still alive at 80 dies abruptly on their 80th birthday)
    - Lump the curve into a rectangle: width of 80, height to be calculated
Approximations (continued)

- Height of the rectangle:
  - Total population of US: 300 million (2008)
  - Easier to divide to 75 than to 80
    - The inaccuracy is not larger than the error made by lumping; might even cancel the lumping error
  - Height: \( \frac{300,000,000}{75} = 4,000,000 \)
- 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 !!
Models: other examples

• **Computer network architecture**
  o Design chosen so that the server services are available at all times (with large probability)
    - Depending on the probability of failure of a computer, or of the network links, a certain design is chosen
    - Different reliability/failure models will lead to different solutions
  o Change the reliability model and the architecture will stop being robust
    - How about somebody shutting off the electricity in the whole building

• **Cryptography**
  o A cryptographic system should be provably secure
    - Prove mathematically that an attacker would not be able to break the system in “reasonable” time
    - What computing power should one assume for the attacker?
    - What knowledge of the system should one assume for the attacker?
    - What level of access to the cryptographic system?

• **Casino gambling**
  o The games/machines are designed in such a way that in average the casino has (slightly) better chances of winning
    - Some gamblers will win, others will lose
  o Change the rules and the chances of winning may change
    - Casinos are very carefully monitoring players with a gambling system
Life inside a cell

- **Simplifications often made by biomodelers**
  - Cell is “like a bag of chemicals floating in water”
  - Metabolites flow around chaotically
  - Metabolites are uniformly distributed
  - Proteins are just like balls (or cubes), DNA is just like a rope
  - 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
  - Highly specific metabolites
  - Very precise recognition of one’s target
  - Energy efficiency optimized
  - Exquisite regulation, 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, ...
We focus in this course on mathematical and computational models

- As we saw, many other types of models exist
- “Model” is indeed a very overloaded word
- In this way, we define a model as a mathematical representation of the reality
- Models that mimic the reality by using the language of mathematics

**Goal of the course**

- An introduction to the process of mathematical modeling
- Give a number of techniques used for:
  - Building a model
  - Analyzing a model
  - Using a model
  - Simulating a model
- Not a course in mathematics, rather in the use of some mathematical techniques for the purpose of modeling
  - How to use various tools
  - Little on the math properties of the tools
Mathematical modeling

• What is a mathematical model
  o Possible definition: An abstract, simplified, mathematical construct related to a part of reality and created for a particular purpose

• Why mathematical modeling?
  o It allows for a precise formulation of the chosen aspects of the reality
    – Must formulate ideas precisely; less implicit assumptions
  o It allows for a precise formulation of the current knowledge of the reality
    – Models as knowledge representations
  o It allows for precise reasoning about the reality
    – Based on the concise language, models can be analyzed in a formal way
  o It allows for some types of analysis that would be impossible to perform on the reality
    – Large body of math tools available
    – Model checking: verify all possible behaviors of the model in time
    – Scenario analysis: verify the behavior of the model in some well-defined extreme scenarios (e.g., disaster scenarios)
  o It allows for predictions
    – Based on computational simulations
Mathematical models

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

- **Formulate the problem**
  - What do you want to know/understand?

- **Outline the model**
  - Divide the world into the three categories in the previous slide
  - Specify the interrelations among the variables
  - A challenging stage; not always clear what is important and what is not

- **How useful is the resulting model?**
  - Can you obtain the needed data for the model?
  - Can the available data be used in the model?
  - If not, reformulate the model and perhaps even the problem
  - Note: sometimes a model needs no data; all included in the assumptions

- **Test the model**
  - Use the model to make predictions that can be checked against known data or common-sense
  - Note: sometimes difficult/expensive/impossible to test a model

- **Use the model**
  - Careful that the situations where it is used were captured in the modeling process
  - Not very different than those where the model was tested
  - Every application of the model is a test for the model
Modeling

Start here

- Real-world data
  - Simplification
  - Verification
  - Predictions / explanations

Model
  - Analysis
  - Mathematical conclusions
  - Interpretation
An example

• **Goal:** build a model for the long term growth of a population
  
  o Choose as independent (external) variables:
    - Birth rate per individual: \( b \)
    - Death rate per individual: \( d \)
    - Population size at \( t=0 \): \( P(0) \)

  o Write the model: the change in the population per unit of time is the gain through birth minus the loss through death:
    \[
    \frac{dP}{dt}(t) = bP(t) - dP(t)
    \]

  o Solve the model:
    \[
    P(t) = P(0)e^{(b-d)t}
    \]

  o Test the model:
    - If \( b<d \): the population eventually dies out: \( \lim_{t \to \infty} P(t) = 0 \)
    - If \( b>d \): the population explodes: \( \lim_{t \to \infty} P(t) = \infty \)
    - If \( b=d \): the population remains constant
    - **Note:** good to test the model in limit cases

  o **Conclusion**
    - The model is fragile: unreasonable conclusion unless it predicts no change in the population
    - May be good for short-term predictions
An example (continued)

- Error in the model formulation:
  - the growth rate in a population is not a constant
  - it should depend on the size of population: e.g., exhaustion of food supply
- Reformulate the model:
  - The growth rate \( r=b-d \) is now a (unknown) function of \( P \): \( r(P) \)
  - \( \frac{dP}{dt}=r(P)P \)
- Model is less useful (incomplete) because the function \( r(P) \) may be hard (impossible) to deduce
  - Obtain rough estimates of \( r(P) \)
  - For example: assume that \( r(P) \) gets to 0 as \( P(t) \) gets close to a maximum value \( M \)
  - For example take \( r(P)=s(M-P(t)) \), for some constants \( s, M \)
- Model analysis
  - The model has \( M \) as a steady state
- Conclusion:
  - Reaching a steady state was built in the choice of \( r(P) \)
  - Weak point: the model shows no fluctuations around the steady state
An example (continued)

• Reformulate the model
  o Introduce a time delay:
    – The death rate \( d \) is the same regardless of the age
    – The birth rate changes in time from 0 to a constant that is reached at age \( p \)
  o Reformulate the model for the interval \([0, p]\)
    – \( \frac{dP}{dt} = -dP(t) + bP(t-p) \)

• Other objections to the model
  o Refine the death and the birth rate depending on the age and the sex division of the population
    – Divide the population into sex and age groups
    – Model the interrelations among these groups
  o Model the seasonality of the birth rate
    – Short term models: consider the details
    – Long terms models: eliminate the problem by averaging over an entire year
• **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
Modeling in science and engineering

- Great traditions of mathematical modeling in science and engineering
  - Physics
  - Chemistry
  - Chemical engineering
  - Computer science
  - Biology
Modeling approaches

• Mathematical models
  o Continuous vs. discrete mathematics
  o Deterministic vs. stochastic mathematics

• Only somewhat elementary techniques to be discussed in this course
  o list on the next slide
  o only elementary math concepts, techniques
  o wide applicability

• A second course on modeling to be offered in period 2
  o based on concepts, techniques from computer science
  o more advanced
  o more focus on software tools, computational simulations
  o fresh material, developed in the last 10-15 years
  o more specialized, narrower applicability
Course content

- Modeling change
- Modeling proportionality, geometric similarity
- Model fitting
- Data-driven modeling
- Simulation-based modeling

- Discrete probabilistic modeling
- Discrete optimization modeling
- Dimensional analysis
- Modeling with ODEs
- Continuous optimization modeling

Course schedule

• Lectures
  o Mondays 13-15, room Catbert (B3028)
  o Wednesdays 10-12, room Catbert (B3028)
  o Starts: September 3
  o Ends: October 26

• Exercises
  o Mondays 15-17, room Cobol (B3040)

• Final exam: in October-November (to be fixed)
• Course webpage: http://www.users.abo.fi/ipetre/compmod/
• Lecturer: Ion Petre, ipetre@abo.fi
• Teaching assistant: Cristi Gratie, cgratie@abo.fi