1

ENGINEERING SYSTEM MODELLING AND SIMULATION

Mustafa Kemal SEVİNDİR, Ph.D

sevindir@yildiz.edu.tr

A Complex World Needs Models

Engineers and Scientists Try To Understand, Develop, or Optimize “Systems”.

“System” refers to the object of interest, which can be a part of nature (such as a plant cell, an atom, a galaxy etc.) or an artificial technological system.

For example, consider the problem of a table which is unstable due to an uneven floor.

This is a technical system and everybody knows what must be done to solve the problem: we just have to put suitable pieces of cardboard under the table legs.

Each of us solves an abundant number of problems relating to simple technological systems of this kind during our lifetime.

Beyond this, there is a great number of really difficult technical problems that can only be solved by engineers.

Characteristic of these more demanding problems is a high complexity of the technical system. We would simply need no engineers if we did not have to deal with complex technical systems such as computer processors, engines, and so on.

Similarly, we would not need scientists if processes such as the photosynthesis of plants could be understood as simply as an unstable table.

The reason why we have scientists and engineers, virtually their right to exist, is the complexity of nature and the complexity of technological systems.

(The complexity challenge) It is the genuine task of scientists and engineers to deal with complex systems, and to be effective in their work, they most notably need specific methods to deal with complexity.

The general strategy used by engineers or scientists to break up the complexity of their systems is the same strategy that we all use in our everyday life when we are dealing with complex systems: simplification.

The idea is just this:

if something is complex, make it simpler.

Consider an everyday life problem related to a complex system: A car that refuses to start.

In this situation, everyone knows that a look at the battery and fuel levels will solve the problem in most cases.
Everyone will do this automatically, but to understand the problem solving strategy behind this, let us think of an alternative scenario.

Assume someone is in this situation for the first time. Assume that “someone” was told how to drive a car, that he has used the car for some time, and now he is for the first time in a situation in which the car does not start.

Of course, we also assume that there is no help for miles around!

Then, looking under the hood for the first time, our “someone” will realize that the car, which seems simple as long as it works well, is quite a complex system.

Fig. 1.1 Car as a real system and as a model.

We know that this simplified picture is appropriate in this given situation, and it guides us to look at the battery and fuel levels and then to solve the problem within a short time.

This is exactly the strategy used by engineers or scientists when they deal with complex systems.

Anyone who wants to understand complex systems or solve problems related to complex systems needs to apply appropriate simplified descriptions of the system under consideration. This means that anyone who is concerned with complex systems needs models, since simplified descriptions of a system are models of that system by definition.

(Role of models) To break up the complexity of a system under consideration, engineers and scientists use simplified descriptions of that system (i.e. models).

Systems, Models, Simulations

In 1965, Minsky gave the following general definition of a model.

(Model) To an observer B, an object A^* is a model of an object A to the extent that B can use A^* to answer questions that interest him about A.

Systems, Models, Simulations

Teleological Nature of Modeling and Simulation

An important aspect of the above definition is the fact that it includes the purpose of a model, namely, that the model helps us to answer questions and to solve problems.

This is important because particularly beginners in the field of modeling tend to believe that a good model is one that mimics the part of reality that it pertains to as closely as possible.

(The best model) The best model is the simplest model that still serves its purpose, that is, which is still complex enough to help us understand a system and to solve problems. Seen in terms of a simple model, the complexity of a complex system will no longer obstruct our view, and we will virtually be able to look through the complexity of the system at the heart of things.

The entire procedure of modeling and simulation is governed by its purpose of problem solving – otherwise it would be a mere l'art pour l'art.

Modeling and simulation is always goal-driven, that is, we should know the purpose of our potential model before we sit down to create it.

It is hence natural to define fundamental concepts such as the term *model* with a special emphasis on the purpose-oriented or *teleological nature of modeling and simulation*.

Systems, Models, Simulations

Modeling and Simulation Scheme

Conceptually, the investigation of complex systems using models can be divided into the following steps:

(Modeling and simulation scheme)

Definition

- Definition of a problem that is to be solved or of a question that is to be answered.
- Definition of a system, that is, a part of reality that pertains to this problem or question.

Systems Analysis

- Identification of parts of the system that are relevant for the problem or question.

Modeling

- Development of a model of the system based on the results of the systems analysis step.

Simulation

- Application of the model to the problem or question.
- Derivation of a strategy to solve the problem or answer the question.

Validation

- Does the strategy derived in the simulation step solve the problem or answer the question for the real system?

In the *car example*

Definition

The problem is that the car does not start and the car itself is the system.

Systems Analysis

The battery and fuels levels as the relevant parts of the system.

Modeling

A model consisting of a battery and a tank such as in Figure 1.1 is developed.

Fig. 1.1 Car as a real system and as a model.

Simulation

The application of this model to the given problem in the “simulation” step of the scheme then leads to the strategy “check battery and fuel level”.

Validation

This strategy can then be applied to the real car in the “validation” step.

If it works, that is, if the car really starts after refilling its battery or tank, we say that the model is *valid or validated.*

In a real modeling and simulation project, the *systems analysis step* of the above scheme can be a very time-consuming step. It will usually involve a thorough evaluation of the literature. In many cases, the literature evaluation will show that similar investigations have been performed in the past, and one should of course try to profit from the experiences made by others that are described in the literature.

Beyond this, the system analysis step usually involves a lot of discussions and meetings that bring together people from different disciplines who can answer your questions regarding the system. These discussion will usually show that new data are needed for a better understanding of the system and for the validation of the models in the validation step of the above scheme. Hence, the definition of an experimental program is also another typical part of the systems analysis step.

The *modeling step* will also involve the identification of appropriate software that can solve the equations of the mathematical model.

In the *validation step*, the model results will be compared with experimental data. These data may come from the literature, or from experiments that have been specifically designed to validate the model.

Usually, a model is required to fit the data not only quantitatively, but also qualitatively in the sense that it reproduces the general shape of the data as closely as possible. But, of course, even a model that perfectly fits the data quantitatively and qualitatively may fail the validation step of the above scheme if it cannot be used to solve the problem that is to be solved, which is the most important criterion for a successful validation.

(Start with simple models!)

To find the best model, start with the simplest possible model and then generate a sequence of increasingly complex model formulations until the last model in the sequence passes the validation step.

Systems, Models, Simulations

Simulation

(Simulation)

Simulation is the application of a model with the objective to derive strategies that help solve a problem or answer a question pertaining to a system.

The term *simulation* originates from the Latin word “simulare”, which means “to pretend”: in a simulation, the model pretends to be the real system.

Systems, Models, Simulations

System

(System)

A *system* is an object or a collection of objects whose properties we want to study.

Systems, Models, Simulations

Conceptual and Physical Models

The model used in the car example is something that exists in our minds only.

We can write it down on a paper in a few sentences and/or sketches, but it does not have any physical reality. Models of this kind are called *conceptual models.*

Conceptual models are used by each of us to solve everyday problems such as the car that refuses to start.

They are also applied by engineers or scientists to simple problems or questions similar to the car example.

If their problem or question is complex enough, however, they rely on experiments, and this leads us to other types of models.

For example; an engineer wants to reduce the fuel consumption of an engine:

In this case, the problem is the reduction of fuel consumption and the system is the engine. Assume that the systems analysis leads the engineer to the conclusion that the fuel injection pump needs to be optimized.

Typically, the engineer will then create some experimental setting where he can study the details of the fuel injection process.

Such an experimental setting is then a model in the sense that it will typically be a very simplified version of that engine, that is, it will typically involve only a few parts of the engine that are closely connected with the fuel injection process. In contrast to a conceptual model, however, it is not only an idea in our mind but also a real part of the physical world, and this is why models of this kind are called *physical models.*

The engineer will then use the physical model of the fuel injection process to derive strategies – for example, a new construction of the fuel injection pump – to reduce the engine’s fuel consumption, which is the simulation step of the above modeling and simulation scheme. Afterwards, in the validation step of the scheme, the potential of these new constructions to reduce fuel consumption will be tested in the engine itself, that is, in the real system.

Physical models are applied by scientists in a similar way. For example, let us think of a scientist who wants to understand the photosynthesis process in plants.

Similar to an engineer, the scientist will set up a simplified experimental setting – which might be some container with a plant cell culture – in which he can easily observe and measure the important variables, such as CO_2, water, light, and so on.

Mathematics as a Natural Modeling Language

Input–Output Systems

Any system that is investigated in science or engineering must be observable in the sense that it produces some kind of output that can be measured (a system that would not satisfy this minimum requirement would have to be treated by theologians rather than by scientists or engineers).

This observability condition can also be satisfied by systems where nothing can be measured directly, such as black holes, which produce measurable gravitational effects in their surroundings.
Most systems investigated in engineering or science do also accept some kind of input data, which can then be studied in relation to the output of the system (Figure 1.2a).

For example, a scientist who wants to understand photosynthesis will probably construct experiments where the carbohydrate production of a plant is measured at various levels of light, CO_2, water supply, and so on.

The light, CO_2, and water levels

Plant cell

The measured carbohydrate production

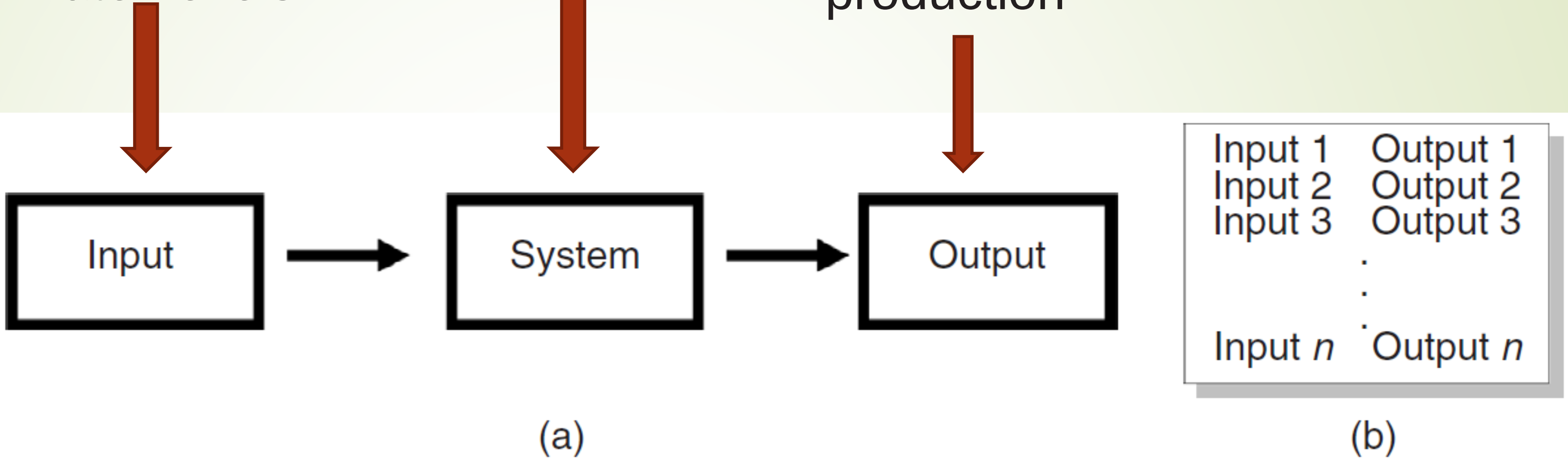


Fig. 1.2 (a) Communication of a system with the outside world. (b) General form of an experimental data set.

(Input–output systems)

Scientists or engineers investigate “input–output systems”, which transform given input parameters into output parameters.

Typically, however, such purely descriptive studies raise questions about the way in which the system works, and this is when input–output relations come into play.

Engineers, on the other hand, are always concerned with input–output relations since they are concerned with technology.

Technology

“the application of scientific knowledge to the practical aims of human life”.

These “practical aims” will usually be expressible in terms of a system output, and the tuning of system input toward optimized system output is precisely what engineers typically do, and what is in fact the genuine task of engineering.

Mathematics as a Natural Modeling Language

General Form of Experimental Data

The experimental procedure is used very generally in engineering and in the (empirical) sciences to understand, develop, or optimize systems. It is useful to think of it as a means to explore *black boxes*.

At the beginning of an experimental study, the system under investigation is similar to such a "black box" in the sense that there is some uncertainty about the processes that happen inside the system when the input is transformed into the output.

Typically, however, the experimenter will have some hypotheses about the internal processes, which he wants to prove or disprove in the course of his study. That is, experimenters typically are concerned with systems as *gray boxes* which are located somewhere between black and white boxes.

Depending on the hypothesis that the experimenter wants to investigate, he confronts the system with appropriate input quantities, hoping that the outputs produced by the system will help prove or disprove his hypothesis.

This is similar to a *question-and-answer game*: the experimenter poses questions to the system, which is the input, and the system answers to these questions in terms of measurable output quantities.

Mathematics as a Natural Modeling Language

Distinguished Role of Numerical Data

Now what is an appropriate method for the analysis of experimental datasets?

In most cases experimental data are numbers and can be quantified.

It is natural to think of a system in mathematical terms.

A system can be naturally seen as a mathematical function, which maps given input quantities x into output quantities $y = f(x)$.

This means that if one wants to understand the internal mechanics of a system “black box”, that is, if one wants to understand the processes inside the real system that transform input into output, a natural thing to do is to translate all these processes into mathematical operations.

If this is done, one arrives at a simplified representation of the real system in mathematical terms.

Now remember that a simplified description of a real system (along with a problem we want to solve) is a model by definition.

The representation of a real system in mathematical terms is thus *a mathematical model* of that system.

(Naturalness of mathematical models) Input–output systems usually generate numerical (or quantifiable) data that can be described naturally in mathematical terms.

Anyone concerned with systems and their input–output relations is also concerned with mathematical problems – regardless of whether he likes it or not and regardless of whether he treats the system appropriately using mathematical models or not. The success of his work, however, depends very much on the appropriate use of mathematical models.

Definition of Mathematical Models

To understand mathematical models, let us start with a general definition.
A mathematical model is a set of mathematical statements $M = \{1, 2, . . . , n\}$.

(Mathematical Model) A mathematical model is a triplet (S, Q, M) where S is a system, Q is a question relating to S, and M is a set of mathematical statements $M = \{1, 2, . . . , n\}$ which can be used to answer Q.

Let us look at another famous example that shows the importance of *Q*. Suppose we want to predict the behavior of some mechanical system *S*. Then the appropriate mathematical model depends on the problem we want to solve, that is, on the question *Q*.

If *Q* is asking for the behavior of *S* at moderate velocities, classical (Newtonian) mechanics can be used, that is, *M* = {*equations of Newtonian mechanics*}. If, on the other hand, *Q* is asking for the behavior of *S* at velocities close to the speed of light, then we have to set *M* = {*equations of relativistic mechanics*} instead.

Examples and Some More Definitions

Suppose we want to know the *mean age of some group of people*.

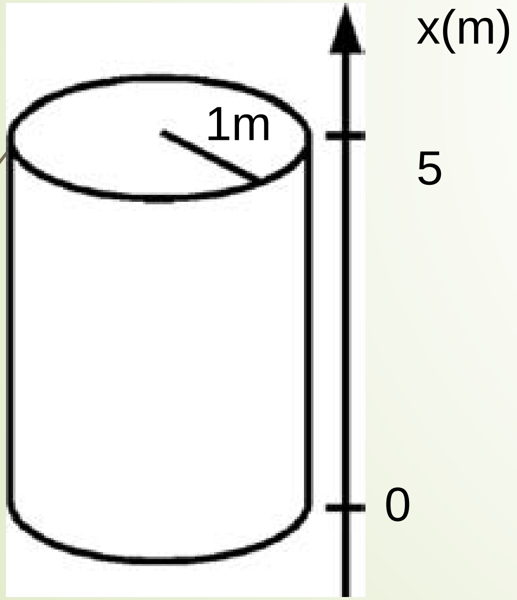
Then, we apply a mathematical model (*S*, *Q*, *M*) where

S is that group of people,

Q asks for their mean age, and

M is the mean value formula $\bar{x} = (\sum_{i=1}^{n} x_i)/n$.

Suppose we want to know the mass X of some substance in the cylindrical tank given a constant concentration c of the substance in that tank.



Then, a multiplication of the tank volume with c gives the mass X of the substance, that is,

$$X = 5\pi c$$

We apply a mathematical model (*S*, *Q*, *M*) where

S is the tank,

Q asks for the mass of the substance, and

M is $X = 5\pi c$.

An example involving more than simple algebraic operations is obtained if we assume that the *concentration c in the tank* of figure depends on the height coordinate, *x*.
In that case

$X = 5\pi c$ turns into

$$X = \pi \cdot \int_0^5 c(x)dx$$

In many mathematical models (S, Q, M) involving calculus, the question Q asks for the optimization of some quantity.
Suppose for example we want to *minimize the material consumption* of a cylindrical tin having a volume of 1 l.
In this case,

$$M = \{\pi r^2 h = 1,\ A = 2\pi r^2 + 2\pi rh \to \text{min}\}$$

can be used to solve the problem.

The second statement requires the surface area of the tin to be minimal, which is equivalent to a minimization of the metal used to build the tin.

The mathematical problem can be solved if one inserts the first equation of

$M = \{\pi r^2 h = 1, A = 2\pi r^2 + 2\pi rh \rightarrow \min\}$

into the second equation of,

$M = \{\pi r^2 h = 1, A = 2\pi r^2 + 2\pi rh \rightarrow \min\}$

which leads to

$A(r) = 2\pi r^2 + \frac{2}{r} \rightarrow \min$

$$r = \sqrt[3]{\frac{1}{2\pi}} \approx 0.54\ dm$$

$$h = \sqrt[3]{\frac{4}{\pi}} \approx 1.08\ dm$$

Examples and Some More Definitions

State Variables and System Parameters

Of course, each of us knows that a cylindrical tin can be described very easily based on its radius *r* and its height *h*.

This means everyone of us automatically applies the correct mathematical model, and hence, everybody automatically believes that the system in the tin problem is a simple thing.

But if we do not apply this model to the tin, it becomes a complex system.

Imagine a *Martian* or some other extraterrestrial being who never saw a cylinder before.

Suppose we would say to this Martian: “Look, here you have some sheets of metal and a sample tin filled with water. Make a tin of the same shape which can hold that amount of water, and use as little metal as possible.”

Then this Martian will – at least initially – see the original complexity of the problem. If he is smart, which we assume, he will note that infinitely many possible tin geometries are involved here.

From this original ("Martian") point of view we thus see that the system *S* of the tin example is quite complex, in fact an infinite-dimensional system.

And we see the power of the mathematical modeling procedure which reduces those infinite dimensions to only two, since the mathematical solution of the above problem involves only two parameters: *r* and *h* (or, equivalently, *r* and *A*).

Originally, the system "tin" in the above example is an infinite-dimensional thing not only with respect to its set of coordinates or the other aspects mentioned above, but also with respect to many other aspects which have been neglected in the mathematical model since they are unimportant for the solution of the problem, for example the thickness of the metal sheets, or its material, color, hardness, roughness and so on.

All the information which was contained in the original system *S*="tin" is reduced to a description of the system as a mere $Sr = \{r, h\}$ in terms of the mathematical model.

(A main benefit) The reduction of the information content of complex systems in terms of *reduced systems* is one of the main benefits of mathematical models.

(State variables) Let (*S*, *Q*, *M*) be a mathematical model. Mathematical quantities $s_1, s_2, \ldots, s_n$ which describe the state of the system *S* in terms of *M* and which are required to answer *Q* are called the *state variables* of (*S*, *Q*, *M*).

(Reduced system and system parameters) Let s_1, s_2, . . . ,s_n be the state variables of a mathematical model (S, Q, M). Let p_1, p_2, . . . ,p_m be mathematical quantities (numbers, variables, functions) which describe properties of the system S in terms of M, and which are needed to compute the state variables. Then $S_r = \{p_1, p_2, . . . , p_m\}$ is the *reduced system* and p_1, p_2, . . . , p_m are the *system parameters* of (S, Q, M).

This means that the state variables describe the system properties we are really interested in, while the system parameters describe system properties needed to obtain the state variables mathematically.

For example, in the *tank problem* above we were interested in the mass of the substance; hence, in this example we have one state variable, that is, $n = 1$ and $s_1 = X$.

To obtain s_1, we used the concentration c; hence, we have one system parameter in that example, that is, $m = 1$ and $p_1 = c$. The reduced system in this case is $S_r = \{c\}$.

By definition, the reduced system contains all information about the system which we need to get the state variable, that is, to answer Q.

In the *tin example*, we needed the surface area of the tin to answer Q, that is, in that case we had again one state variable $s_1 = A$.

On the other hand, two system parameters $p_1 = r$ and $p_2 = h$ were needed to obtain s_1, that is, in this case the reduced system is $S_r = \{r, h\}$.

(Importance of experiments) Typically, the properties (parameters) of the reduced system are those which need experimental characterization. In this way, the modeling procedure guides the experiments, and instead of making the experimenter superfluous (a frequent misunderstanding), it helps to avoid superfluous experiments.

Examples and Some More Definitions

Using Computer Algebra Software

Let us make a few more observations relating to the “1 l tin” example above. The mathematical problem behind this example can be easily solved using software. For example, using the *computer algebra software Maxima*, the problem can be solved as follows:

```
1: A(r):=2 *%pi *r ˆ 2 +2/r;
2: define(A1(r),diff(A(r),r));
3: define(A2(r),diff(A1(r),r));
4: solve(A1(r) =0);
5: r:rhs(solve(A1(r)=0)[3]);
6: r,numer;
7: A2(r)>0,pred;
```

Examples and Some More Definitions

The Problem Solving Scheme

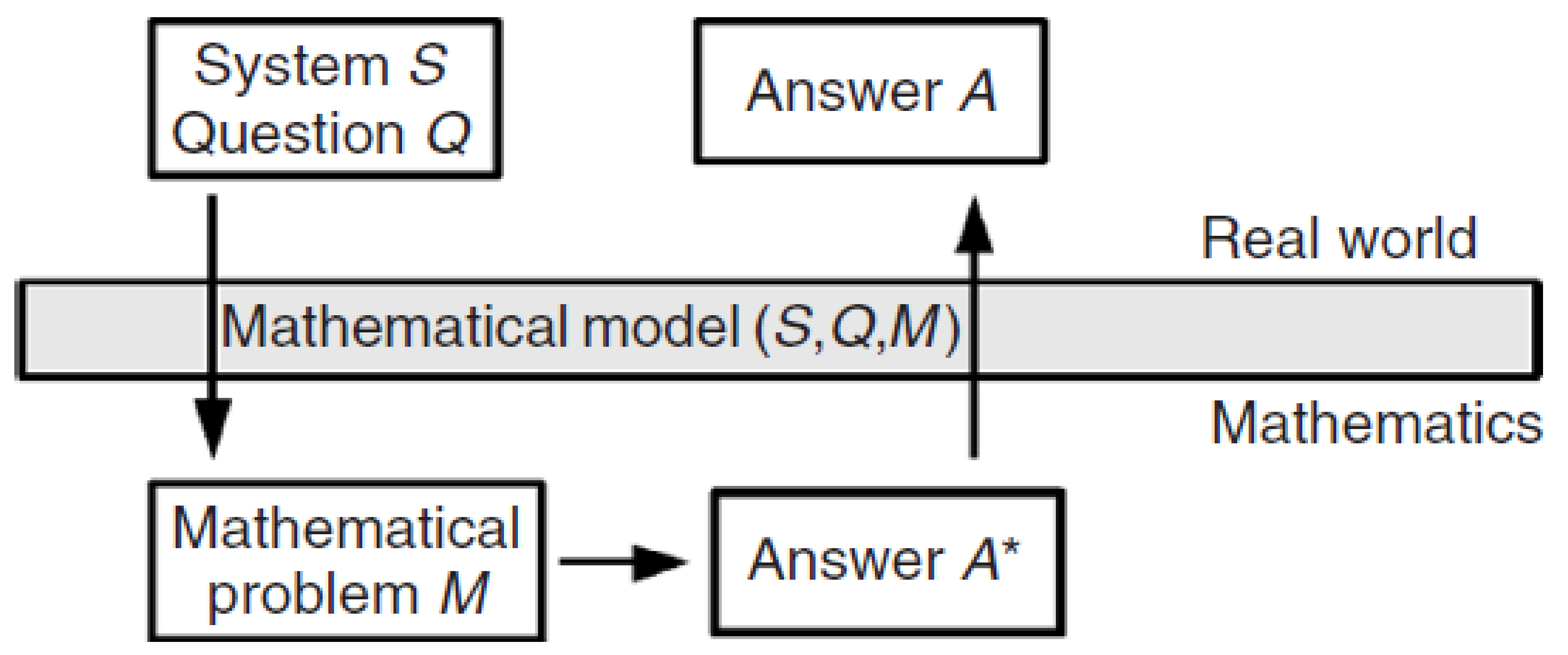
One can clearly distinguish between the formulation of a mathematical model on the one hand and the solution of the resulting mathematical problem on the other hand, which can be done with appropriate software.

This means that it is not necessary to be a professional mathematician if one wants to work with mathematical models. Of course, it is useful to have mathematical expertise. Mathematical expertise is particularly important if one wants to solve more advanced problems, or if one wants to make sure that the results obtained with mathematical software are really solutions of the original problem and no numerical artifacts.

The tin example shows another important advantage of mathematical modeling.

After the tin problem was formulated mathematically, the powerful and well-established mathematical methods of calculus became applicable.

Using the appropriate software, the problem could then be solved with little effort. Without the mathematical model for this problem, on the other hand, an experimental solution of this problem would have taken much more time.



Problem solving scheme

The starting point is a real-world system *S* together with a question *Q* relating to *S*. A mathematical model (*S*, *Q*,*M*) then opens up the way into the “mathematical universe”, where the problem can be solved using powerful mathematical methods. This leads to a problem solution in mathematical terms (*A**), which is then translated into an answer *A* to the original question *Q* in the last step.

The role of mathematics in problem solving scheme figure can be described like a subway train: since it would be a too long and hard way to go from the system *S* and question *Q* to the desired answer *A* in the real world, smart problem solvers go into the “mathematical underground”, where powerful mathematical methods provide fast trains toward the problem solution.

Examples and Some More Definitions

Strategies to Set up Simple Models

In many cases, a simple three-step procedure can be used to set up a mathematical model.

Problem 1:

Which volumes of fluids A and B should be mixed to obtain 150 l of a fluid C that contains $70 gl^{-1}$ of a substance, if A and B contain $50\ gl^{-1}$ and $80\ gl^{-1}$, respectively?

$x + y = 150$

$50x + 80y = 70 \cdot 150$

(Three steps to setup a model)

- *Step 1:* Determine the number of unknowns, that is, the number of quantities that must be determined in the problem. In many problem formulations, you just have to read the last sentence where the question is asked.
- *Step 2:* Give *precise* definitions of the unknowns, including units. It is a practical experience that this should not be lumped with step 1.
- *Step 3:* Reading the problem formulation sentence by sentence, translate this information into mathematical statements which involve the unknowns defined in step 2.

Let us apply this to *Problem 1* above.

In *step 1* and *step 2*, we would ascertain that *Problem 1* asks for two unknowns which can be defined as
- *x*: volume of fluid A in the mixture [l]
- *y*: volume of fluid B in the mixture [l]

These steps are important because they tell us about the unknowns that can be used in the equations. As long as the unknowns are unknown to us, it will be hard to write down meaningful equations in step 3.

Indeed, it is a frequent beginner’s mistake in mathematical modeling to write down equations which involve unknowns that are not sufficiently well defined. People often just pick up symbols that appear in the problem formulation – such as *A*, *B*, *C* in *problem 1* above – and then write down equations like

$$50A + 80B = 70$$

This equation is indeed *almost* correct, but it is hard to check its correctness as long as we lack any precise definitions of the unknowns.

The intrinsic problem with equations such as $50A + 80B = 70$ lies in the fact that A, B, C are already defined in the problem formulation.

There, they refer to the names of the fluids, although they are (implicitly) used to express the volumes of the fluids in $50A + 80B = 70$.

Thus, let us now write down the same equation using the unknowns x and y defined above:

$$50x + 80y = 70$$

Now the definitions of *x* and *y* can be used to check this equation.

What we see here is that on the left-hand side of $50x + 80y = 70$, the unit is (grams), which results from the multiplication of 50 gl^{-1} with *x* [l]. On the right-hand side, however, the unit is grams per liter.

So we have different units on the different sides of the equation, which proves that this is a wrong equation. At the same time, a comparison of the units may help us to get an idea of what must be done to obtain a correct equation.

In this case, it is obvious that a multiplication of the right-hand side with some quantity expressed in liter would solve the unit problem.
The only quantity of this kind in the problem formulation is the 150 l volume which is required as the volume of the mixture, and multiplying the 70 with 150 indeed solves the problem in this case.
A major problem in *step 3* is to identify those statements in the problem formulation which correspond to mathematical statements, such as equations, inequalities, and so on. The following note can be taken as a general guideline for this:

(Where are the equations?)

The statements of the problem formulation that can be translated into mathematical statements, such as equations, inequalities, and so on, are characterized by the fact that they impose restrictions on the values of the unknowns.

Let us analyze some of the statements in *Problem 1* above in the light of this strategy:

- *Statement 1*: 150 l of fluid C are required.
- *Statement 2*: Fluid A contains 50 gl^{-1} of the substance.
- *Statement 3*: Fluid B contains 80 gl^{-1} of the substance.
- *Statement 4*: Fluid C contains 70 gl^{-1} of the substance.

Obviously, *statement 1* is a restriction on the values of x and y, which translates immediately into the equation:

$$x + y = 150$$

Statement 2 and *statement 3*, on the other hand, impose no restriction on the unknowns. Arbitrary values of x and y are compatible with the fact that fluids A and B contain 50 gl^{-1} and 80 gl^{-1} of the substance, respectively.

Statement 4, however, does impose a restriction on *x* and *y*.

For example, given a value of *x*, a concentration of 70 gl^{-1} in fluid C can be realized only for one particular value of *y*.

Mathematically, *statement 4* can be expressed

$50x + 80y = 70 \cdot 150$

You may be able to write down this equation immediately. If you have problems to do this, you may follow a heuristic (i.e. not 100% mathematical) procedure, where you try to start as close to the statement in the problem formulation as possible.

In this case, we could begin with expressing *statement 4* as

{Concentration of substance in fluid C} = 70

Then, you would use the definition of a concentration as follows:

$$\frac{\{\text{Mass of substance in fluid C}\}}{\{\text{Volume of the mixture}\}} = 70$$

The next step would be to ascertain two things:

- The mass of the substance in fluid C comes from fluids A and B.
- The volume of the mixture is 150 l.

$$\frac{\{\text{Mass of substance in fluid } A\} + \{\text{Mass of substance in fluid } B\}}{\{\text{Volume of the mixture}\}} = 70$$

The masses of the substance in A and B can be easily derived using the concentrations given in *Problem 1* above:

$$\frac{50x + 80y}{150} = 70$$

(Heuristic procedure to set up mathematical statements) If you want to translate a statement in a problem formulation into a mathematical statement, such as an equation or inequality, begin by mimicking the statement in the problem formulation as closely as possible. Your initial formulation may involve nonmathematical statements. Try then to replace all nonmathematical statements by expressions involving the unknowns.

Examples and Some More Definitions

Strategies to Set up Simple Models - Mixture Problem

Problem 2:

Suppose the fluids *A*, *B*, *C*, *D* contain the substances S_1, S_2, S_3 according to the following table (concentrations in grams per liter):

	A	B	C	D
S_1	2.5	8.2	6.4	12.7
S_2	3.2	15.1	13.2	0.4
S_3	1.1	0.9	2.2	3.1

What is the concentration of S_3 in a mixture of these fluids that contains 75% (percent by volume) of fluids *A* and *B* and which contains 4 gl^{-1} and 5 gl^{-1} of the substances S_1 and S_2, respectively?

Referring to *step 1* and *step 2* of the three-step procedure, it is obvious that we have only one unknown here which can be defined as follows:

- x: concentration of S_3 in the mixture (grams per liter)

Now *step 3* requires us to write down mathematical statements involving x.

Three statements of this kind can be identified:

- *Statement 1*: 75% of the mixture consists of *A* and *B*.
- *Statement 2*: The mixture contains 4 gl^{-1} of S_1.
- *Statement 3*: The mixture contains 5 gl^{-1} of S_2.

Each of these statements excludes a great number of possible mixtures and thus imposes a restriction on *x*.

Beginning with *statement 1*, it is obvious that this statement cannot be formulated in terms of *x*.

We are here in a situation where a number of auxiliary variables is needed to translate the problem formulation into mathematics.

(Auxiliary variables) In some cases, the translation of a problem into mathematics may require the introduction of *auxiliary variables*. These variables are "auxiliary" in the sense that they help us to determine the unknowns.
Usually, the problem formulation will provide enough information such that the auxiliary variables *and* the unknowns can be determined (i.e. the auxiliary variables will just increase the size of the system of equations).

In this case, we obviously need the following auxiliary variables:

- x_A: percent (by volume) of fluid A in the mixture
- x_B: percent (by volume) of fluid B in the mixture
- x_C: percent (by volume) of fluid C in the mixture
- x_D: percent (by volume) of fluid D in the mixture

Now the above *statement 1* can be easily expressed as

$$x_A + x_B = 0.75$$

Similar to above, *statement 2* and *statement 3* can be formulated as

{Concentration of S_1 in the mixture} = 4

and

{Concentration of S_2 in the mixture} = 5

Based on the information provided in the above table (and again following a similar procedure as in the previous section), these equations translate to

$$2.5x_A + 8.2x_B + 6.4x_C + 12.7x_D = 4$$

and

$$3.2x_A + 15.1x_B + 13.2x_C + 0.4x_D = 5$$

Since x is the concentration of S_3 in the mixture, a similar argumentation shows

$$1.1x_A + 0.9x_B + 2.2x_C + 3.1x_D = x$$

So far we have the four equations and five unknowns x, x_A, x_B, x_C, and x_D, that is, we need one more equation. In this case, the missing equation is given implicitly by the definition of x_A, x_B, x_C, and x_D. These variables express percent values, and hence, we have

$$x_A + x_B + x_C + x_D = 1$$

Altogether, we have now obtained the following system of linear equations:

$$x_A + x_B = 0.75$$

$$2.5x_A + 8.2x_B + 6.4x_C + 12.7x_D = 4$$

$$3.2x_A + 15.1x_B + 13.2x_C + 0.4x_D = 5$$

$$1.1x_A + 0.9x_B + 2.2x_C + 3.1x_D = x$$

$$x_A + x_B + x_C + x_D = 1$$

Examples and Some More Definitions

Linear Programming

All mathematical models considered so far were formulated in terms of equations only. A mathematical model may involve any kind of mathematical statements. For example, it may involve inequalities.

One of the simplest class of problems involving inequalities are linear programming problems that are frequently used e.g. in operations research.

Consider the following problem taken from the linear programming article of *Wikipedia.org:*

Linear programming example

Suppose a farmer has a piece of farm land, say A square kilometers large, to be planted with either wheat or barley or some combination of the two. Furthermore, suppose the farmer has a limited permissible amount F of fertilizer and P of insecticide which can be used, each of which is required in different amounts per unit area for wheat (F_1, P_1) and barley (F_2, P_2). Let S_1 be the selling price of wheat, and S_2 the price of barley. How many square kilometers should be planted with wheat versus barley to maximize the revenue?

Denoting the area planted with wheat and barley with x_1 and x_2 respectively, the problem can be formulated as follows:

$x_1, x_2 \geq 0$ (the farmer cannot plant a negative area)

$x_1 + x_2 \leq A$

(no more than the given A square kilometers of farm land can be used)

$F_1x_1 + F_2x_2 \leq F$ (fertilizer limits)

$P_1x_1 + P_2x_2 \leq P$ (insecticide limits)

$S_1x_1 + S_2x_2 \rightarrow \max$ (required revenue maximization)

S → the farm

Q → “How many square kilometers should be planted with wheat versus barley to maximize the revenue?”

M → Above Equations

A mathematical model (*S*, *Q*, *M*) is obtained.

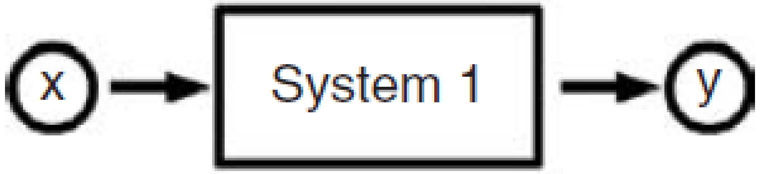
Examples and Some More Definitions

Modeling a Black Box System

It was mentioned that the systems investigated by scientists or engineers typically are “input–output systems”, which means they transform the given input parameters into output parameters.

In the tin example, the radius and height of the tin are input parameters and the surface area of the tin is the output parameter.

The exploration of an example input–output system in some more detail will now lead us to further important concepts and definitions. Assume a “system 1” as in the figure which produces an output length *y* (centimeters) for every given input force *x* [N].



(a)

x (N)	10	20	30	40	50
y (cm)	3	5	11	12	16

(b)

(a) System 1 with input *x* (N) and output *y* (cm)
(b) System 1 data

Furthermore, assume that we do not know about the processes inside the system that transform *x* into *y*, that is, let this system be a "black box" to us as described above.

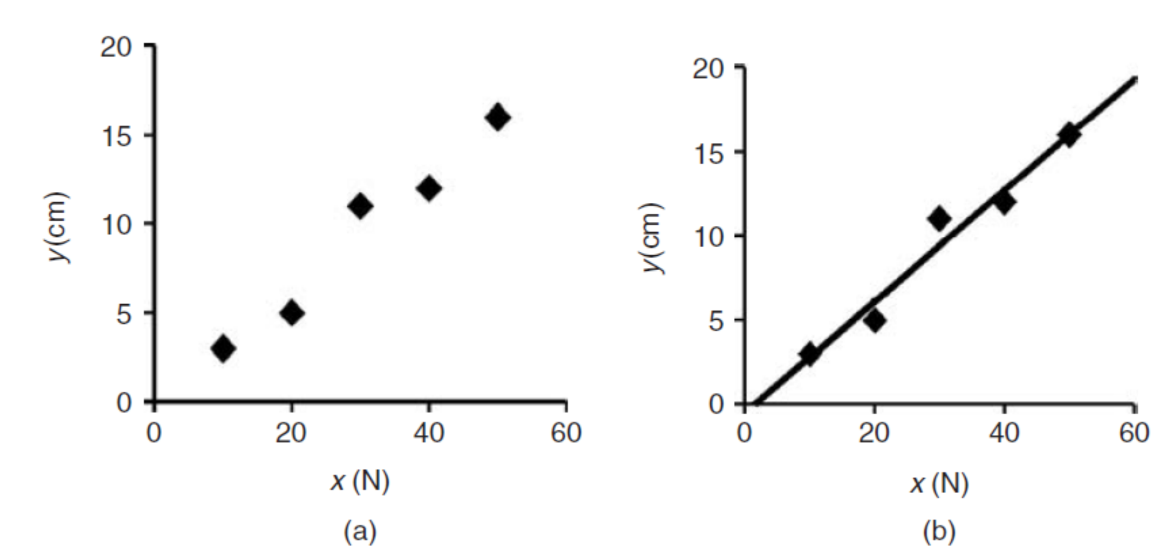
Consider the following problem:

Q: Find an input *x* that generates an output *y* = 20 cm.
S: System 1
We are now looking for an appropriate set of mathematical statements *M* that can help us to answer *Q*.

All that the investigator of system 1 can do is to produce some data using the system, hoping that these data will reveal something about the processes occurring inside the "black box".

To see what happens, the investigator will probably produce a plot of the data as in the figure.

(a) Plot of the data



(b) System 1 data with regression line.

Mathematically, this means that the function $y = f(x)$ behind the data is a straight line:

$f(x) = ax + b$

Now the investigator can apply a statistical method called *linear regression* to determine the coefficients *a* and *b* of this equation from the data, which leads to the “regression line”

$f(x) = 0.33x - 0.5$

Figure b shows that there is a good coincidence or, in statistical terminology, a good “fit” between this regression line and the data.

$f(x) = 0.33x - 0.5$
can now be used as the *M* of a mathematical model of system 1.

The question *Q* stated above ("Which system input *x* generates a desired output $y = 20$ cm? ") can then be easily answered by setting $y = f(x) = 20$, that is,

$$20 = 0.33x - 0.5$$

which gives $x \approx 62.1$ N.

Of course, this is just an *approximate result* for several reasons.
First of all, figure shows that there are some deviations between the regression line and the data. These deviations may be due to measurement errors, but they may also reflect some really existing effects.

The example shows the *importance of statistical methods* in mathematical modeling.
First of all, statistics itself is a collection of mathematical models that can be used to describe data or to draw inferences from data.

Beyond this, statistical methods provide a necessary link between nonstatistical mathematical models and the real world.

In mathematical modeling, one is always concerned with experimental data, not only to validate model predictions, but also to develop hypotheses about the system, which help to set up appropriate equations.

In the example, the data led us to the hypothesis that there is a linear relation between x and y.

Even More Definitions

Phenomenological and Mechanistic Models

The mathematical model used above to describe system 1 is called a *phenomenological model* since it was constructed based on experimental data only, treating the system as a black box, that is, without using any information about the internal processes occurring inside system 1 when *x* is transformed into *y*.

On the other hand, models that are constructed using information about the system S are called *mechanistic models*, since such models are virtually based on a look into the internal mechanics of *S*.

(Phenomenological and mechanistic models)

A mathematical model (*S*, *Q*, *M*) is called

- *phenomenological*, if it was constructed based on experimental data only, using no a priori information about *S*,

- *mechanistic*, if some of the statements in *M* are based on a priori information about *S*.

Phenomenological models are also called *empirical models*, *statistical models*, *data-driven models* or *black box models* for obvious reasons.

Mechanistic models for which all necessary information about *S* are available are also called *white box models*.

Most mechanistic models are located somewhere between the extreme black and white box cases, that is, they are based on some information about *S* while some other important information is unavailable. Such models are sometimes called *gray box models* or *semi-empirical models.*

Phenomenological models are also called *empirical models*, *statistical models*, *data-driven models* or *black box models* for obvious reasons.

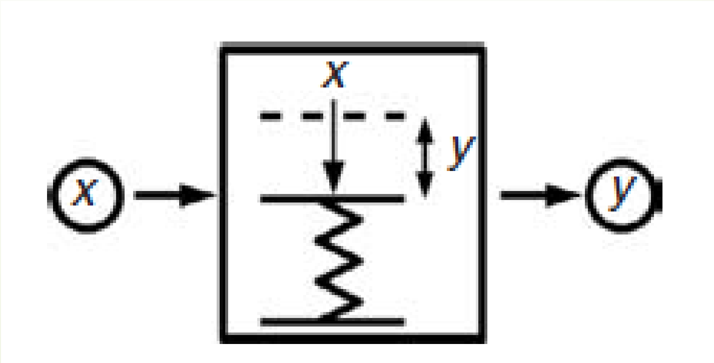
Mechanistic models for which all necessary information about *S* are available are also called *white box models*.

Most mechanistic models are located somewhere between the extreme black and white box cases, that is, they are based on some information about *S* while some other important information is unavailable. Such models are sometimes called *gray box models* or *semi-empirical models.*

Assume that system 1 is a mechanical spring, *x* is a force acting on that spring, and *y* is the resulting elongation. This is now an *a priori information* about system 1, and it can be used to construct a mechanistic mathematical model based on elementary physical knowledge.

Mechanical springs can be described by Hooke's law.

$$x = k \cdot y$$



Now the following mechanistic mathematical model (*S*, *Q*, *M*) is obtained:

- *S*: System 1
- *Q*: Which system input *x* generates a desired output of *y* = 20 cm?
- *M*: $x = k \cdot y$

The mechanistic model of system 1 has several important advantages compared to the phenomenological model, and these advantages are *characteristic advantages of the mechanistic approach*.

Mechanistic models do also allow *better predictions* of modified systems.

Another advantage of mechanistic models is the fact that they usually involve *physically interpretable parameters*, that is, parameters which represent real properties of the system.

(Phenomenological vs. mechanistic)

Phenomenological models are universally applicable, easy to set up, but limited in scope.

Mechanistic models typically involve physically interpretable parameters, allow deeper insights into system performance and better predictions, but they require a priori information on the system and often need more time and resources.

Even More Definitions

Stationary and Instationary models

(Stationary/instationary models)

A mathematical model (S,Q, M) is called

- *instationary*, if at least one of its system parameters or state variables depends on time and

- *stationary* otherwise.

Even More Definitions

Distributed and Lumped models

(Distributed/lumped models)

A mathematical model (S, Q,M) is called

- *distributed*, if at least one of its system parameters or state variables depends on a space variable,

- *lumped* otherwise.

Classification of Mathematical Models

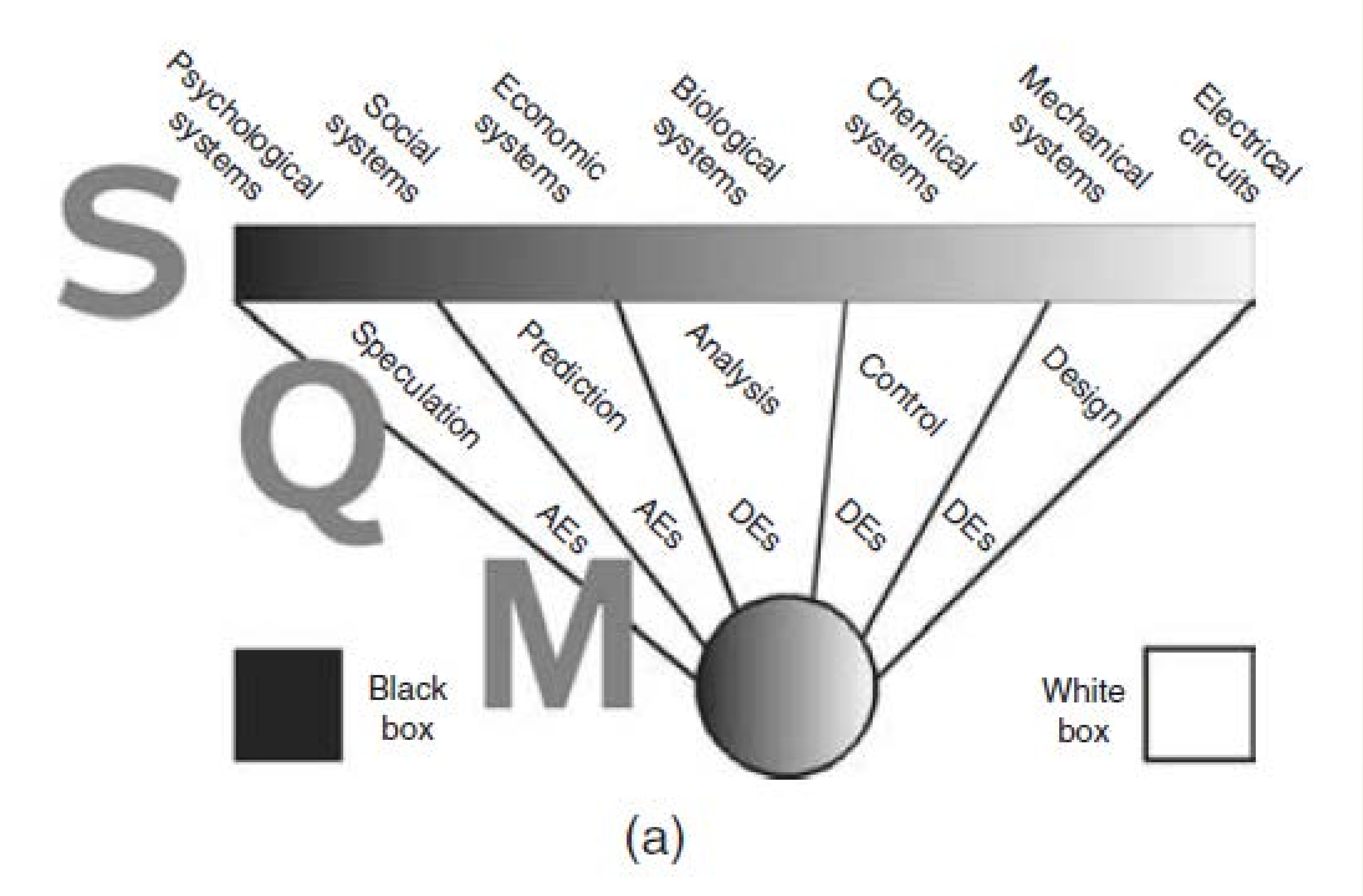
The practical use of a classification of mathematical models lies in the fact that you understand “where you are” in the space of mathematical models, and which types of models might be applicable to your problem beyond the models that you have already used.

Classification of Mathematical Models

From Black to White Box Models

The "space of mathematical models" evolves naturally from definiton, where we have defined a mathematical model to be a triple (S,Q,M) consisting of a system S, a question Q, and a set of mathematical statements M.

Based on this definition, it is natural to classify mathematical models in an *SQM space*.



(a) Classification of mathematical models between black and white box models

Note that the three dimensions of a mathematical model (*S*, *Q*, *M*) can be seen in the figure:
the systems (*S*) are classified on top of the bar, immediately below the bar there is a list of objectives that mathematical models in each of the segments may have (which is *Q*), and at the bottom end there are corresponding mathematical structures (*M*) ranging from algebraic equations (AEs) to differential equations (DEs).

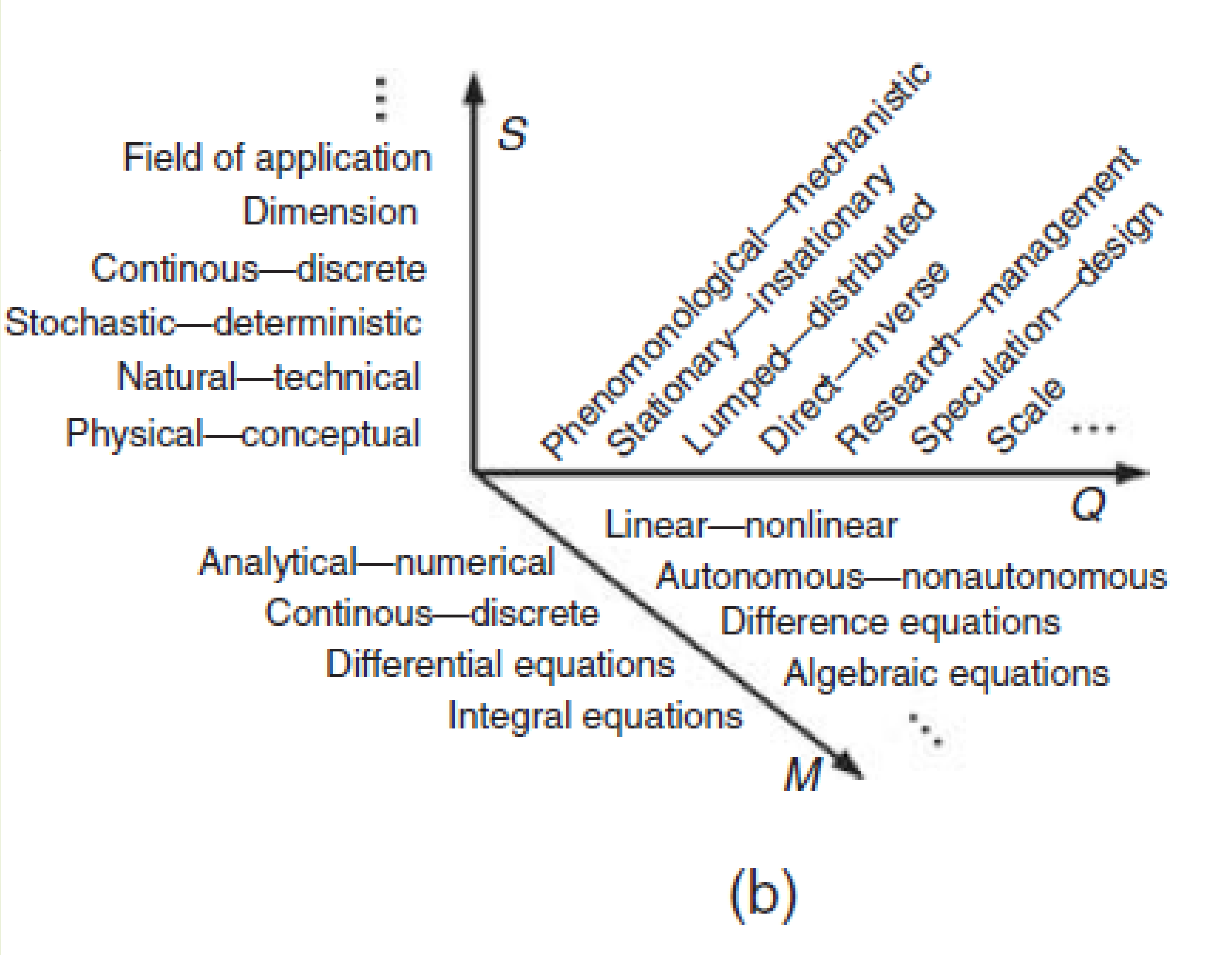
$f(x) = ax + b$ *is an example of AEs*

Classification of Mathematical Models

SQM Space Classification: S Axis

Since mathematical models are characterized by their respective individual *S*, *Q* and *M* “values”, one can also think of each model as being located somewhere in the “SQM space” of figure.

On each of the *S*-, *Q*- and *M*-axes of the figure, mathematical models are classified with respect to a number of criteria which were compiled based on various classification attempts in the literature.



(b) Classification of mathematical models in the SQM space.

Physical – conceptual.

Physical systems are part of the real world, for example, a fish or a car. Conceptual systems are made up of thoughts and ideas, for example, a set of mathematical axioms. We focus entirely on physical systems.

Natural – technical.

Naturally, a natural system is a part of nature, such as a fish or a flower, while a technical system is a car, a machine, and so on.

Stochastic – deterministic.

Stochastic systems involve random effects, such as rolling dice, share prices and so on.

Deterministic systems involve no or very little random effects, for example, mechanical systems, such as the planetary system, a pendulum, and so on. In a deterministic system, a particular state A of the system is always followed by one and the same state B, while A may be followed by B, C or other states in an unpredictable way if the system is stochastic.

Continuous – discrete.

Continuous systems involve quantities that change continuously with time, such as sugar and ethanol concentrations in a wine fermenter.

Discrete systems, on the other hand, involve quantities that change at discrete times only, such as the number of individuals in animal populations.

Note that on the M axis of figure, continuous systems can be represented by discrete mathematical statements and vice versa (e.g. a continuous mathematical formulation is used to describe the discrete predator–prey system).

Dimension.

Depending on their spatial symmetries, physical systems can be described using 1, 2, or 3 space variables. The number of space variables used to describe a physical system is called its *dimension* (frequently denoted 1D, 2D, or 3D).

Field of application.

We can distinguish between chemical systems, physical systems, biological systems, and so on.

Classification of Mathematical Models

SQM Space Classification: Q Axis

Phenomenological – mechanistic.

This has been discussed in detail before.

Stationary – instationary.

It depends on the question which we are asking (i.e. on the “*Q*” of a mathematical model (*S*, *Q*, *M*)) whether a stationary (time-independent) or instationary (time-dependent) model is appropriate.

Lumped – distributed.

It depends on the question which we are asking (i.e. on the “*Q*” of a mathematical model (*S*, *Q*, *M*)) whether a lumped (space-independent) or distributed (space-dependent) model is appropriate.

Direct – inverse.

Consider an input–output system. If *Q* assumes given input and system parameters and asks for the output, the model solves a so-called *direct problem.*

If, on the other hand, *Q* asks for the input or for parameters of *S*, the model solves a so-called inverse problem.

If *Q* asks for parameters of *S*, the resulting problem is also called a *parameter identification problem*.

If *Q* asks for input parameters, the resulting problem is also called a *control problem*, since in this case the problem is to *control* the input in a way that generates some desired output.

Research – management.

Research models are used if *Q* aims at the understanding of *S*; management models, on the other hand, are used if the focus is on the solution of practical problems related to *S*. Research models tend to be more complex and less manageable from a practical point of view.

Depending on *Q*, the same mathematical equations can be a part of a research or of a management model. For example, the predator–prey model is a research model if the investigator just wants to understand the oscillations of the predator and prey populations, and it is a management model if is used to control the predator and prey populations.

Speculation – design.

Scale.

Depending on *Q*, the model will describe the system on an appropriate scale. For example, depending on *Q* it can be appropriate to virtually follow a fluid particle on its way through the complex channels of a porous medium, or just to compute the pressure drop across a porous medium based on its permeability.

Obviously, these cases correspond to a description of a porous medium on two scales (microscopic/macroscopic).

Classification of Mathematical Models

SQM Space Classification: M Axis

Linear – nonlinear.

In linear models, the unknowns (or their derivatives) are combined using linear mathematical operations only, such as addition/subtraction or multiplication with parameters.

Nonlinear models, on the other hand, may involve the multiplication of unknowns, the application of transcendental functions, and so on. Nonlinear models typically have more (and more interesting) solutions but are harder to solve.

Analytical – numerical.

In analytic models, the system behavior can be expressed in terms of mathematical formulas involving the system parameters. Based on these models, qualitative effects of parameters and the entire system behavior can be studied theoretically, without using concrete values for the parameters.

Numerical models, on the other hand, can be used to obtain the system behavior for specific parameter values.

Autonomous – nonautonomous.

This is a mathematical classification of instationary models. If an equation does not depend explicitly on time, it is called *autonomous*, otherwise nonautonomous.

Continuous – discrete.

In continuous models, the independent variables may assume arbitrary (typically real) values within some interval. For example, many of the ODE models use time (within some time interval) as the independent variable. In discrete models, on the other hand, the independent variables may assume some discrete values only.

Difference equations.

In difference equations, the quantity of interest is obtained as a sequence of discrete values. Usually, this is expressed in terms of recurrence relations in which each term of the sequence depends on previous terms. Difference equations are frequently used to describe discrete systems.

Differential equations.

Differential equations are equations involving derivatives of an unknown function. They are a main tool to set up continuous mechanistic models.

Integral equations.

Integral equations are equations involving an integral of an unknown function.

Algebraic equations.

AEs are equations involving the usual algebraic operations such as addition, subtraction, division, and so on.

Phenomenological Models

Elementary Statistics - Descriptive Statistics

The first thing that is usually done with a given dataset is *descriptive statistics*, that is, the application of methods that summarize and describe the data.

Table: Spring data.

x	10	20	30	40	50
y	3	5	11	12	16

The simplest thing that one can do with a dataset is to compute *measures of position*, which characterize the approximate location of the data in various ways.

The most well known and most frequently used measure of position is the *arithmetic mean*, which is defined as follows:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

For example,the arithmetic means of the *x* and *y* are *x* = 30 and *y* = 9.4, respectively, which basically says that the *x* and *y* data spread around these values.

After computing measures of position, the next step usually is to look at *measures of variation* (or measures of *statistical dispersion*), which basically measure how widely spread the values in a dataset are. The most popular measure of variation is the *sample standard deviation*

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x_i})^2}{n-1}}$$

s ≈ 15.8 and *s* ≈ 5.3 when applied to the *x* and *y* data, respectively.

The sample standard deviation measures the variability of the data in terms of the deviations from the mean.
Basically, the sample standard deviation expresses an average of these (squared) deviations. To understand the meaning of this expression a little more, let us look at what is meant by a “sample” here.
Statistical investigations typically focus on a well-defined collection of objects which constitute what is called a *population.*

For example, if an investigator wants to characterize the impact of a nutrient on a particular plant species, then his investigation will involve a population consisting of all plants of this species. In many cases, it will be impossible and inefficient to investigate the entire population due to limited time and resources (e.g. the plant species under investigation may cover most of the earth’s surface).

Statistical investigations will thus typically be restricted to a subset of a population which is called a *sample*. A number of strategies such as *random sampling* (each member of the population has an equal chance of being selected) or *stratified sampling* (which uses a division of the population into subgroups sharing the same characteristics such as gender or age) are used to make sure that the sample represents the entire population as good as possible.

Other frequently used measures of variation include

- the (sample) **range**, that is, the difference between the maximum and minimum values in the sample;

- the (sample) **average deviation**, that is, the mean of the absolute deviations $|x_i - x|$;

- various dimensionless measures such as the (sample) **coefficient of variation** $c_v = s/\bar{x}$,

Phenomenological Models

Elementary Statistics - Random Processes and Probability

Suppose you are interested in some quantity which we denote by X, and which may be temperature, the concentration of some substance, and so on.

You will usually need to have precise measurements of that quantity, so let us assume that you have a new measurement device and want to know about the measurement errors produced by that device.

Then, a standard procedure is to repeatedly measure that quantity in a situation where the correct result is known (e.g. by using standardized solutions if X is the concentration of some substance).

Assuming that the true value of the quantity of interest is 20, the data produced in this way may look like this:

20.13443 19.83828 20.01702 19.99835 19.94526 20.01415 19.96707

What we see here is that the measurement values oscillate in a random way around the true value. Most measurement devices produce random errors of this kind, which is no problem as long as the amplitude of these oscillations is small enough.

Now a natural question regarding the above data is this:

What is the probability with which the deviations of the measurement value from the true value will be less than some specified value such as 0.1?

In statistical terms, we would say that the above data have been generated by the *random variable X.*

(Random variable)

A random variable is a variable that has a single numerical value, determined by chance, for each outcome of a procedure.

Perhaps the most classical example is the random variable

X_1 : result of a dice

X_2: waiting time at a bus stop if you arrive there without knowing the time table and many other examples.

Let us ask for the probability with which a random variable attains certain values.
This is an easy thing if one is concerned with simple systems such as a dice.
Everyone of us knows that the probability of getting a "3" in a dice play is 1/6 or 16.7%. In statistics, this is usually written as

$$P(X_1 = 3) = 1/6$$

(Events and sample space)

- An *event* is any collection of results or outcomes of a procedure.

- A *simple event* is an outcome or an event that cannot be further broken down into simpler components.

- The *sample space* for a procedure consists of all possible *simple* events.

In the dice example, the sample space would be

$S = \{1, 2, 3, 4, 5, 6\}$

and all subsets $A \subset S$ such as

$A_1 = \{1, 2\}$ ("dice result is below 3") or
$A_2 = \{1, 3, 5\}$ ("dice result is an odd number")
would be events in the sense of the above definition.

Examples of simple events would be $A3 = \{2\}$, $A4 = \{5\}$, and so on.

In the bus-waiting-time example, the sample space would be

$S = \{x \in R | 0 \leq x < 15\}$

if we assume that the buses arrive in 15-min intervals, and a possible event would be [0, 2[(“the waiting time is below 2min”).

The probability function *P* is usually defined based on axioms.

(Probability)

Given a sample space *S*, the *probability function P* assigns to each event $A \subset S$ a number $P(A) \in [0, 1]$, called the *probability* of the event *A*, which will give a precise measure of the chance that A will occur.

(Classical approach to probability)

Assume that a given procedure has *n* different simple events and that each of those simple events has an equal chance of occurring. If event *A* can occur in *s* of these *n* ways, then

$$P(A) = s/n$$

This formula works well for the dice and many other similar *discrete random variables* that involve a finite number of equally likely possible results (note that discrete random variables may also involve countable infinitely many possible results).

It does not work, however, for *continuous random variables* with an infinite number of possible results similar to the random variable X_2 discussed above that describes the bus waiting time.

$S = \{x \in R | 0 \leq x < 15\}$ of this example indeed involves an infinite number of continuously distributed possible results between 0 and 15 min. In this case, the following formula can be used:

(Relative frequency approximation)

Assume that a given procedure is repeated n times, and let $f_n(A)$ denote the relative frequency with which an event A occurs. Then,

$$P(A) = \lim_{n \to \infty} f_n(A)$$

This means that if we, for example, want to approximate the probability of bus waiting times between 0 and 2 min (i.e. the probability of A = [0, 2[), the following approximation can be used

$$P(A) \approx f_n(A)$$

and the quality of this approximation will increase as n is increased.