The diversity of model functions and data/models relationships in history and today

Franck Varenne – Associate Professor of Epistemology
University of Rouen (France), ERIAC (UR 4705) & IHPST (UMR 8590)
franck.varenne@univ-rouen.fr

Séminaire du Réseau ModStatSAP
ISC Paris - 16 mai 2024

Method and Claims

- Method:
  - History of science: a brief recall of old but still existing practices
  - Comparative and applied epistemology: the role of computers in the emergence of novel modeling practices

Method and Claims

- Method:
  - History of science: a brief recall of old but still existing practices
  - Comparative and applied epistemology: the role of computers in the emergence of novel modeling practices

- Claims on modeling:
  - a multifaceted and ancestral practice
  - a recent diversification of its functions due in particular to the expansion of computer simulations and to other computerized modeling techniques (e.g.: data mining, ontologies)
  - perhaps interesting for the practitioner to hear from some discriminating concepts suggested by applied epistemology of models and simulation in order to
    - have the possibility to clearly discern the kind of knowledge she/he gains through a model
    - discriminate further between the advantages and the drawbacks (limitations) of such and such type of model

Outline

- I- A short story of data/model/prediction relationships
- II- “Model” : a broad characterization for today and 5 main functions of models
- III- On 20 distinct functions of models
- IV- Limitations of models : a sample
- V- Toward the notion of “simulation”
- VI- Epistemic functions of computational models & simulations
- VII- Examples of integrative simulations in biology

Conclusions on Models and Simulations after the computational turn

I- A short story of data/model/prediction relationships

- Some definitions: model, data, theory, law, prediction

I- A short story of data/model/prediction relationships

- Some definitions: model, data, theory, law, prediction
- Epistemological hypotheses behind the practices of prediction:
  - One example: the “law of falling bodies”
  - Galilean hypotheses
  - Epistemological hypotheses behind classical approaches in mechanics
  - Epistemological hypotheses behind data analysis
  - Epistemological hypotheses behind analyses and prediction of complex systems

Some definitions

Beware : In this section, only **formal models** (related to precise prediction and data)

- A formal **model** is a formal construct possessing a kind of unity and formal homogeneity so as to conveniently satisfy a specific request : prediction, explanation, communication, decision, etc.

Some definitions

Beware : In this section, only **formal models** (related to precise prediction and data)

- A formal **model** is a formal construct possessing a kind of unity and formal homogeneity so as to conveniently satisfy a specific request : prediction, explanation, communication, decision, etc.

- The **data**: from *datum*, *do, dare* (lat.): things or properties that are given (by observation, experiment, measure), not (completely) built, not artifactual, elementary, not structured ; plural: ***data*** = a set a given things, then as such possibly structured (“atomic facts”, “observational sentences” in the positivist philosophy of science)

Some definitions

Beware : In this section, only **formal models** (related to precise prediction and data)

- A formal **model** is a formal construct possessing a kind of unity and formal homogeneity so as to conveniently satisfy a specific request : prediction, explanation, communication, decision, etc.

- The **data**: from *datum*, *do, dare* (lat.): things or properties that are given (by observation, experiment, measure), not (completely) built, not artifactual, elementary, not structured ; plural: ***data*** = a set a given things, then as such possibly structured ("atomic facts", "observational sentences" in the positivist philosophy of science)

- A **theory**: ≠ a model, a set of sentences (axioms and rules of transformation) written in a given language - be it formalized or not - that permits the translation and the derivation of a whole set of observational sentences (among them : empirical laws) about a whole domain of entities and properties

Some definitions

Beware : In this section, only **formal models** (related to precise prediction and data)

- A formal **model** is a formal construct possessing a kind of unity and formal homogeneity so as to conveniently satisfy a specific request : prediction, explanation, communication, decision, etc.

- The **data**: from *datum*, *do, dare* (lat.): things or properties that are given (by observation, experiment, measure), not (completely) built, not artifactual, elementary, not structured ; plural: ***data*** = a set a given things, then as such possibly structured ("atomic facts", "observational sentences" in the positivist philosophy of science)

- A **theory**: ≠ a model, a set of sentences (axioms and rules of transformation) written in a given language - be it formalized or not - that permits the translation and the derivation of a whole set of observational sentences (among them : empirical laws) about a whole domain of entities and properties

- A **law**: ≠ a model, a constant, universal and necessary (or statistically significant) relationship between properties
  - U = R.I - Ohm's law
  - P.V = n.R.T – Boyle-Mariotte's law
  - A law can be induced by experiments and deduced by a theory
  - A law can be approximated or sketched by a kind of model (belonging to the category of *intelligible presentations* : model functions # 7, 8 or 9, *i.e.* phenomenological, explanative or comprehensive model)

Some definitions

Beware : In this section, only **formal models** (related to precise prediction and data)

- A formal **model** is a formal construct possessing a kind of unity and formal homogeneity so as to conveniently satisfy a specific request : prediction, explanation, communication, decision, etc.

- The **data**: from *datum*, *do, dare* (lat.): things or properties that are given (by observation, experiment, measure), not (completely) built, not artifactual, elementary, not structured ; plural: ***data*** = a set a given things, then as such possibly structured ("atomic facts", "observational sentences" in the positivist philosophy of science)

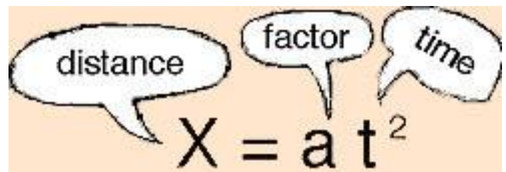
- A **theory**: ≠ a model, a set of sentences (axioms and rules of transformation) written in a given language - be it formalized or not - that permits the translation and the derivation of a whole set of observational sentences (among them : empirical laws) about a whole domain of entities and properties

- A **law**: ≠ a model, a constant, universal and necessary (or statistically significant) relationship between properties
    - U = R.I - Ohm's law
    - P.V = n.R.T – Boyle-Mariotte's law
  - A law can be induced by experiments and deduced by a theory
  - A law can be approximated or sketched by a kind of model (belonging to the category of *intelligible presentations* : model functions # 7, 8 or 9, *i.e.* phenomenological, explanative or comprehensive model)

- A **prediction**: the derivation of some observational sentences from the application of a law (or a model approximating a law) to a particular situation or to particular initial conditions.
  - In the positivist perspective (e.g. the Deductive-Nomologic model of explanation after Hempel), prediction is only a kind of explanation: an explanation of the future
  - The predictive law can have a statistical form

Epistemological hypotheses behind the practice of prediction: the “law of falling bodies”

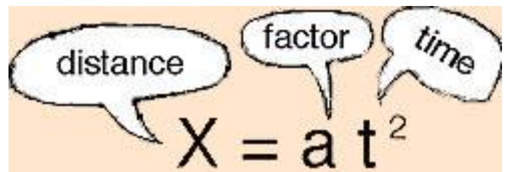
- Question: What did Galileo newly achieve with his mathematical « law of falling bodies » (1638)?



Source : Virtual Museum of Science, Technology and Culture – Tel-Aviv

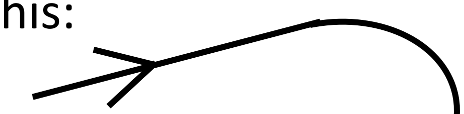
Epistemological hypotheses behind the practice of prediction: the "law of falling bodies"

- Question: What did Galileo newly achieve with his mathematical « law of falling bodies » (1638)?



Source : Virtual Museum of Science, Technology and Culture – Tel-Aviv

- Answer: The novel and precise conjunction of explanation **and** prediction (Koyré, Crombie, Clavelin, Hempel)
- Before Galileo (Antiquity, Middle Age) :
  - engineers were conceiving and using catapults and bombards with the help of empirical knowledge summarized in some tables or abaci to approximate the parabolic form of the movements made by projectiles
  - whereas philosophers still taught that movement could only be linear or circular or a succession of the two like this:



Galilean Hypotheses

- 1. Nature is regular and constant
- 2. It is “written in mathematical language” (1623)
- 3. Hence “Laws of Nature” exist that permit prediction
- 4. In mechanics, processes (“causal chains”) are additive then separable: abstract laws permit prediction (the friction or resistance effect of the liquid or the gas surrounding the falling body easily can be subtracted or added to the gravity effect)

Epistemological hypotheses behind classical approaches

- Regularity of nature, thanks to:
  - Causes: essentialism
  - Forces (dispositions, propensities): realism of forces (Newton, f=m.a)
  - Laws: positivism (Comte, Mach, Hempel)
  - Mechanisms (in biology, social sciences, Machamer *et al*. 2000): a current positivist conception of forces or dispositions or propensities (objective probabilities)
- Contestations on "data": theory ladenness of data (Hobbes, Hanson, Kuhn), properties of the subject, not of the world in itself (Kant), bias (psychology)

Epistemological hypotheses behind multivariate analysis (R.A. Fisher, 1922)

"On the mathematical foundations of theoretical statistics", Phil. Trans. A

- **Variability** in living beings (biometrics), **non nominality of data** ≠ mechanics
- **Intricacy** of causal chains, **non additivity** ≠ mechanics
- As Laplace and contrary to Pearson, Fisher does not reject causes
- But he critics naïve bayesian approaches: **bayesians mistake the *hypothetical population* with the *sample extracted from this population***
- **Against this mistake:**
    - Introduction of an intermediary construct: the **"hypothetical law"** (1922) based on a frequentist interpretation of probabilities (*i.e.* probability is the asymptote of an experimental frequency).
    - The "design of experiment" (1924): controlled/uncontrolled factors, randomisation
    - Statistical analysis = comparison of the "hypothetical law" with the null hypothesis
    - As a consequence:

        - infinite is necessary to handle, then fictions too: statistical models seen as a "filter for information" (fn #5)

        - we are right to assume that chance is structured (Central Limit Theorem → a priori normal distribution) and that, followingly, estimation of predictor paremeters is possible (≠ Rare Events, Black Swans, cf. Taleb, 2007)
    - Hypotheses: "causal matrix" (Fisher), multivariate analysis, (multi)linear algebraic approach, parametric inferential statistics

The cycles of oppositions and mergings between empirico-inductivism and hypothetico-deductivism in data analysis

- **Empirico-Inductivism (EI)** (J.P. Benzécri in France, Geometrical Analysis of Data)
  - Data first, observation first: they speak by themselves
  - Data freely show a pattern that suggests a law that permit to predict other data
- **Hypothetico-Deductivism (HD)** (J. Neyman)
  - No data without a firstly conceptualized and theorized format of data (unity, range,..)
  - No experiment without the test of a hypothetical pattern (even implicitly)
- **Multivariate analysis** (Fisher, etc.) **merges the two:**
  - You have to make assumption on the structure of the hypothetical infinite population of your data (hypothetico-deductivism/modelism)…
  - …in order to have access to a structure of data in the less constraining way (inductivism)

The cycles of oppositions and mergings between empirico-inductivism and hypothetico-deductivism in data analysis

- **Empirico-Inductivism (EI)** (J.P. Benzécri in France, Geometrical Analysis of Data)
  - Data first, observation first: they speak by themselves
  - Data freely show a pattern that suggests a law that permit to predict other data
- **Hypothetico-Deductivism (HD)** (J. Neyman)
  - No data without a firstly conceptualized and theorized format of data (unity, range,..)
  - No experiment without the test of a hypothetical pattern (even implicitly)
- **Multivariate analysis** (Fisher, etc.) **merges the two:**
  - You have to make assumption on the structure of the hypothetical infinite population of your data (hypothetico-deductivism/modelism)…
  - …in order to have access to a structure of data in the less constraining way (inductivism)
- **Today cycling objections:**
  - **EI oriented objection:** such a "parametric estimation" still can be seen as a too constraining and deforming access to data (too much HD) : non parametric statistical inference is a solution (but see Denis & Varenne, 2022, p. 293)
  - **HD oriented objection:** both non parametric and parametric statistical inference don't give concepts, understandings, real knowledge in this sense
  - **IE oriented answer:** 1) weak : this is not the role of data mining ; 2) strong : on the contrary, GA or neural networks approaches can be seen as more realistic (in the connectionist/anti-symbolic approach of intelligence and AI)
  - **HD oriented reply:** Most data mining programs/courses seem to promise what they cannot give : a gain in knowledge. Cf. *e.g.* Daniel T. Larose (2005): ***Discovering Knowledge in Data – An Introduction to Data Mining***

Revue Ouverte d'Intelligence Artificielle

CHRISTOPHE DENIS, FRANCK VARENNE
Interprétabilité et explicabilité de phénomènes prédits par de l'apprentissage machine

Volume 3, nº 3-4 (2022), p. 287-310.

<http://roia.centre-mersenne.org/item?id=ROIA_2022__3_3-4_287_0>

“Le caractère agnostique du modèle d’apprentissage ne signifie toutefois pas qu’un tel modèle ne repose sur aucune hypothèse mathématique préalable. Comme le montre S. Shalev-Schwartz & S. Ben-David (*Understanding Machine Learning : From Theory to Algorithms*, Cambridge University Press, 2014), les modèles d’apprentissage agnostiques de type ≪ PAC ≫ (probably approximately correct), par exemple, reposent sur **des hypothèses mathématiques minimales et nécessaires** comme l’hypothèse de plongée préalable du domaine d’objets dans une sigma algèbre (ou tribu*), munie des bonnes propriétés qui nous sont nécessaires (en termes de théorie de la mesure), ainsi que **l’hypothèse de définissabilité d’une distribution** quelconque sur ce domaine”, Denis, Varenne, ROIA, 2022, p. 293

* « une tribu sur un ensemble X est un ensemble non vide de parties de X, stable par passage au complémentaire et par union dénombrable (donc aussi par intersection dénombrable) », Wikipedia

Revue Ouverte d'Intelligence Artificielle

Christophe Denis, Franck Varenne
Interprétabilité et explicabilité de phénomènes prédits par de l'apprentissage machine

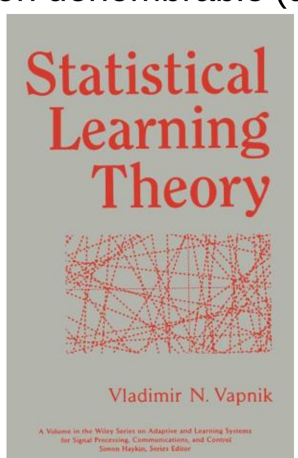
Volume 3, nº 3-4 (2022), p. 287-310.

<http://roia.centre-mersenne.org/item?id=ROIA_2022__3_3-4_287_0>

"Le caractère agnostique du modèle d'apprentissage ne signifie toutefois pas qu'un tel modèle ne repose sur aucune hypothèse mathématique préalable. Comme le montre S. Shalev-Schwartz & S. Ben-David (*Understanding Machine Learning : From Theory to Algorithms*, Cambridge University Press, 2014), les modèles d'apprentissage agnostiques de type ≪ PAC ≫ (probably approximately correct), par exemple, reposent sur **des hypothèses mathématiques minimales et nécessaires** comme l'hypothèse de plongée préalable du domaine d'objets dans une sigma algèbre (ou tribu*), munie des bonnes propriétés qui nous sont nécessaires (en termes de théorie de la mesure), ainsi que **l'hypothèse de définissabilité d'une distribution** quelconque sur ce domaine", Denis, Varenne, ROIA, 2022, p. 293

* « une tribu sur un ensemble X est un ensemble non vide de parties de X, stable par passage au complémentaire et par union dénombrable (donc aussi par intersection dénombrable) », Wikipedia

Voir aussi Vapnik, 2005 :



Epistemological hypotheses behind analyses and prediction of complex systems

- Negative statements:
  - Non additivity
  - Non linearity
  - Deterministic chaos: even a good theory can be uncomputable for a precise prediction
  - Emergence
  - Downward or more generally inter level causations
  - Axiomatic Heterogeneity of aspectual models or laws interacting in the same target system (Varenne, 2010, 2013, 2018)

Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy

Epistemological hypotheses behind analyses and prediction of complex systems

- Negative statements:
  - Non additivity
  - Non linearity
  - Deterministic chaos: even a good theory can be uncomputable for a precise prediction
  - Emergence
  - Downward or more generally inter level causations
  - Axiomatic Heterogeneity of aspectual models or laws interacting in the same target system (Varenne, 2010, 2013, 2018)

- Strategies:
  - Focus on bifurcations instead of trajectories
  - Instead of the *form of trajectories,* focus *on* the *structure of the phase space* (= abstract space of dynamic variables)
  - Tackle some generic complex behavior through mathematical tools (dynamic systems) or computational tools (exotic statistical physics)
  - Don't predict but think in probability and through projections of weighted scenarios (IPCC)
  - More and more : simulations comes first, then model (Varenne, 2007). There is an increasing search for models on realistic simulations. Simulations are no more subsidiary for models. Since the 1990's, it tends to be the contrary.
  - The captation and stabilization of experimental data in some complicated or even complex systems of simulations:
    - Multi-scale pluri-axiomatized/pluri-formalized simulations: e.g. in the complex *Functional Structural Plant Models* of AMAP (CIRAD/INRA/INRIA), (Varenne, 2007)
    - Phenomenological reconstruction: a multidisciplinary well informed task of intertwining empirical, interpretative and theoretical approach of a multi-physics, multi-aspectual biological system (Olivier, Peyriéras et al. 2010 in *Science* : Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy
    - Blue Brain project !
  - The search for ontologies

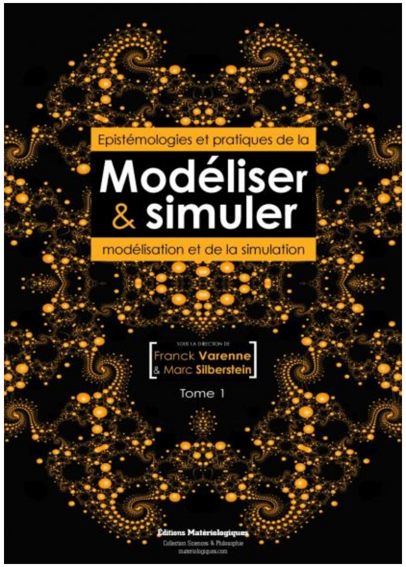
II-A- « Model »: a broad and pragmatic characterization (source: Varenne, 2013, 2018)

- **Minsky (1965) :** "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"

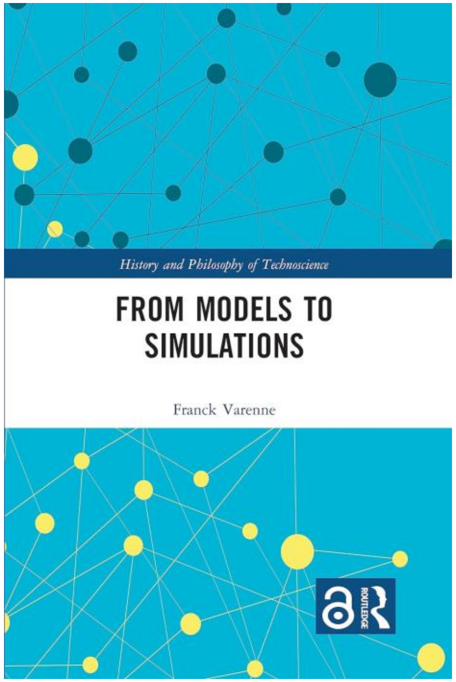
- **Consequences:**
    - Not necessarily a representation
    - A double relativity: 1st to the observer, 2nd to the question asked by the observer
    - The model is an "object" : i.e. an entity with ontological independence and autonomy.
        - It is not only a linguistic product, nor an expression nor a metaphor
    - Its general function in sciences : to facilitate a mediation in the context of a cognitive questioning, a cognitive inquiry (practical cognition or theoretical cognition)

II-B- The 5 Main Functions of Models

(sources: Varenne, 2013, 2018)



Modéliser & simuler. Epistémologies et pratiques de la modélisation et de la simulation, F. Varenne, M. Silberstein, Matériologiques, 2013. Sur Cairn : https://www.cairn.info/modeliser-et-simuler-tome-1--9782373612707.htm



From Models to Simulations, Routledge, 2018, free online: https://doi.org/10.4324/9781315159904

II-B- The 5 Main Functions of Models

(source: Varenne, 2013, 2018)

- I- to facilitate an observation or, more generally, an experiment (observation + controlled interaction)

- II- to facilitate the coining of an intelligible representation

- III- to facilitate theorization

- IV- to facilitate the mediation between discourses and disciplines

- V- to facilitate decision and action

III- On 20 functions of models (1/5)

Source : Varenne (2013, 2018)

- **I- to facilitate an observation or an experiment:**

  - 1) To make sensible (1:1 scale mockup of the human body (wax anatomical and surgical model), solar system with balls,…)

  - 2) To make memorizable (simplifying pedagogical diagrams,…)

  - 3) To facilitate experiments through indirect experiments on **experimental models**: experiment on real or imaginary (thought experiments) objects or on experimental living models (Arabidopsis, E. coli, drosophila, pigs, mice, rabbits, monkeys…)
    - That are easily accessible for material (space, time, number, rapidity), financial (cost), technical (availability in a given context), ethical or deontological reasons
    - Pb of model validity: specificity; face validity (same pheno, symptoms) ; construct validity (homology, same cause for sickness: e.g. virus inoculation. Source: Depaulis, INSERM, 2012)

  - 4) To facilitate the presentation, through its abbreviation, of the data and controlled variables of an experiment (not the representation of the experimented object): *e.g.* through a statistical “model of data” (statistics, data analysis, models of analysis) (in France: J.P. Benzécri: strong inductivism). Summarize data : Principal Comp Analysis

III- On 20 functions of models (2/5)

Source : Varenne (2013, 2018)

- **II- to facilitate an intelligible presentation through a mental representation or a conceptualization:**
  - 5) To facilitate the **compression of data** to build a first kind of conceptualization through "models of data".
    Question: Are the statistical moments of data conceptual constructs (HD) or real properties of data (EI)?
  - 6) To facilitate the **selection and classification of** relevant entities and properties in a given domain: conceptual models, models of knowledge, ontologies (see the growing number of ontologies in integrative or systemic biology: not to be confused with theories!)
  - 7) To facilitate the **reproduction or production of an observable dynamic**: phenomenological models, face value model, **predictive models** (parallel dynamics without explanative factor: no mechanisms). E.g.: a fitted polynomial equation (see: "instrumentalism" on models by Friedman, 1952), "classification predictive models", "regression predictive models", Finlay, 2014, p. 10. See "models of decision" #19-20 (belong to "predictive models" of classification)
  - 8) To facilitate the **explanation** of a phenomenon though the visualization and reproduction of mechanisms of elementary interactions : explanative models, mechanistic models
    - e.g.: mechanistic models (pump for the heart): communication of movements through contacts
    - electrical model for the biological neuron (Hodgkin-Huxley 1952): transmission of signals ;
    - energetic models for the nutrition (Biological cycles…): communication, transport and transformation of energies
    - individual-based models for diffusion phenomena or phenomena arising in assemblies of cells, organs or organisms: in ecology, epidemiology, endocrinology, formal neural network models for biological neural networks… : transportation of matters, energies and signals
    - computational models
  - 9) To facilitate the **comprehension** of a phenomenon by formulating the general principles that rule a dynamic that looks like the observed one: theoretical models. Examples:
    - cybernetic models, robots, animats (behavior-based robots)
    - topological models of morphogenesis (Thom, Zeeman, catastrophe theory)
    - thermodynamics of open systems, bifurcation theory (Prigogine), synergetics (Haken)
    - fractals (Mandelbrot);autopoiesis (Maturana, Varela)
    - …

III- On 20 functions of models (3/5)

Source : Varenne (2013, 2018)

- **III- To facilitate a theorization**
    - A theory ≠ A model
    - A theory : ***a set of sentences (axioms and principles/rules of transformation) written in a given language - be it formalized or not – that permits the translation and the derivation of a whole set of observational sentences (among them : empirical laws) about a whole domain of entities and properties***
    - Subsidiary question: are there genuine theories (fundamental laws) in biology and not only mechanisms ? See Canguilhem 1963.
  - 10) To facilitate the building of a still not mature theory: first formulation of regularities and derivations, but not founded on proper principles and axioms
  - 11) To interpret a theory, to show its visualizability in terms of mental images or of thought experiments (Boltzmann)
  - 12) To illustrate a given theory by another one (Maxwell): search for mathematical analogies (geometrical models) to facilitate the calculus
  - 13) To test the inner coherence of a given theory (link with the mathematical theory of models in logic and mathematics
  - 14) To facilitate the application of the theory, *i.e.* its reconnection with the data. E.g.:
    - An intermediate model between the theory and the model of data (rules of correspondence).
    - Approximate heuristic models or asymptotical models of Navier-Stokes equations in fluid dynamics
  - 15) To facilitate the hybridation and co-calculation of heterogeneous theories
    - e.g.: models of polyphase systems in physico-chemistry or chemistry of combustion (liquid / solid / gaseous phase).

III- On 20 functions of models (4/5)

Source : Varenne (2013, 2018)

- **IV- To facilitate the mediation between discourses about a complex - in the sense of multidimensional - phenomenon (to facilitate the formulation of the questioning not of the formulation of an hypothetical answer):**

  – 16) to facilitate communication between disciplines and researchers (data-base sharing)

  – 17) to facilitate deliberation and dialogue.

    E.g. : in environmental sciences RAINS models (Kieken, 2004) on acid rains

  – 18) to facilitate the co-construction of the management of mix systems, of socio-natural systems.

    E.g.: companion modeling (CIRAD), participative modeling, interactive modeling of agronomic and agricultural systems

III- On 20 functions of models (5/5)

Source : Varenne (2013, 2018)

- **V – To facilitate the formulation not of the questioning nor of the answer to the question but of the final decision only, i.e. to help determinate the kind of possible pre-established actions (to vaccinate or not, to buy or not…):**
  - 19) To facilitate a rapid decision in a complex context of emergency.
    - E.g. model for the management of epidemics, model for the management of catastrophe
  - 20) To facilitate a decision in a context where models of decision finally can be counted as explanative too because, on the run, they are becoming not only descriptive but prescriptive (in that they act as self-fulfilling representations). Confusions with functions #8 and #9 often.
    - E.g.: models of decision in psychology, theory of decision, economics, mathematical models of derived products in finance (MacKenzie 2004)(Aglietta 2008)

IV-Limitations of models : a sample

- Validation : variety of procedures
- Range of their validity: a quasi-circular problem. A few remarks.
  - There exist no formal nor general theory nor model of what it implies for a model to be a good one. Methodology, know-how, art (epistemology) ?
  - Induction must be reinforced, multiplied, multiplexed
  - See: multi-aspectual models, multiscale models, integrative models
  - See the rise of so-called “cross-validation” techniques

IV-Limitations of models : a sample

- Validation : variety of procedures
- Range of their validity: a quasi-circular problem. A few remarks.
  - There exist no formal nor general theory nor model of what it implies for a model to be a good one. Methodology, know-how, art (epistemology) ?
  - Induction must be reinforced, multiplied, multiplexed
  - See: multi-aspectual models, multiscale models, integrative models
  - See the rise of so-called "cross-validation" techniques

- Does a model always have to be simple?
  - No.
  - For a very long time a confusion has been done between the role of facilitation the model must have and its supposed property to be simple
  - In fact, it depends on its avowed function: theoretical or empirical, …?
  - This is the recent rise of "simulations" and "computational models" that has permitted to make the difference and to reveal the traditional confusion (Varenne, 2007, 2008)
  - Let's have a look now on recent trends due to the spreading of simulations

V- Toward the notion of simulation (1/5)

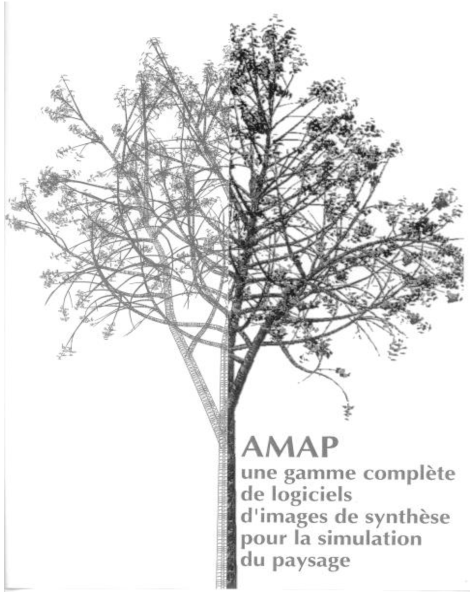
Sources of this section : Phan & Varenne, 2010 ; Varenne, 2010 ; Varenne, 2013

- Computer simulations depend on formal models (helps to solve, calculate, validate)
- A formal **model** is a formal construct possessing a kind of unity and formal homogeneity so as to satisfy a specific request : prediction, explanation, communication, decision, computability, etc.
- Concerning **simulation**, traditional definitions need to be generalized.
- It is often said that “a simulation is a model in time”, a ”process that mimics the (supposed to be the more) relevant characteristics of a target process”, Hartmann (1996). But consider:
  - The **variety** of types of contemporary CSs.
  - Today, CSs rarely are the dynamic evolution of **a single formal model**.
  - CSs in the sciences of complex objects are most of the time CSs of complex ***systems of models***.
  - Moreover, there exist ***various kinds of CSs*** of the same model or of the same system of models.

V- Toward the notion of simulation (2/5)

- Last but not least, the criterion of the “temporal mimicry” is in crisis too: it is not always true that the dynamic aspect of the simulation imitates the temporal aspect of the target system. Some CSs can be said to be *mimetic in their results but non-mimetic in their trajectory* (Varenne, 2007) (Winsberg 2008).

- For instance, it is possible to simulate the growth of a botanical plant sequentially and branch by branch (through a non-mimetic trajectory) and not through a realistic parallelism, i.e. burgeon by burgeon (through a mimetic trajectory), and to obtain the same resulting and imitating image (Varenne 2007).



Source : Simulated Poplar - Plant Architecture Modelling Laboratory (CIRAD/France)

V- Toward the notion of simulation (3/5)

- The problem: the temporal aspect is itself dependent on the persistent - but vague - notion of imitation or similitude.
- But, in fact, it is possible to give a minimal characterization of a CS (not a definition) referring neither to an absolute similitude (formal or material) nor to a dynamical model.
- Let's say that **a simulation is minimally characterized by a strategy of symbolization taking the form of at least one step by step treatment.** This step by step treatment proceeds in two major phases:
  - **1st phase (operational phase):** a certain amount of operations running on symbolic entities (taken as such) which are supposed to denote either real or fictional entities, reified rules, etc.
  - **2nd phase (observational phase):** an observation or a measure or any mathematical or computational re-use of the result of this amount of operations taken as given through a visualizing display or a statistical treatment or any kind of external or internal evaluations.
  - e.g., in some CSs, the simulated "data" are taken as genuine data for a model or another simulation, etc.

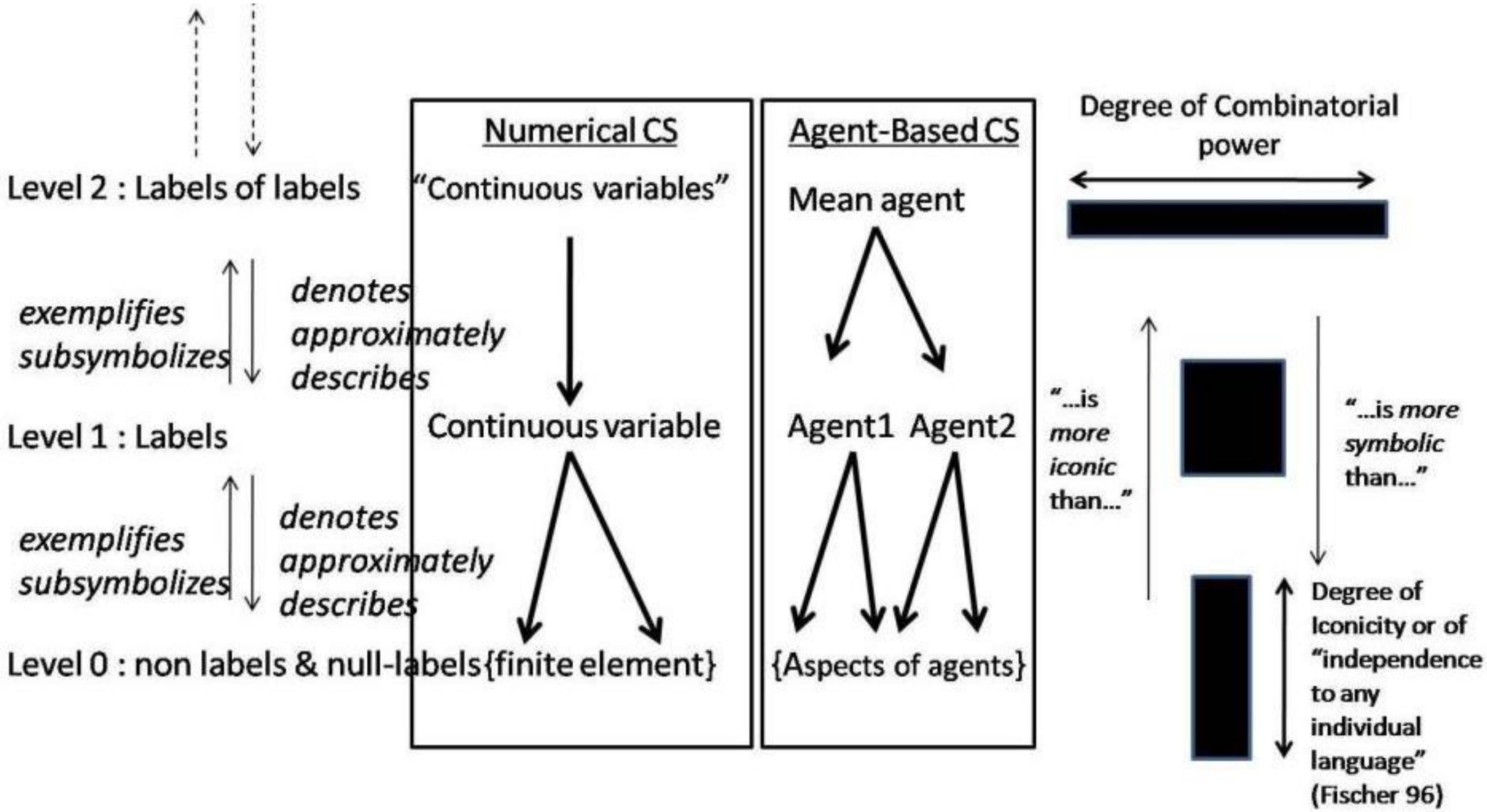
V- Toward the notion of Simulation (4/5)

Sub-symbolhood in computer simulations

- Concerning the two phases in simulation (operative, observational):
  - During the observational phase, marks which were first treated as genuine symbols, i.e. as denoting entities, are finally treated as *sub-symbols*: Why? They are treated at another level at the one they first operated.
  - At the end of process, it is the result observed - as a whole - which gains a proper and new symbolic nature
  - And this is relatively to this new symbol or system of symbols that the first symbols become sub-symbols.
  - Let's recall that, according to (Smolensky 1988), subsymbols operate in a connectionist network at a lower level than the symbols. As such, they can be seen as constituents of symbols.
    - Subsymbols "participate in numerical – not symbolic – computation": the kinds of operation on symbols (computations) are not the same at each level.

V- Toward the notion of simulation (5/5)

Simulations and hierarchies of symbols



- We can draw a parallel between the hierarchy of levels of symbols in a symbols' hierarchy and the similar hierarchies in numerical simulations and in agent-based simulations.
- The relation of subsymbolization can be interpreted in terms of an exemplification whereas the relation of denotation can be interpreted in terms of an approximate description.

VI- Epistemic functions of M&S (1/5)

Simulations of Models

- According to (Ören 2005) & (Yilmaz *et al*. 2006), **“simulation has two different meanings:**
  - **(a) imitation of a target system and**
  - **(b) goal-directed experimentation with dynamic models”**

- The previous conceptual analyses confirm and explain further this matter of fact:
  - First. We are right to say that a computer simulation is a “simulation of a model” when its specific strategy of subsymbolization essentially is taken as a strategy of *subsymbolizing* the dynamic of the model.
    - From this viewpoint, a lapse of time taken in the dynamic of the model is *iconically denoted* by a lapse of time of computation in the CS. An iconic semiotic relation takes place here because *a lapse of time is denoted through another lapse of time*.
    - This iconic relation is not an “imitation” of a property of a target system but an imitation of an aspect of the time-consuming dynamic of the model by a time-consuming activity: a computation.
    - This hidden imitation is what permits to characterize the second meaning of “simulation” - according to (Yilmaz *et al*. 2006) - as a kind of *experimentation* (on a model or system of models).

VI- Epistemic functions of M&S (2/5)

Simulations of Target Systems

- Second. A CS can be called a simulation for another reason:
  - It can be seen as *a direct simulation of an external target system* and not as a simulation of model.
  - Here, we find what (Yilmaz *et al.* 2006) call the first meaning of simulation: *imitation*.
  - In this case, it is implicitly assumed that symbols at stake in the simulations are entering in some direct iconic relations to some external properties of the external target objects.
  - From this viewpoint, contrary to what prevails in the last case, external relations between symbols and target entities or target symbols or labels have to be taken into account.

VI- Epistemic functions of M&S (3/5)

Models as experiments

- **To what extent models can be seen as some kind of experiment?**

- CA models have an empirical aspect because some causal factors are denoted through symbols of which partial iconicity is patent and can be reasonably recognized as a sufficiently "realistic" conjecture in the argumentative approach of the domain specialist, i.e. the biologist (e.g.: a formal cell = a biological cell).

- On the contrary, models are seen from an instrumentalist standpoint when the level of iconicity of their symbols is weak (the remoteness of reference is important) and when this is their combinatorial power at a high level in the denotational hierarchy which is requested

VI- Epistemic functions of M&S (4/5)

Computer Simulation as experiment ON a model

- **How and why can a CS be seen as an experiment on a model?**
- Because of the analogy between any subsymbolization (such as any change of level in the DH) and the canonic relations between the formal and the material stuffs in the empirical sciences, a CS can be said to be an experiment on the model or on the system of models
- But if we focus on some residual symbolic aspects of the subsymbols at stake, we can speak of such a CS of model as a conceptual exploration. And it is because we put the stress on the residual combinatorial /symbolic power of the subsymbols that we see such a CS as a delegation of an intellectual /conceptual practice.
- It follows that external validity is a matter of degree and depends on the strength of the alleged iconic aspects. If this iconic aspect is extremely stabilized and characterized, the simulation can even be compared to an exemplification. In this case, external validity is not far from an internal one.

VI- Epistemic functions of M&S (5/5)

Computer Simulation as an experiment in itself

- **To what extent a CS can be seen as an experiment in itself?**

- A CS can first lend its empiricity from an experiencing, that is, from a comparison with the target (external validity): and those are (1) the empiricity regarding the causes (of the computation) and (2) the empiricity regarding the effects (of the computation).

- But its empiricity can be decided not from an experiencing but from an experimenting on the interaction between levels of symbols, i.e. between controlled and uncontrolled changing factors: and here are (1) the empiricity regarding the intrication of the denotational routes (interactions of mechanisms), and (2) the empiricity regarding the defect of any a priori epistemic status.

- Through this particular experimenting dimension, **software-based CS** (in **integrative biology** for instance) gain some particular kind of empiricity which seems to give them a similar **epistemic power** than ordinary experiments.

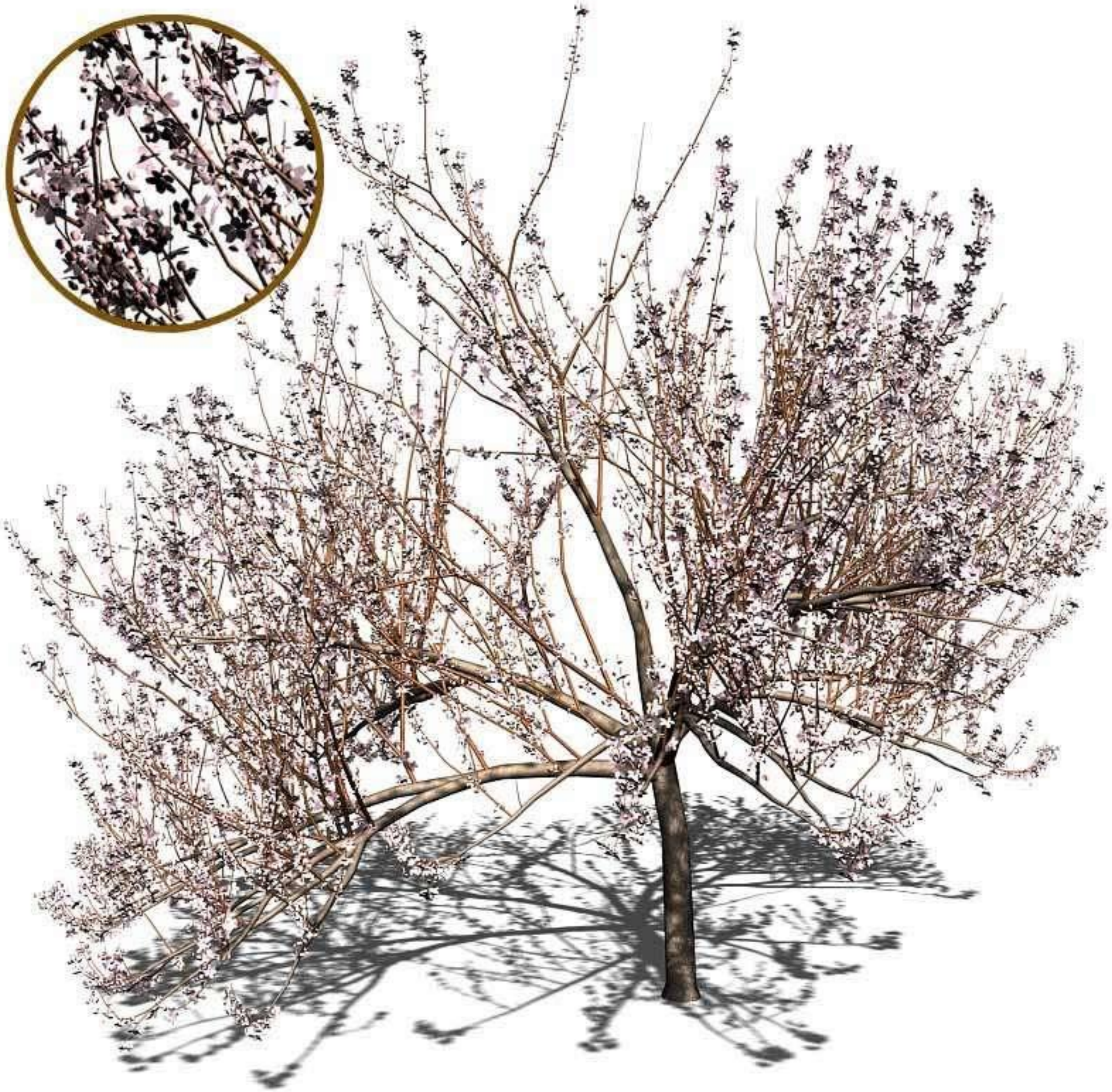
VII- Examples of integrative simulations in biology

And some remarks on their methodological and epistemological consequences

Aulne - Source : Bionatics (http://www.bionatics.com)

Rapidly growing tree mature at about 60 years with long trunk and narrow crown. Distinctive outline in winter. Height 20m or more.

Accacia Lahia - Source : Bionatics (http://www.bionatics.com)
A perennial flat-topped species of tree found in Africa.

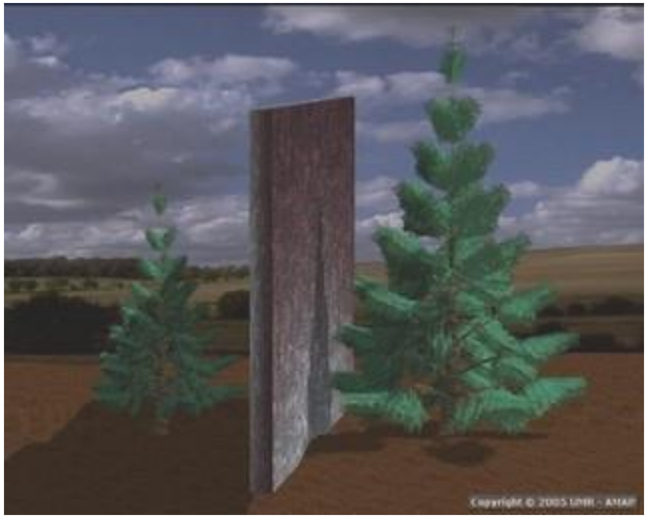


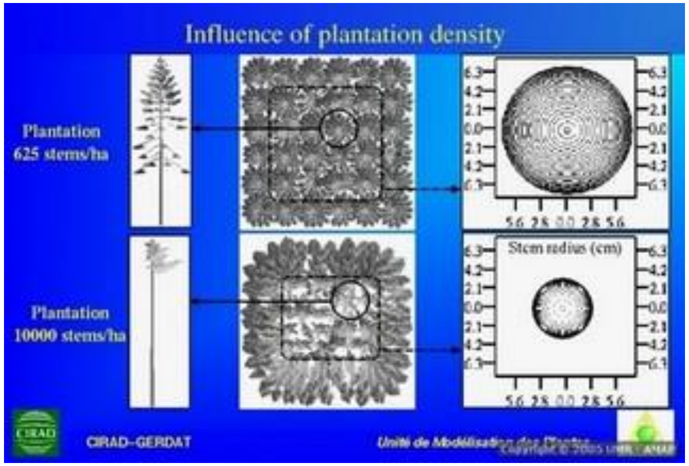
Abricotier japonais - Source : Bionatics (http://www.bionatics.com)
Low spreading tree with pink flowers in spring.

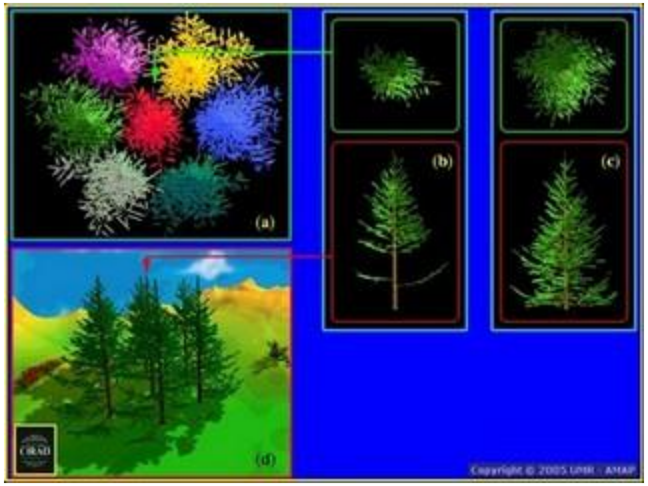
What for ?

Application in architecture - Bionatics : http://www.bionatics.com

Internal and External Interactions

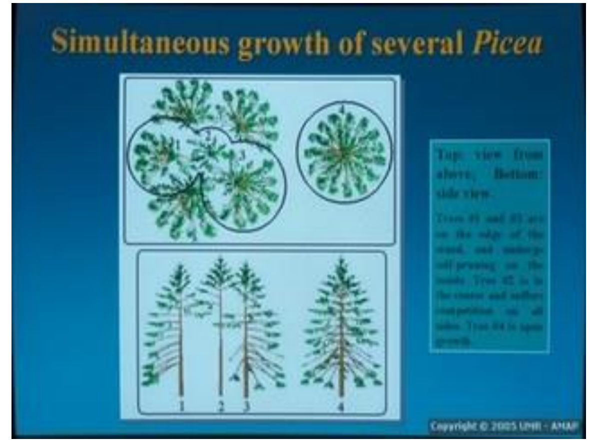


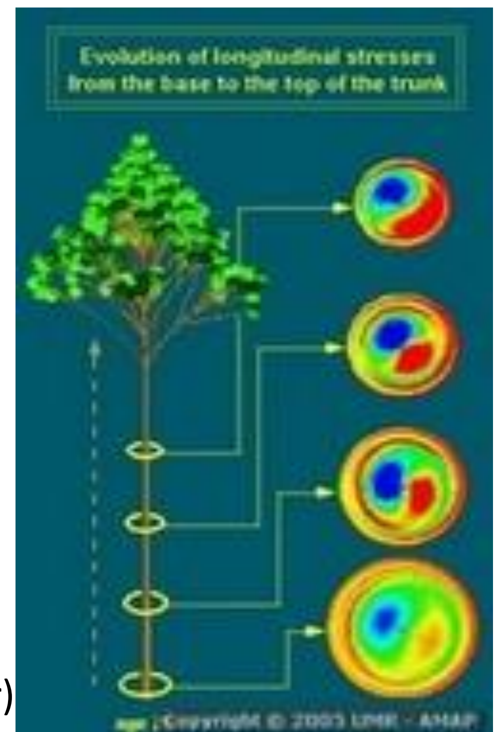




Source: AMAP (CIRAD, INRIA, INRA, IRD, CNRS, Montpellier)

Interactions, flexions, mechanical constraints → prediction of wood quality



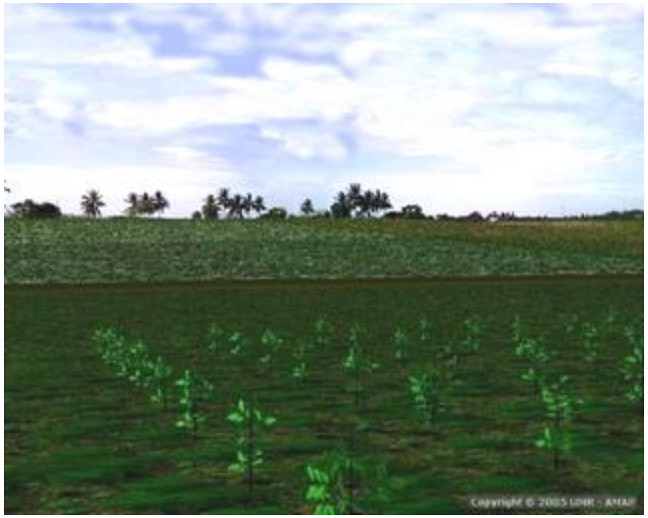


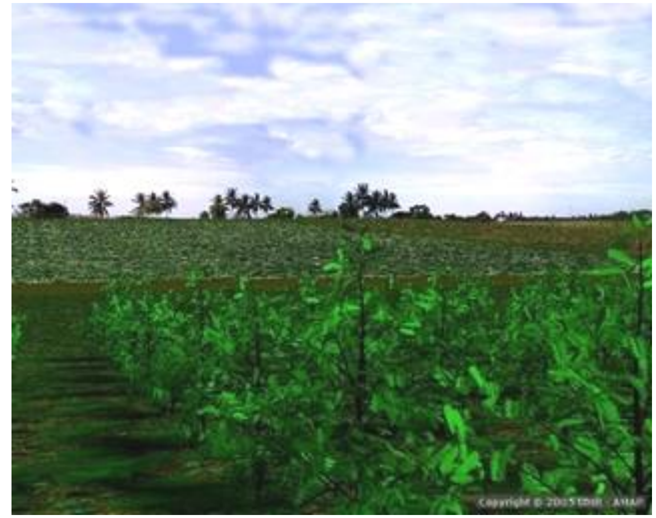
Source: AMAP (CIRAD, INRIA, INRA, IRD, CNRS, Montpellier)

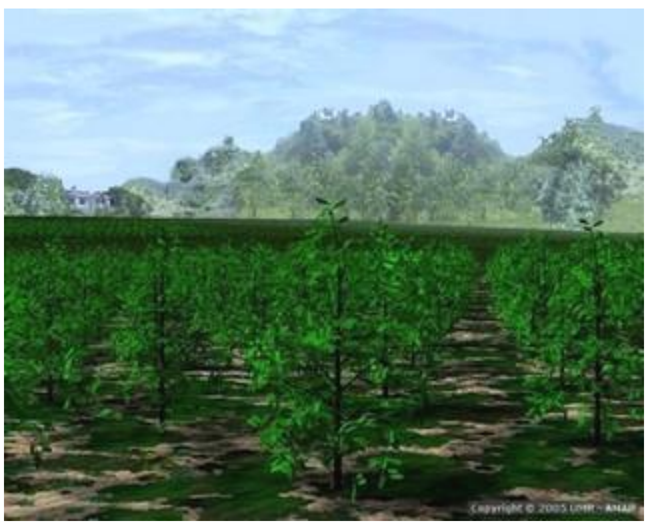
Applications in predictive agronomy : coffee, corn, …

Source : Philippe de Reffye (Digiplante-Inria-ECP-INRA, Amap Cirad)

Applications in predictive agronomy







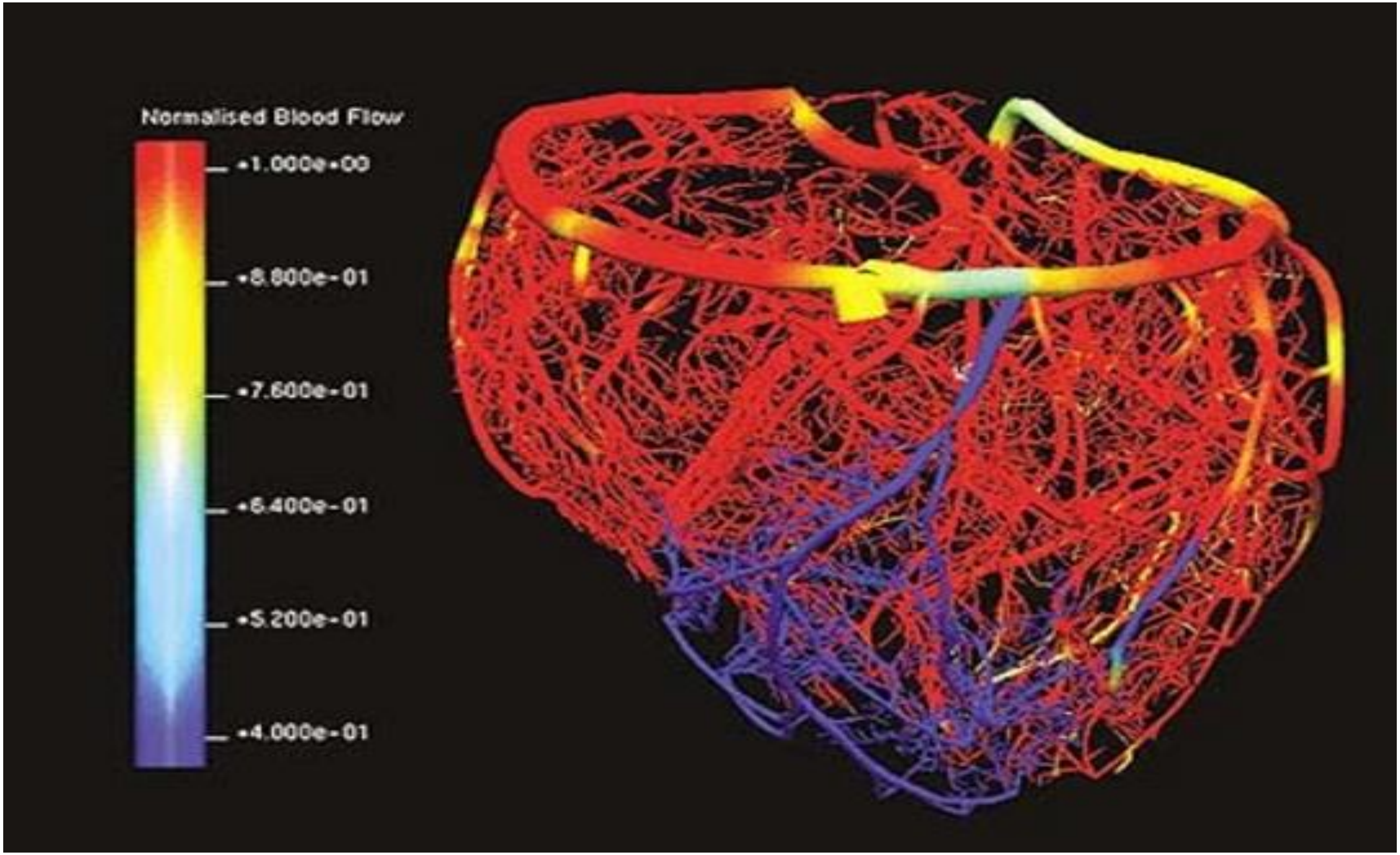
Source: AMAP (CIRAD, INRIA, INRA, IRD, CNRS, Montpellier)

Applications in urbanism…

Source: AMAP (CIRAD, INRIA, INRA, IRD, CNRS, Montpellier)

Application in paysagism

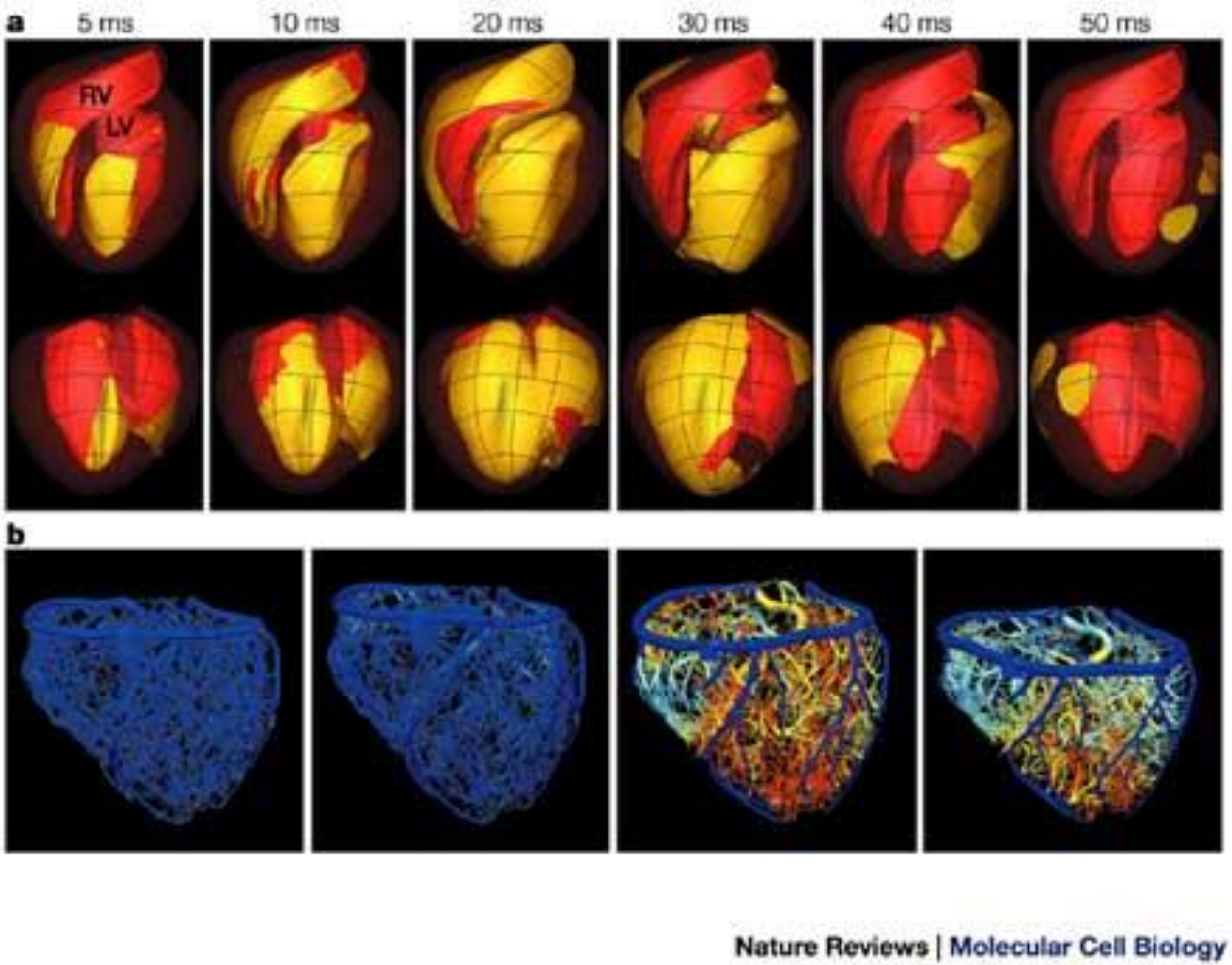
Source: AMAP (CIRAD, INRIA, INRA, IRD, CNRS, Montpellier) : réhabilitation d’une carrière



Virtual heart - Denis Noble *et al*. (Oxford – Physiome Project)

"*The 'Oxford Cardiac Electrophysiology Group' led by Professor Denis Noble is an example for having developed a virtual model of the human heart, which integrates the kinetic characteristics of the molecular and cellular mechanisms of heart activity into detailed anatomical heart models and allows forecasts to be made on the physiology and pathophysiology of the heart*", Dr. Roland Eils (German Cancer Research).

Source : http://bio-pro.de/magazin/thema/00173/index.html?lang=en&artikelid=/artikel/03079/index.html



Physiome Project : Auckland, Oxford, San Diego

Source : http://www.nature.com/nrm/journal/v4/n3/box/nrm1054_BX2.html

The era of « Multis »

-1- Multi-aspectual

-2- Multiscale

- 3- Multi-physical : electrical, mechanical, chemical phenomena

- 4- Multidisciplinary: chemistry, mechanics, electricity, biology…

- 5- Multifield

Specifically, if this integrative simulation is not reducible to data-fusion (data-fusion = "action / decision oriented" integrative simulation for detection of targets or weapons**), this multiplicity implies too :**

- Multi-scale

- Multiplicity of **epistemic status** of the submodels of each scale or each aspect : (Varenne, 2007, 2010, 2013)
  - explanative submodels with verified or hypothesized mechanisms
  - phenomenological submodels (stochastic processes, Monte Carlo…)
  - digitalization of captured scenes
  - IRM scannings
  - …

-Inter-models explanation : explanation through
  - "Emergence"
  - or not : only interactions at runtime between elementary mechanisms: "phenomenological reconstruction" [see (Peyrieras *et al.*) on the first steps of the ontogenesis of the zebra fish embryo: filiations cell by cell]

Conclusions on the functions of M&S

in particular after the computational turn

- New ways of representing, of imaging, of predicting
- New ways of explaining: "Science without laws" (Giere, 1999). Generative methods = growing a morphogenesis on the computer is like explaining without laws (from "laws" to "iterate mechanisms", Epstein, Axtell…).
- The search for universality is sometimes replaced by the search for iteration (of mechanisms) + interactions.
- New ways of experimenting: experimenting on interactions between symbols that have not always the same level of conventionality, hence the role of pluriformalized simulations and of phenomenological reconstruction before any global theorization.
  - Pb: computational emergence // biological emergence ? (Varenne 2012 "La reconstruction phénoménologique par simulation : vers une épaisseur du *simulat*" ; Varenne, 2013 : Chains of Reference in Computer Simulations)
- New ways of communicating the scientific understanding of phenomena to students: letting them experiment on computational models
- The "computerization" of interdisciplinary conceptual models and data communications between experts: thanks to ontologies in the sense of 'ontologies' in computer science (*Open Biomedical Ontologies* de IFOMIS : Grenon & Smith, see *Biologie systémique*, Roux-Rouquié, Ecrin, 2007)
- Homogenization of formalisms *ex post* (simulations) or *ex ante* (ontologies).

Thank you !

- Some references:

  - Achinstein, P. (1968), *Concepts of Science – A philosophical analysis*, London, The Johns Hopkins Press.
  - Black, M. (1962), *Models and Metaphors – Studies in Language and Philosophy*, Ithaca and London, Cornell University Press.
  - Cartwright, N. (1983), *How the Laws of Physics lie*, Oxford, Oxford University Press.
  - Carusi A., Burrage K. and Rodriguez (2013), Model Systems in Computational Systems Biology. Juan Duran and Eckhart Arnold (Eds.): *Computer Simulations and the Changing Face of Scientific Experimentation*, Cambridge Scholars Publishing, p.118-144.
  - Finlay, S., *Predictive Analytics, Data Mining and Big Data – Myths, Misconceptions and Methods*, Palgrave-Macmillan, 2014.
  - Goodman, N. (1981), “Routes of Reference”, *Critical Inquiry*, vol. 8, n°1, 121-132.
  - Hempel, C.G. (1965), *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, New York, The Free Press.
  - Larose, D.Y. (2005), *Discovering Knowledge in Data – An Introduction to Data Mining*, Hoboken, Wiley.
  - Legay, Jean-Marie, *L’expérience et le modèle*, Paris, INRA Editions, 1997.
  - Minsky, M. (1965), “Matter, Mind and Models”, *Proc. of IFIP Congress*, 45-49.
  - MacKenzie, D. A. (1981), *Statistics in Britain*, Edinburgh, Edinburgh University Press.
  - Mäki, U. (ed.)(2002), *Fact and Fiction in Economics - Models, Realism and Social Construction*, Cambridge (UK), Cambridge University Press.
  - Morgan, M. S. & Morrison, M. (eds.) (1999), *Models as Mediators*, Cambridge (UK), Cambridge University Press.
  - Olivier, N. et al. (2010). « Cell lineage reconstruction of early zebrafish embryos using label-free nonlinear microscopy » *Science*, 2010, 329, 967–971.
  - Phan, D. & Amblard, F., *Agent-based Modelling and Simulation in the Social and Human Sciences*, Oxford, The Bardwell Press, 2007.
  - Phan, D. & Varenne, F., “Agent-Based Models and Simulations in Economics and Social Sciences: from conceptual exploration to distinct ways of experimenting", *Journal of Artificial Societies and Social Simulation*, 13(1), 5, 2010. Free access: http://jasss.soc.surrey.ac.uk/13/1/5.html
  - Schmid, A. F. (1998), *L’âge de l’épistémologie*, Paris, Kimè.
  - Van Fraassen, B. (1980), *The scientific Image*, Oxford, Clarendon Press.
  - Varenne, F., *Du modèle à la simulation informatique*, Paris, Vrin, 2007.
  - Varenne, F., "Models and Simulations in the Historical Emergence of the Science of Complexity", *in* M.A. Aziz-Alaoui & C. Bertelle (eds), *From System Complexity to Emergent Properties*, Springer, 2009, pp. 3-21. Preprint: Models & Simulations in the Historical Emergence of the Science of Complexity
  - Varenne, F., "Framework for M&S with Agents in regard to Agent Simulations in Social Sciences: Emulation and Simulation", in *Activity-Based Modeling and Simulation*, A. MUZY, D. R. C. HILL & B. P. ZEIGLER (eds.), Clermont-Ferrand, Presses Universitaires Blaise Pascal, 2010, pp. 53-84. Free access: FMSA in regard to Agent Simulations
  - Varenne, F., “Modèles et simulations dans l’enquête scientifique : variétés traditionnelles et mutations contemporaines” *in* F. Varenne & M. Silberstein, *Modéliser & Simuler – Epistémologies et pratiques de la modélisation et de la simulation*, Paris, Editions Matériologiques, 2013, pp. 11-49. Preprint: Introduction M&S 1
  - Varenne, F., “Chains of Reference in Computer Simulations”, FMSH selected working paper, FMSH-WP-2013-51, GeWoP-4, octobre 2013, 32 pages, Chains of Reference in Computer Simulations
  - Varenne, F., *From Models to Simulations*, New York, Routledge, 2018.
  - Winsberg E. (1999), Sanctioning models: the epistemology of simulation, *Science in Context*, 12, pp. 275-292.
  - Winsberg E. (2009), A Tale of Two Methods, *Synthese*, 169 (3), pp. 575-592.
  - Zeigler B. P., Praehofer H. and Kim Tag G. (2000), *Theory of Modeling and Simulation. Integrating Discrete Event and Continuous Complex Dynamic Systems*, New York, Academic Press, 2000, 1st edition: 1976.