

Computer Simulations in Social Sciences

by

Marco Valente

`mv@business.auc.dk`

Content

- Making science as generating models
- “Hard” vs. “Soft” sciences
- Simulating historical phenomena
- Technical requirements for scientific simulations
- Simulation languages pro's and con's
- *Laboratory for Simulations Development - Lsd*
- Model elements in Lsd

Making science as generating models

The goals of scientific research are:

- Making sense observed events
- Forecasting future events
- Providing tools to desirable results

Making science as generating models

When making research we need to build a model. A model is a simplified representation of one or more real events that we can use for one of the three goals, or all of them.

The importance of a model is measured by simplicity, generality and accuracy. That is, less elements it uses, the larger the number of real events it encompasses, and the more precise its predictions.

Making science as generating models

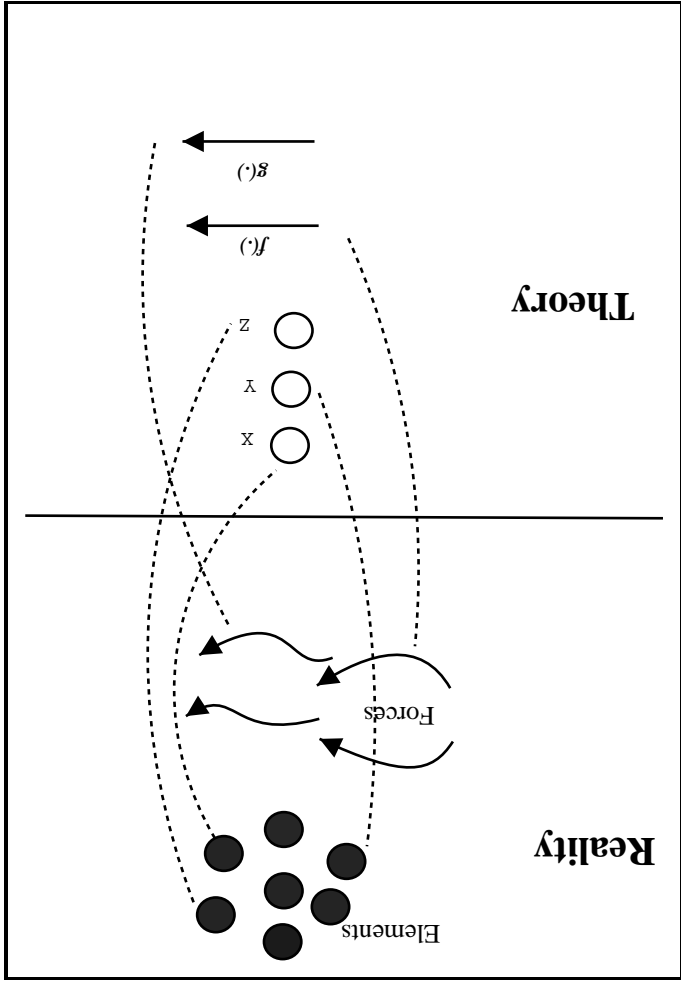
Let's see how a model is generated and used. We will see why mathematics a powerful modelling tool, but has limitations in cases where it can be applied. We will also define in which cases simulations are a better modelling strategy.

Making science as generating models

A scientific research concerns some reality. We can define such reality as entities and actions that modify the entities. Let's call the *objects* and *forces* that influence objects.

The first step when building a model consists in individuating which objects and forces are relevant and which are not. That is, we select from the reality some objects and forces that consider crucial, and decide to ignore others that we assume irrelevant for our goals.

Selecting elements of the model



Making science as generating models

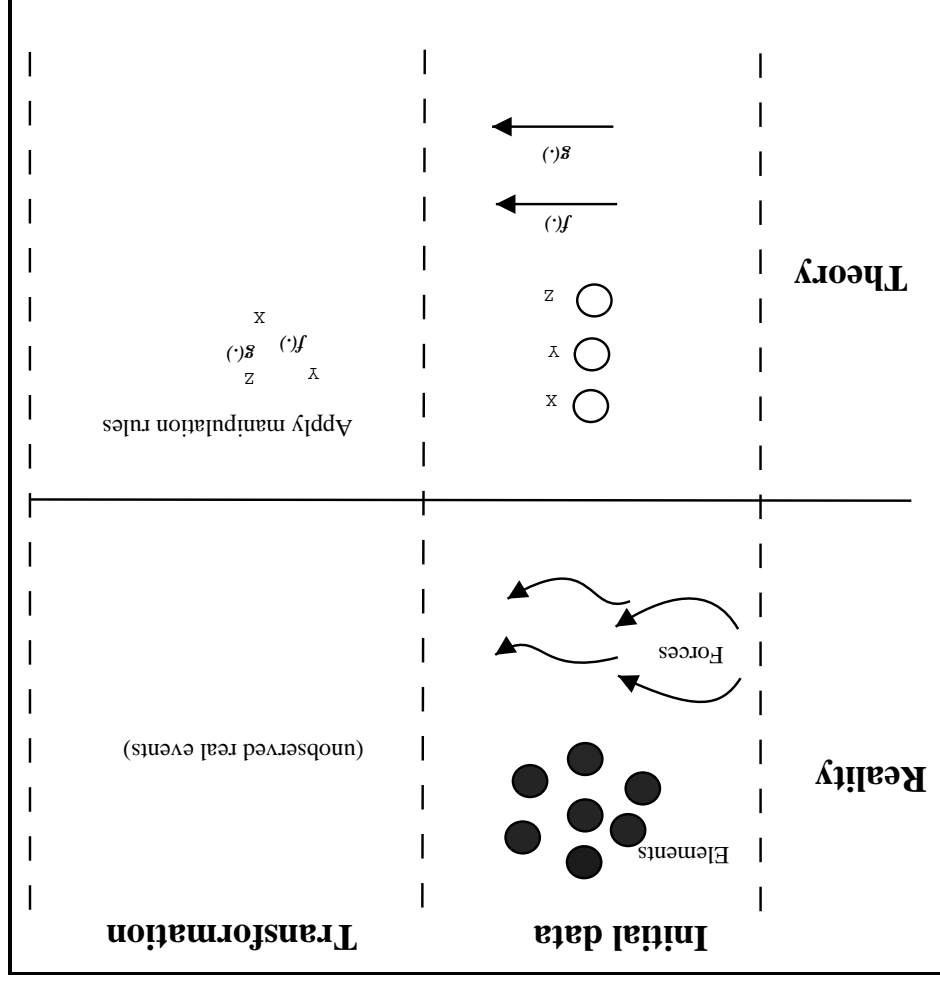
We can label the operation of selecting the elements of interest and deciding their representation as a process of *abstraction*.

Note that a model always simplifies the reality, deciding to leave out of the analysis some elements. Moreover, the representation used in the model is an approximation, that consider only certain aspects of the real world elements and ignores others.

Making science as generating models

The second step consist in manipulating the elements of the model according to the rules of a logical language.
The elaboration is made only on the symbols, and may, or may not, reflect the actual events in real world.

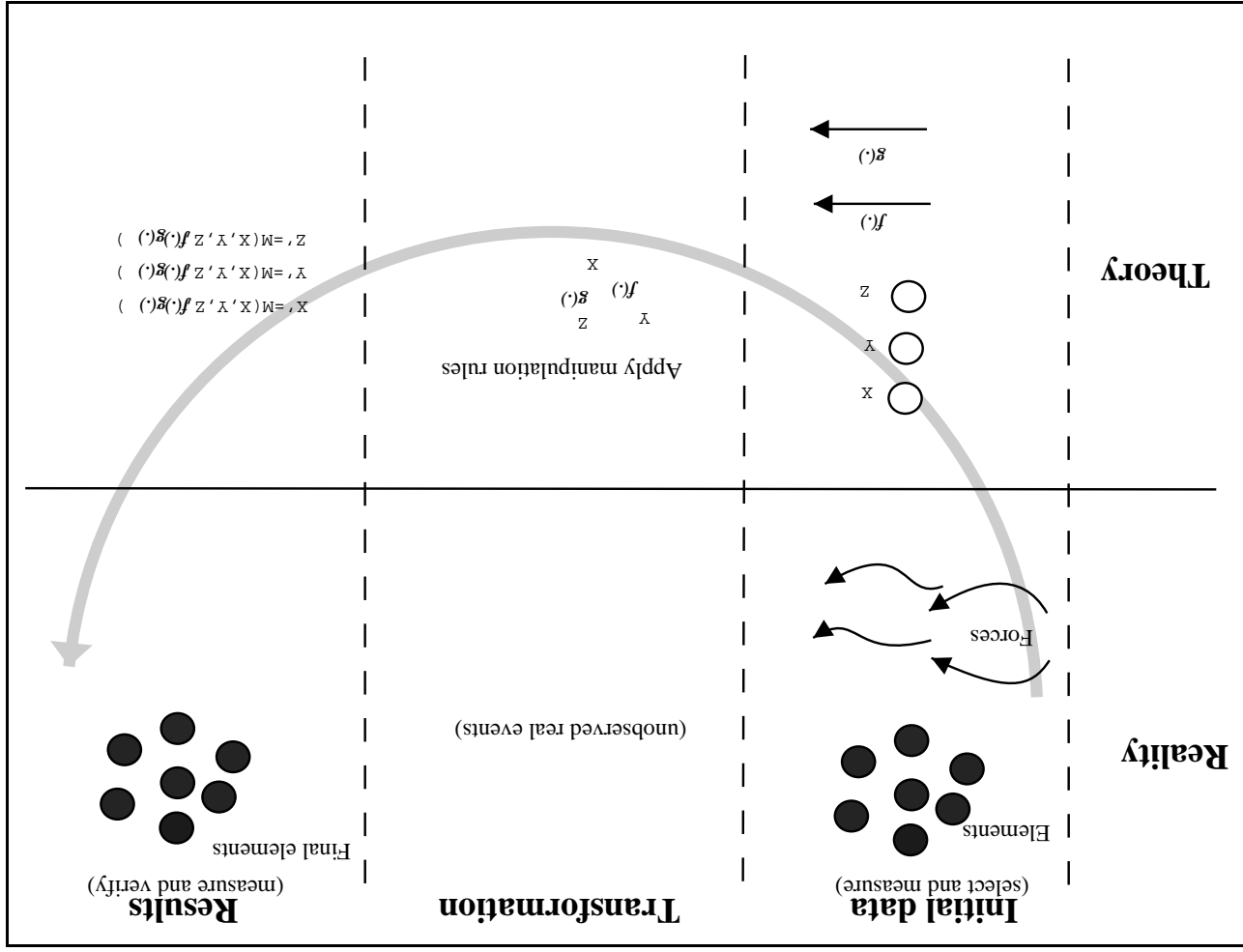
Manipulating the elements of the model



Making science as generating models

Eventually, we obtain the *results* of the model as new symbolic elements. These results are compared with the observation of the real elements to test the prediction of the model, or any other criterion used to assess the validity of the model.

Verifying the results of the model



Making science as generating models

In summary, making a model consists in considering two observations of the real world, which may be the same system in two different times, or a system observed from two different perspectives. The model is built through the abstraction process that determines the symbols to use. The elaboration on the symbols generates a new set of symbols, or a different configuration of the model, as result. Then, the result is tested with the second observation by comparing its result with another observation.

Making science as generating models

Whenever one tries to make sense of some real-world events it is necessary to build some sort of model. We may say that building models *is* the process of understanding the reality. In this perspective, even a history, well told, is itself a kind of a model: the historian selects some events, and explains how they interact to generate a subsequent event.

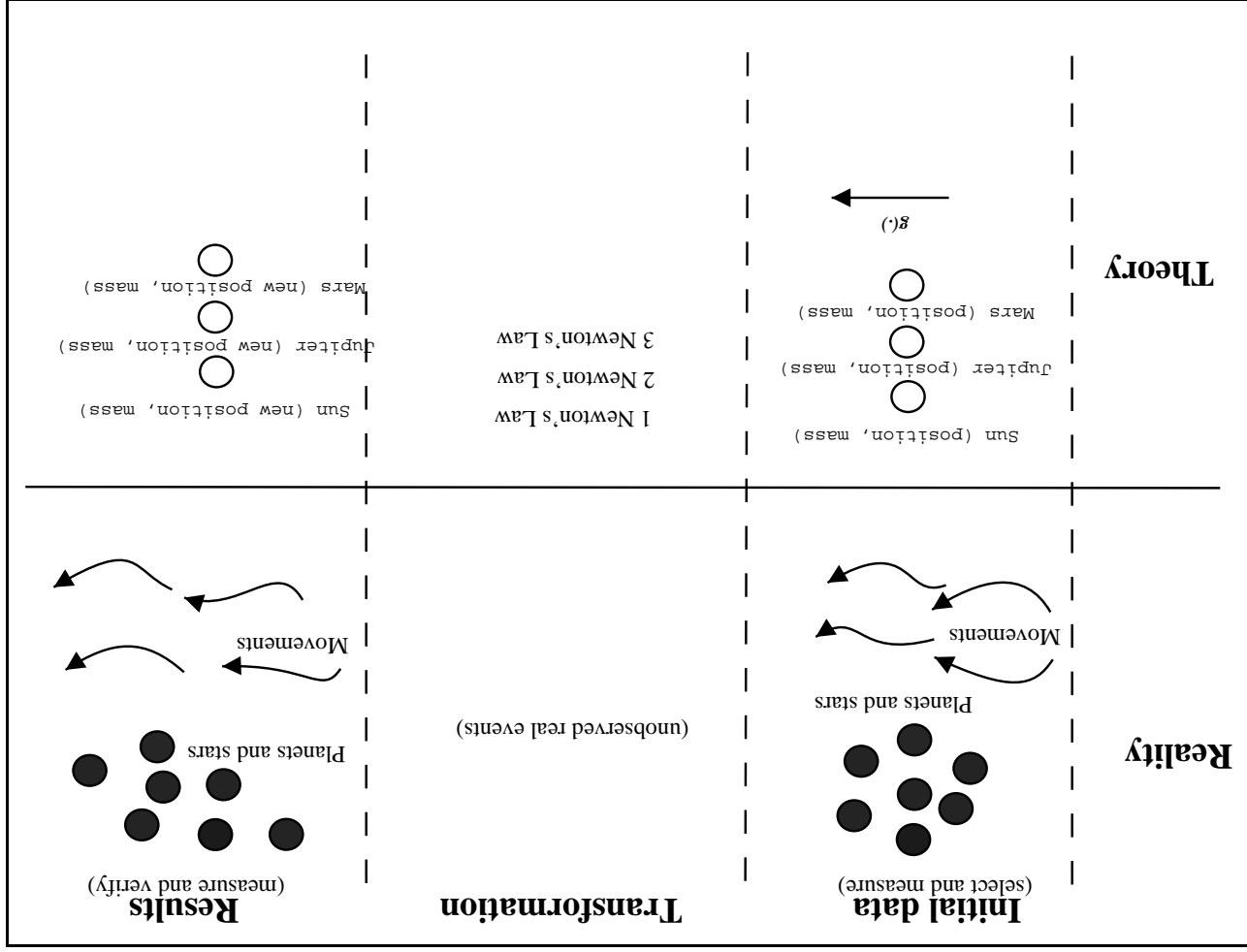
Mathematical modelling

The most successful type of modelling in science make use of mathematics. That is:

- Interpret real objects as vectors of variables;
- Define equations linking variables' values

This modelling approach produced basically all modern science, so that it is considered *the only* scientific methodology. For example, consider the classical application of planets' movements.

Example of mathematical approach



Mathematical modelling as purely quantitative

The standard scientific methodology focuses on the capacity to generate quantitative results that are robustly supported by the observation. The model is correct if it can consistently generate the same relation among the quantities observed in the real world.

Note that the mathematical models do not any claim on the realism of the generating process. The equation of the model are not “true” or not in respect of the real world. They simply work, in that they produce quantitative relations confirmed by the observation.

Weaknesses of mathematical modelling

The mathematical approach requires two properties in order to produce successful results:

- The set of variables representing objects must not change;
- The relevant events are perfectly defined by quantities.

In a number of fields, these properties are hardly shown. For example, consider a biological system and a social system.

Examples of non-mathematical systems

In a biological systems an evolving species changes the very elements under study, in a way that quantitative measures cannot fully represent.

We can measure some aspects of the *results* of the evolving system, but cannot predict or fully understand what is going on. A biological theorem cannot even be stated.

Also, in an economic system the elements under study, firms, industries, consumers etc., change their nature and behaviour such that different variables become relevant or irrelevant through time. In any case, any quantitative relation is unreliable, and subject to changes that do not depend only on strictly economic events.

Complex research fields

We can state some properties that make a system hard to treat mathematically:

- Open systems. Biological as well as economic events take place in environments that heavily influence the units under study, but yet cannot be fully modelled without making it intractable.
- Dynamical systems. When time is a relevant aspect, then it is more likely that our objects change their nature, and therefore no quantitative representation is reliable.

We can call these system **complex**. In a complex system the relevant properties stem from the interactions among elements, are very sensible to small changes of the environmental conditions, and change through time in an irreversible way.

What results on models of complex systems?

Mathematical models provide detailed quantitative results. Models of complex systems generally fail to give the same type of results.

Yet, models of complex systems can aim at reproducing some of the complexity of their real counterparts. The researchers hope that the study of the modelled complexity can help to better understand, forecast and improve real-world systems.

What results on models of complex systems?

The aim of the study (understanding, modelling) complex system consists in discovering and describing the relative influence that some forces have on particular properties of the system. Any real complex model will not confirm our pure data, since, by definition, a complex system is *open*, subject to uncontrolled influences. The theoretical analysis on complex systems can help to explain parts of the causes of some events.

The model of a complex system will systematically fail to conform to any real world observation, but will explain part of its processes.

Mathematical vs complex models

- **Mathematical models.** Given the initial values of variables the model predicts the values of the same or other variables. A model is accepted by purely quantitative assessments.
- **Complex models.** Given the dynamics of the environment and the interactions among model elements, the model partly explains how a given pattern is determined.

Using complex models

The model of a complex system cannot be limited to produce final values. It must provide explanations for all the states assumed by the model. Any observation of a real system will differ from one another, and from any model reproduction. But the model can explain what are the relevant forces at work, how a system can be made to deviate from unwanted patterns, what are the likely outcomes.

Methodology for complex models

Complex models are generally built as computer programs, i.e. computer simulation models. However, they cannot be left to programming experts, while the scientists limit their activity in providing instructions and reading final results. The reason is that a simulation models' inner workings contain the actual results the researcher is aiming for, and a technician will not be able to recognize them.

Methodology for complex models

Compare the building of a simulation program with the proving of a theorem. Once a theorem has been proved as correct, anybody can use its results without understanding the proof. So, an economist can state the enunciation of a tentative theorem, passing it to a mathematician who proves the theorem correct or false. The economist can then use the theorem in explaining economic phenomena.

Not so simulation models. Building the model requires a deep knowledge of the real system, and the very programming logic tells the scientists whether early hypothesis are coherent or not. A computer program needs to be logically and temporally coherent, and the scientists largely *learns* his/her model while writing it.

Methodology for complex models

Moreover, when a simulation code is written and a simulation run is observed, there are rarely single variables explaining what is going on. The answers to *why does this happen?* must be discovered with a lengthy process of trial and error, detailed observation of many series, data analysis, re-runs with slightly different initializations, etc. All of these requires the scientists to be with the hands on keyboard and mouse, and cannot be outsourced to a professional programmer.

Methodology for complex models

Lastly, scientific research requires necessarily the possibility to prove your results to fellow scientists in order to have your work assessed and continued. When using a simulation model, you then need to pass you expertise of the model to other people, with documentation and comments.

Methodology for complex models

In summary, using a simulation model to study a complex (i.e. open and dynamical) system requires:

- **Write your own code.** The researcher must produce the simulation program.
- **Run your own simulation.** The researcher needs to run the simulations observing the results, generating statistics, and in general making sense of series produced.
- **Packing the simulation program.** Other researchers need not only to run the program, but also understanding how it was built, and be able to exploit it to the same extent of the original author.

User requirements of a simulation language

Respecting the guidelines for simulations of complex systems may seem overly tough for people who, after all, are not interested in a PhD in computer science. But failing to comply with them seriously undermine the potential usefulness of simulation models.

In the software industry it is well known that for any 5% of code devoted to perform a given task, it is necessary to write 95% of code for user interfaces and error controls. Simulation programs are not exceptions: the scientifically relevant part of a model is generally very simple to express with any programming language; the very tough programming tasks concern the interfaces to access the model and the controls of errors.

User requirements of a simulation language

A language for non-programming experts needs to provide automatically all the interfaces that may be necessary to exploit the model. But, at the same time, need not to constrain the type of model that can be expressed with it.

Laboratory for Simulation Development (Lsd) has been built with this perspective. Lsd is not a model, but a pure language (based on C++, a very common language), so that any conceivably computational structure can be implemented with it. Lsd wraps around the C++ computational layer a complete set of interfaces required to manage simulation models according to the protocol suggested above.

Components of a simulation model

A simulation model runs through a simulated time: a “clock” creates instants of time, so that any operation in the simulation refers to a generic time step of the simulation, indicated with t .

During the simulation the model generates series of values as results of computations. Let's call a series a **Variable**, and the values of the series as produced by the **equation** of the variable. The equations elaborate with logical and mathematical operation the values from other variable, possibly from previous steps, and from **parameters**, variables producing at each time step the same value.

Components of a simulation model

For example, in a model we may have a variable X that at each time t assume the resulting value from the equation

$$X_t = \text{IF } (Y_t > 0) \text{ THEN } (X_{t-1} \times \alpha) \text{ ELSE } (X_{t-1} \times \beta), \text{ where the}$$

equation uses the same-step value of variable Y , the latest value of X itself, and two parameters α and β .

Components of a simulation model

Generally, models include many copies of some variables. For example, consider a model representing a market with many firms. All the variables representing values of the firms will be labelled with the same name, and computed with the same equation. Generally modelers use *Object Based* languages, that allow to define **Objects** as containers for set of variables, parameters, and other objects, resembling real-world entities.

For example, we may have an object *Country* containing country-specific variables and parameters, and a set of objects *Market* with market specific elements, in turn containing sets of objects *Firm*. The grouping of variables and parameters in objects is not technically necessary, but it has been shown to be a very useful approach to build simulation programs.

Components of a simulation model

To kick start a simulation we need to have the initial values for the parameters. Moreover, some equations make use of past, or *lagged*, variables values that, in the very first steps of the simulations, need to be provided by the modeller. As initial data we also need to specify how many copies of each set of objects must be included in the model (and therefore how many copies of the variables, parameters and sets of objects).

Components of a simulation model

In summary, the elements required to run a simulation model are:

- **Objects:** containers of variables, parameters and other objects for each time step
- **Variables:** names associated to equations producing one value during a simulation run
- **Parameters:** names associated to values that do not change
- **Initial data:** values for the number of objects, initial values of lagged variables and parameters