DE LA RECHERCHE À L'INDUSTRIE

Very short introduction to Validation from the Simulation point of view

CEA/CESTA-DSGA
| Marc Sancandi

A right definition of Simulation

Pritsker (*Compilation of definitions of simulation*, in Simulation August 1979, pp 61-63) gave 21 different def.

"A PROCESS WHICH MIMICS THE RELEVANT FEATURES OF A TARGET PROCESS"

Hartmann, *The World as a process*, 1996

This definition includes many different "tangible" targets :

- Physical ones, that is all the real phenomena which surround us
  Hereafter named "World of Interest" (WoI)
- And non physical ones : Economics, Social Networks, Artificial Life ...

while exluding some intensive computational processes

- Games, Virtual Reality
- Most part of purely numerical algorithms (ex. Monte Carlo integration), ...

Extensive view is given in the Stanford Encyclopedia of Philosophy (on line), entry *"Computer simulations in science"*, Winsberg

One can mimic the target process by

- Theoritical and Computational means
  - What one usually think, when speaking about “simulation”
  - Example of a pure abstract model : Burger equation mimics highway traffic
- Experimental (= Empirical) means
  - More commonly seen as “observation” instead as “simulation”
  - Counter-example : laboratory simulation of Jupiter’s great red spot
- Or, simultaneously both !. . .
  - Complementarity and mutual enhancement of the two imitation processes

Remember that a representation (= Image) is not the reality (= Target process)
“A map is not the territory” Korzybski

The two Scientific Images of the World

- The Empirical Image : made of material experiments (not necessarly theory-based)
- The Rational Image (by opposition, from Kant, *Ciritcs of the pure reason*) : essentially a theory-based image which rests on (supposed-to-be) physical models and derived codes (models are scientific images of a theory).

So one can speak of Rational Simulation and of Empirical Simulation

The two Scientific Images of the World

- The Empirical Image : made of material experiments (not necessarly theory-based)
- The Rational Image (by opposition, from Kant, *Ciritcs of the pure reason*) : essentially a theory-based image which rests on (supposed-to-be) physical models and derived codes (models are scientific images of a theory).

So one can speak of Rational Simulation and of Empirical Simulation

What is "Numerical Simulation" (NS) ? .... (~ Computer Simulation)

Following Hartmann, Korb says
Philosophy of computer simulation, http ://www.csse.monash.edu.au/~korb/lmps.pdf

"Process $\mathbb{S}$ is a **Computer Simulation** iif it is a Simulation and a computer process

Not truly satisfactory for it seems to concern only Rational Simulation ...
... analysis of celestial images or DNA sequencing (empirical simulation) need huge NS means

Alternative definitions say too much or too less.

Examples of scientific images

- Flight of an airplane
  - Rational : use of NS equations plus turbulence models
  - Empirical : Scale model (= wind tunnel)
- Fatigue breaking of a piece of metal :
  - Rational : inference about the microscopic mechanical processes involved
  - Empirical : accelerated ageing, macroscopic observations
- Tumor growth :
  - Rational : discrete WoI continuously representated through a set of coupled PDE
  - Empirical : analogical model (mice = men)
- Portfolio management :
  - Rational : stochastic modeling
  - Empirical : Heads or tails ?
- Urban gang extension :
  - Rational : individuals thought as particles submitted to simple interacting rules.
  - Empirical : polls, police reports

The reality of "Simulation as a whole"

- Different actors may form different images of WoI (either rational and empirical)
  van Frassen, *The scientific image*, 1980, Oxford Univ. Press
- Rational and empirical images may be irreducible the one to the other
  (see above the example of an airplane flight)
- Each image has its own models, facilities, rules and practices . . .which are dependent on the knowledge and the skill of the operators

So, to reduce the subjectiveness of human factor, and ensure that "Simulation as a whole" is *valid*, (=legitimate) some controls are needed at each sensible point.

To avoid using connotated words (see below) , I introduce the acronym "SGA" *(Simulation Goodness Assurance)* to refer to this global control process

The three facets of a SGA program

1. Organizational *(not evoked here, a key point in engineering activities)*
   - ✓ Identification of tasks, teams, objectives, traceability, ...
   - ✓ A "paperwork" job ...but a fundamental one !
2. "Technical"
   - ✓ Here we are interested in answering theoretical questions
   - ✓ State the problem : Conceptualization + Specification
   - ✓ Next solve it !
3. Decisionnal *(not evoked here)*

   Here some decisions are taken, based on objective grounds

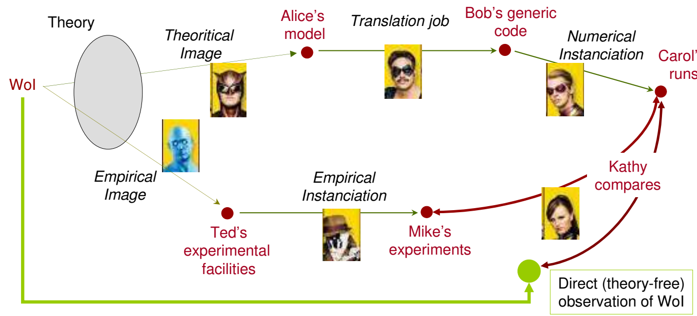
   - ✓ consistency of informations, objective costs, ...

   and often subjective ones

   - ✓ strategy (nuclear deterrent), policy (high scale protective inoculation)
   - ✓ an even social or psychological considerations, ...

An illustration of Simulation and its associated SGA process

The controllers team



" Is Alice's scientific image a good one ? "

"Is Ted's scientific image a good one ? "

" Does Bob's team correctly translate Alice's model ? "

....

Who watches the Watchmen ?

Each  do a specific job ; some of them recieved conventional names

- Accreditation ($\simeq$ Certification) decides on the relevancy of "Simulation as a whole" as a receivable mimicry of the target process

- Code Verification : *"How correct is the physics-to-code translation ?"*)
  (math. formulation, schemes, algorithmic ...)

  Calculation Verification : *"How correct is the code instanciation ?"*
  (mesh, time step, CFL, AMR, ...)

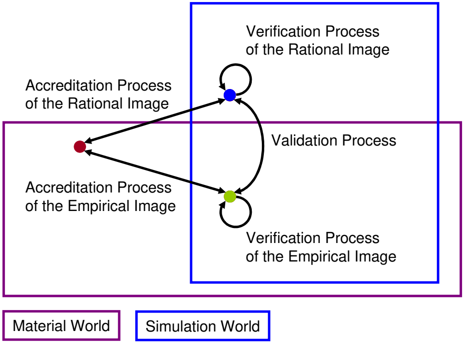
  Each are pure mathematical processes, self-sufficient & unambiguous

- Validation decides on the *correctness* of the rational image ($\sim$ the physical model).
  Conclusions about validation are drawn from comparisons between code predictions and experimental results (precisions given next slides).

Observe the dissymetry Validation introduces between rational and empirical images

The place where some control processes lie

- Target Process
- Rational Imitation Process
- Empirical Imitation Process



Some questioning the word "Validation" already raises (1/2)

The previous *definition* of Validation doesn't set identical status to numerical and empirical results.

Up to some measurement errors :

experimental results are implictly assume to say the "truth"

This may be problematic when

- The Empirical image doesn't faithfully represents WoI (ex. wind tunnel)
- Empirical and Rational representations differs in essence (see above) :

  Ex. of Probe atmospheric re-rentry in early stages : physical modeling based upon kinetic assumptions (=discontinuous matter) ; but measurements operate at macroscopic level (=continuous matter).

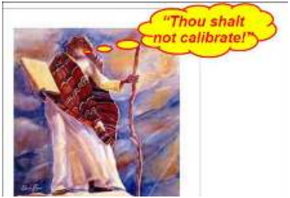
  Other one : Emergent behaviors modeled at entities level (agent-based modeling) while observation occurs on collective level. . .

Some questioning the word "Validation" already raises (2/2)

In the end, a lot of people identify Validation with Calibration

Big mistake : Calibration cares about the past,
Validation about the future !

"Prediction is very difficult especially if it's about the future", Niels Bohr, Nobel Prize of Physics, 1922



Thanks to F. Hémez

In Validation, predictions must agree with the target process.

In Calibration, predictions must agree with the empirical images of some target-like processes

Add-on 1 : A focus on the rational branch step-by-step

1. Choose a theoritical framework to represent WoI .... *WoI as a continuous media*
2. Use general theories and principles to derive a coarse model .... *continuum mechanics*
3. Refine it by introducing detailed features (laws) .... *constitutive laws*
4. Derive a mathematical computer tractable formulation .... *FE formulation*
5. Develop the corresponding generic code .... *Abaqus* ^{TM}
6. Instanciate this later .... *data file, mesh, ...*

Note that some empiricism appears here : step 3 constitutive laws are often semi-empirical formulas ; step 6 instanciation relates to empirical observations of WoI

Each step 1 to 6 gives rise to a new kind of model : what then means the word “model” ?

Add-on 2 : What means the expression "Modeling and Simulating" (M&S)

- Straightforwardly, Modeling means "building a formal representation (aka model) of WoI"
  Among the models seen above, what are those "Modeling" then covers ?

- Simulating implicitly mean "evaluate THE model".
  (Is "THE model" the last one [step 6 above] ?)

Fuzziness of word "model" makes acronym"M&S" unclear : where do we stop modeling and start simulating ?

Add-on 3 : The *"algebra"* of validation

Let us introduce the following logical propositions

- Val(M_n) is "The model M_n developped at step n of SGA is valid"
- Ver(M_n, M_{n+1}) is "The rewritting rule from model M_n to model M_{n+1} is correct"

Then one considers that Ver(M_n, M_{n+1}) & Val(M_{n+1}) $\models$ Val(M_n)

The first model is formal, and so often intractable

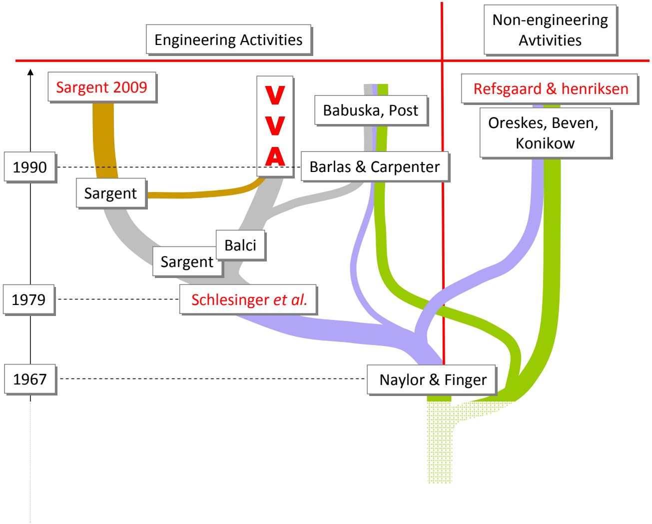
Recursively applying this rule up to the "user-code model" gives a way to infer the validity of this first model

$$\left[\text{Ver}(M_1, M_2)\ \&\ \ldots\ \&\ \text{Ver}(M_{N-1}, M_N)\right]\ \&\ \text{Val}(M_N) \models \text{Val}(M_1)$$

...and so we speak of "code validation" (= Val(M_N)) instead of "model validation" (= Val(M_1))

A simplified and subjective genealogy of some trends in SGA

Some references are given at the end of this presentation



Validation and Verification : What's what ?

Source	Vérification	Validation
Systems Eng. [1]	Proof of compliance with specifications. Verification may be determined by test, analysis, demonstration, and inspection.	Proof that the product accomplishes the intended purpose. Validation may be determined by a combination of test, analysis, and demonstration.
Software Eng. [2]	Software verification is a software engineering activity that demonstrates that the software products meet specified requirements.	Software validation is a software engineering activity that demonstrates that the as-built software product or software product component satisfies its intended use in its intended environment.
M&S [3]	The process of determining that a computational model accurately represents the underlying mathematical model and its solution from the perspective of the intended uses of M&S.	The process of determining the degree to which a model or a simulation is an accurate representation of the real world from the perspective of the intended uses of the model or the simulation.

[1] :*NASA systems engineering processes and requirements*, URL http ://nodis3.gsfc.nans.gov
[2] :*IEEE standard dictionary of electrical and electronics term*, ANSI/IEEE Std 100-1984 (1984)
[3] :*NASA standards for models and simulations*,NASA-STD-2009, 11 juillet 2008

M&S means *Modeling Simulation*

Validation and Verification : What's what ?

Source	Vérification	Validation
Systems Eng. [1]	Proof of compliance with specifications. Verification may be determined by test, analysis, demonstration, and inspection.	Proof that the product accomplishes the intended purpose. Validation may be determined by a combination of test, analysis, and demonstration.
Software Eng. [2]	Software verification is a software engineering activity that demonstrates that the software products meet specified requirements.	Software validation is a software engineering activity that demonstrates that the as-built software product or software product component satisfies its intended use in its intended environment.
M&S [3]	The process of determining that a computational model accurately represents the underlying mathematical model and its solution from the perspective of the intended uses of M&S.	The process of determining the degree to which a model or a simulation is an accurate representation of the real world from the perspective of the intended uses of the model or the simulation.

[1] :*NASA systems engineering processes and requirements*, URL http ://nodis3.gsfc.nans.gov
[2] :*IEEE standard dictionary of electrical and electronics term*, ANSI/IEEE Std 100-1984 (1984)
[3] :*NASA standards for models and simulations*,NASA-STD-2009, 11 juillet 2008

M&S means *Modeling Simulation*

The core question in any prediction activity (whether it be done via a code or a cristal ball) is

How far may I trust the result of this prediction ?

Convention

> The terms Qualification, Qualify(-ing) will be use in a decision-maker, or standardization sense.
> Thus, saying "I qualify a NS, or model, or a product" I will mean that the NS, model or product is considered as good, face to some normalization standards.
>
> Probably this sentence would necessitate some elucidation ?

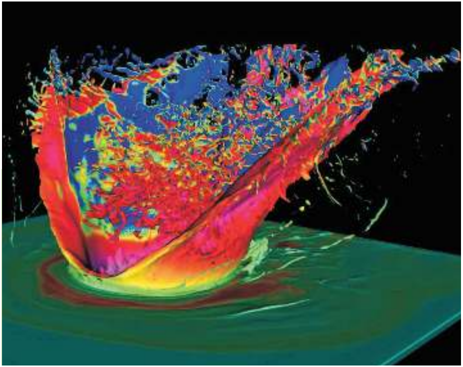
Simulation contexts differ from a field to another

	Geology	Epidemiology	Engineering
Purpose	Understand the Chicoxulub meteorit impact	Contain the spreading of some virus (ie. H1N1 flue)	Conceive some industrial product
Goals	Was it responsible of the extinction of dinosaurs ?	Minimize the number of infected people	Maximize performances while minimizing costs
Empirical grounds	Only one observation ; deduction, hypothetizing	Real time observations of spreading evolution	Many dedicated experiments
Confirmatory experiment	I hope not ! ! !	The end of the infection	Many experiments
Do it again	No	No	Yes
Faith enhanced if	New discoveries confirm theoritical computations	Not assessable [a]	If experiments agree with numerical simulations
External constraints	No one	Public opinion, Media, Crisis management	Markets, consumers attitude, costs

a. What would happened if we had did things differently ?

The famous Chicoxulub meteorit impact

From Post & Votta, *Computational science demands a new paradigm*, Physics Today 35, 2005



Yucatan, 65 million years ago,

Beautiful, but was it reality ...if reality maters ?

Qualification of an Engineering Concept (something manufactured)

- Simulations provide a mean to shorten the development time (to spend money ?)

  Car (crash test), airplane (or part of it), nuclear reactor (partialy)

  Here : classical trials (sim. exp.) and tests (material exp.) situation ; Simulation is just another way !

- Simulations enable to extend the experimental domain out of its boundaries

  Nuclear reactor (core melting), space probe, nuclear weapon

  A new organization of trials and tests is necessary ; Simulation can be the only way to proceed

Qualification of an Engineering Concept (something manufactured)

- Simulations provide a mean to shorten the development time (to spend money ?)
  Car (crash test), airplane (or part of it), nuclear reactor (partialy)
  Here : classical trials (sim. exp.) and tests (material exp.) situation ; Simulation is just another way !
- Simulations enable to extend the experimental domain out of its boundaries
  Nuclear reactor (core melting), space probe, nuclear weapon
  A new organization of trials and tests is necessary ; Simulation can be the only way to proceed

Qualification of a Natural Concept (by opposition to the previous one)

- Ozone layer damage caused by CFC, Chicoxulub impact, Supernova blast
  Uncontrollable, untestable, partially observable systems
- Also (up to some extents) : portfolio management, infection spreading

Is SGA necessary for Qualifying ?

- Yes if one can't qualify the concept in some other way
  - Often the case for Natural Concepts
    Don't free viruses to see if your code correctly simulates their spreading !
  - And sometimes for Engineering Concepts
    Don't launch a nuclear missile to verify code predictions !

- For engineering concepts, the "proof" is generally supplied through ad'hoc experiments

A two million dolars crash test of a Bugatti Veyron !

- While not necessary, SGA may participate to Qualifying
  The fact Veyron numbers 1 to 5 succeeded crash tests say there were good, but now they are good only for trash, and this proves nothing about the capability of Veyron numbers 6-100 to behave still correctly. SGA may help "proving" that it's the case (in a statistical sense).

The Great Schism

CS&E : Computational Science and Engineering ; S&NS : Social and Natural Sciences

CS&E

"It is possible to validate a model when the situation being modeled satisfies four prerequisites . . ."

[Shortly : it must be observable and measurable, ample data may be collected, it must exhibit some structural constancies (so as to we are able justify predictions of not already observed situations)

As a counterexample, a chaotic system doesn't meet these requirements]

Hodges & Dewar, *Is it you or your model talking ? A framework for model Validation*, Report R-41 14-AIAF/OSD, RAND Corp. 1992

Application domain : simulation of weapons systems in battlefield scenarios

SN&S

"Models in social and policy sciences generally fail to satisfy these criteria and therefore cannot be validated"

Oreskes, *Evaluation (not validation) of quantitative models*, Environmental Health Perspectives, Vol 106, Supp^{nt} 6, 1998

Consensus in validation isn't just around the corner

> *"Verification and validation are two extensively used terms in simulation. They are widely used in science in general, both in the natural and the social sciences.*
>
> *They have plethora of different methodological significances, in diverse epistemological perspectives, upon different beliefs, and expectations. They are used often with the same or interchangeable meanings.*
>
> *They are the subject of numerous scientific and philosophical debates, and connected to diverse disciplinary, interdisciplinary and multidisciplinary contexts.*
>
> *In spite of recalcitrant debates, a standard meaning is unlikely to emerge ...Terminological disputes seem unlikely to be useful. Consensus in meaning seems improbable."*
>
> David, *Validation and verification in social simulation : patterns and clarifications of terminology*

Some examples are given in the appendix.

Some dominent positions

- **SN&S : Ecosystems studies**

  Unlike the scientific hypothesis, a model is not verifiable directly by experiment.

  A model is a set of hypotheses that can only be refuted.

- **SN&S : Hydrogeology Studies**

  A model has no necessarily deductive capacity Environmental models cannot be validated but only confirmed

  The terms validation and verification lead to a false impression of model capability'

- **CS&E**

  Validation is the process of :
  determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model

  confirming that the predictions of a code adequately represent measured physical phenomena.

complete citations are given in the appendix

If I had to choose only one definition to use in CS&E field , it would be this last

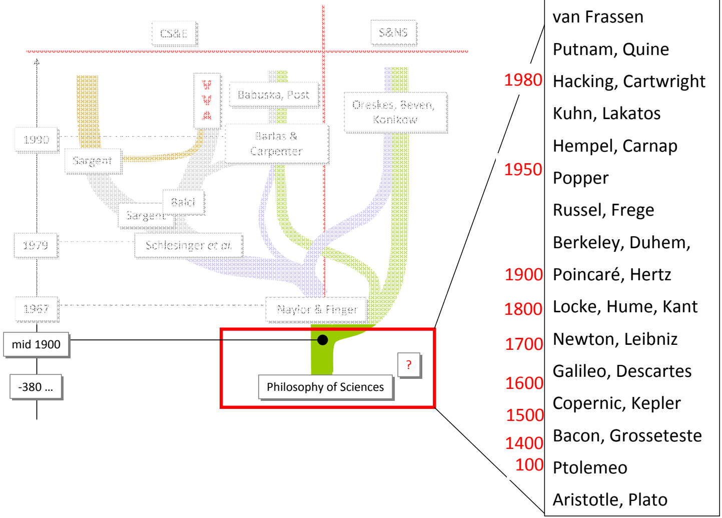
> [Validation] :*"The process of determining the extent to which an M&S is an accurate representation of the real world from the perspective of the intended use of the M&S.*
>
> *Validation methods include expert consensus, comparison with historical results, comparison with test data, peer review, and independent review."*

Department of Army Regulation (AR) 5-11, *Management of Army Models & Simulations*, July 10, 1997.

This definition emphasizes the role of external (aka *"non functional"*) constraints on any validation program ... some of them may be very far from scientific considerations !

1. First steps in Validation
   - Simulation in a nutshell
   - A few words and a lot of meanings
2. The ideas and the tools of Validation
   - Validation as a philosophical question
   - Validation as a probabilistic question
   - Validation as a formal question
   - Validation as a pragmatic (operational) question
3. Ideas face reality . . .
   - The Legacy
   - A short look on VV&A (CS&E field)
4. Conclusion
   - Mistakes and advices

All began 25 centuries ago



Early philosophers (Plato, Aristotle, ...) considered that questions relative to the validity of theories might be reduced to questions of logic alone...

This point of view has dominated the scientific community until the beginning of the 20^{th} century

Deduction *.vs.* Induction

Deduction

Every man is mortal, Socrates is a man, thus Socrates is mortal (Aristotle ~ -350)

Deduction = from the general to the particular

Induction

This raven is black, so is this one, and also this one ..., thus "all ravens are black"

Induction = from many particular "Positive Instances" to the general

Deduction *.vs.* Induction

Deduction

Every man is mortal, Socrates is a man, thus Socrates is mortal (Aristotle ~ -350)
Deduction = from the general to the particular

Induction

This raven is black, so is this one, and also this one ..., thus "all ravens are black"
Induction = from many particular "Positive Instances" to the general

Falsification Principle

Grosseteste ~ 1200

Accumulation of "positive" instances (*ie.* "raven x_n is black") doesn't entail the truth of the theory (hypothese). On the other and, only one "negative" instance ("raven x_n is white") is enough to discredit it

Popper ~ 1937

A theory ($\sim$ a model) can only be refuted by experimental observations, never validated. As long it hasn't be refuted (or falsified), we say it is CORROBORATED
Corroborated $\sim$ some kind of provisional validity

One modern translation

"Surely the predictive value of a calibrated model is precisely zero ...because calibrating a model to measurements does not say anything regarding the ability of that model to match the next set."

Hémez, *15 Years of verifying and validating structural dynamics simulation at Los Alamos*, LA-UR-07-2213, 2007

The Logic of Falsification

$\mathbb{M}(x)$: "Model $\mathbb{M}$ correctly predicts for conditions x"

$O(x)$: "Observation at x equals the predicted value"

$\longrightarrow$

Formally

$[\mathbb{M}(x) \vDash O(x)]$

Inverting the direction of the cause

Accept $\mathbb{M}$ as "valid" if $O(x)$ is true :

$$[\mathbb{M}(x) \vDash O(x)] \Leftrightarrow [O(x) \vDash \mathbb{M}(x)]$$

The Logic of Falsification

$\mathbb{M}(x)$: "Model $\mathbb{M}$ correctly predicts for conditions x"

$O(x)$: "Observation at x equals the predicted value"

$\longrightarrow$ Formally

$[\mathbb{M}(x) \vDash O(x)]$

~~Inverting the direction of the cause~~

~~Accept $\mathbb{M}$ as "valid" if $O(x)$ is true :~~

~~$[\mathbb{M}(x) \vDash O(x)] \Leftrightarrow [O(x) \vDash \mathbb{M}(x)]$~~

FALSIFICATION

reject $\mathbb{M}$ as valid if $O(x)$ false :

$[\mathbb{M}(x) \vDash O(x)] \Leftrightarrow [\neg O(x) \vDash \neg\mathbb{M}(x)]$

The Logic of Falsification

$\mathbb{M}(x)$: "Model $\mathbb{M}$ correctly predicts for conditions x"

$O(x)$: "Observation at x equals the predicted value"

$\longrightarrow$ Formally

$$[\mathbb{M}(x) \vDash O(x)]$$

~~Inverting the direction of the cause~~

~~Accept $\mathbb{M}$ as "valid" if $O(x)$ is true :~~

~~$[\mathbb{M}(x) \vDash O(x)] \Leftrightarrow [O(x) \vDash \mathbb{M}(x)]$~~

FALSIFICATION

reject $\mathbb{M}$ as valid if $O(x)$ false :

$$[\mathbb{M}(x) \vDash O(x)] \Leftrightarrow [\neg O(x) \vDash \neg \mathbb{M}(x)]$$

Popper

Strong corroboration follows from prediction of really new facts (*i.e.* unknown before predictions) By contrast, prediction of well known facts only slightly increases the corroboration of the model.

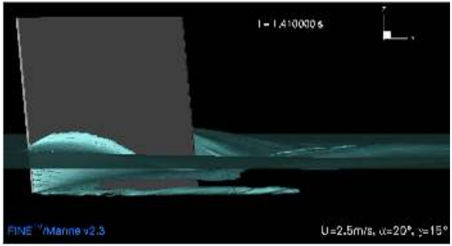
More or less the account for prior knowledge about the *"occurrability"* of the facts ...see below Bayesian Confirmation Theory

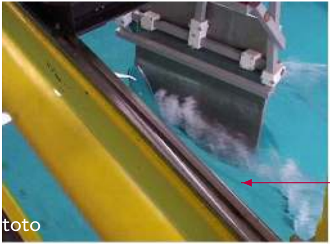
An exemple of prediction of an unknown fact

By courtesy of Michel Visonneau : Visonneau, Deng, Queutey & Wackers, *Anisotropic grid adaptation for RANS simulation of ship flow*, FAST 2011, 11th In. Conf. on Fast Sea Transportation Honolulu, USA, Sep. 2011

Simulation with ISIS-CFD code of a breaking bow wave created by an inclined flat plane (leeward side).

Note the odd air tube below the free surface





Experiment (Ecole Centrale de Nantes)

Prediction of the odd "air-tube" has lead to close observation of the film, thus revealing the existence of a "real" air-tube

Falsification and Auxiliary Hypothesis

The role of Auxiliary Hypothesis $\mathbb{A}$ (Hume ~ 1750, Hempel 1950) ; exemple :

$\mathbb{M}(x)$: "The true value should be 10"

$\mathbb{A}$: "Meas. errors :~Unif$(-1, 1)$"

$O(x)$: "Observation should lie in $[9, 11]$

$\longrightarrow$

Formally

$[[\mathbb{M}(x) \,\&\, \mathbb{A}] \models O(x)]$

Falsification and Auxiliary Hypothesis

The role of Auxiliary Hypothesis $\mathbb{A}$ (Hume ~ 1750, Hempel 1950) ; exemple :

$\mathbb{M}(x)$: "The true value should be 10"

$\mathbb{A}$: "Meas. errors :~Unif$(-1, 1)$"

$O(x)$: "Observation should lie in $[9, 11]$"

⟶

Formally

$$[[\mathbb{M}(x) \,\&\, \mathbb{A}] \vDash O(x)]$$

$O(x)$ may be false because

- ✓ $\mathbb{M}(x)$ si false
- ✓ or $\mathbb{A}$ is !

$$\big[\,[\mathbb{M}(x) \,\&\, \mathbb{A}] \vDash O(x)\,\big] \Leftrightarrow \big[\,\neg O(x) \vDash \neg[\mathbb{M}(x) \,\&\, \mathbb{A}]\,\big]$$
$$\Leftrightarrow \big[\,\neg O(x) \vDash [\neg\mathbb{M}(x) \vee \neg\mathbb{A}]\,\big]$$

Falsification and Auxiliary Hypothesis

The role of Auxiliary Hypothesis $\mathbb{A}$ (Hume ~ 1750, Hempel 1950) ; exemple :

$\mathbb{M}(x)$: "The true value should be 10"

$\mathbb{A}$: "Meas. errors :~Unif$(-1, 1)$"

$O(x)$: "Observation should lie in $[9, 11]$

→

Formally

$$[[\mathbb{M}(x) \,\&\, \mathbb{A}] \vDash O(x)]$$

$O(x)$ may be false because

- ✓ $\mathbb{M}(x)$ si false
- ✓ or $\mathbb{A}$ is !

$$\big[\,[\mathbb{M}(x)\ \&\ \mathbb{A}] \vDash O(x)\,\big] \Leftrightarrow \big[\,\neg O(x) \vDash \neg[\mathbb{M}(x)\ \&\ \mathbb{A}]\,\big]$$
$$\Leftrightarrow \big[\,\neg O(x) \vDash [\neg\mathbb{M}(x) \vee \neg\mathbb{A}]\,\big]$$

One is generally able to save a refuted code by invoking falsity of some (ad'hoc ?) "Auxiliary Hypotheses" (Model Error in Bayesian calibration is an illustration)

Pragmatism : drop out the code only if you have another one which succeed in the falsification test !

Where Logics comes up against common sense : The Paradox of Ravens

From Hempel, or "*How to be an ornithologist without leaving home ?*"

My theory $\mathbb{T}$ is that *"Every Raven is Black"*

so : $[\forall\ x\ :\ R(x) \vDash B(x)]$

Logical implication $\boxed{[R(x) \vDash B(x)] \Leftrightarrow [\neg B(x) \vDash \neg R(x)]}$ means that *"Every non-Black is a non-Raven"* ...and so

Where Logics comes up against common sense : The Paradox of Ravens

From Hempel, or "*How to be an ornithologist without leaving home ?*"

My theory $\mathbb{T}$ is that *"Every Raven is Black"*

so : $[\forall\ x\ :\ R(x) \vDash B(x)]$

Logical implication $\boxed{[R(x) \vDash B(x)] \Leftrightarrow [\neg B(x) \vDash \neg R(x)]}$ means that *"Every non-Black is a non-Raven"* ...and so

This non-Black, which of course is a non-Raven (!), confirms $\mathbb{T}$ as does the observation of a Black Raven.

The moral of this is : *Stay at home and count all non-Raven non-Black around you : that will enhance your confidence in* $\mathbb{T}$

For a general survey of the Raven Paradox see http ://en.wikipedia.org/wiki/Raven_paradox

What paradoxes say to us ?

From common sense, enhancing $\mathbb{T}$ from its contraposition seems stupid !
At the same time, it is imparable from the rules of Logics

> Transposing the paradox of ravens to code corroboration could give something like this
>
> Direct My meterology code is corroborated because it predicted the today temperature would be between 11 and 12 Celsius degres. . . what it appears it is.
>
> Reversed My code is corroborated too, for today temperature is not equal to 100 Celsius degres . . . and for my code didn't predicted it would be.

Many philosophers of science developed arguments to explain how to solve this conflict.

All of them rest on probabilities and on the prior explanatory capability of every empirical observation (high capability has black raven, low one has green apple).

So, keep in mind that "validation" is all but reducible to formal Logics : Probability is required

Don't worry, be happy

- *"It is true that we cannot logically prove that a model is true. But maybe their way* [ref to Oreskes et al.] *of defining* [validation] *is too strict. Do we really want that absolute certainty ?"*

  From P. Suppe, reported by Irobi et al. in *Correctness criteria for models' validation - A philosophical perspective*

- *"Don't take underdetermination, and assumption-ladenness of simulation models, too seriously."* Levin (1966)

Don't worry, be happy

- *"It is true that we cannot logically prove that a model is true. But maybe their way* [ref to Oreskes et al.] *of defining* [validation] *is too strict. Do we really want that absolute certainty ?"*

  From P. Suppe, reported by Irobi et al. in *Correctness criteria for models' validation - A philosophical perspective*

- *"Don't take underdetermination, and assumption-ladenness of simulation models, too seriously."* Levin (1966)

Not very far from George P.E. Box' thought

> *"Essentially, all models are wrong but some are usefull"*

1. First steps in Validation
   - Simulation in a nutshell
   - A few words and a lot of meanings
2. The ideas and the tools of Validation
   - Validation as a philosophical question
   - **Validation as a probabilistic question**
   - Validation as a formal question
   - Validation as a pragmatic (operational) question
3. Ideas face reality ...
   - The Legacy
   - A short look on VV&A (CS&E field)
4. Conclusion
   - Mistakes and advices

What did we learn ?

- That Deduction is the only (while out of reach) scientific principle for establishing proofs
- That we can't base anything on Induction, while it is easy to use.
- That Validation is an unreachable ideal and that a correct concept is Corroboration

The pregnant question then becomes :
How can we reason correctly thanks to the Induction principle while building a satisfactory "weak-sense-theory" of validation ?

We're going to see that Bayesian Confirmation Theory (BCT) is an attempt to answer this question from a probabilistic point of view

Simple ideas about Confirmation measures : **(1)** Metrics (1/2)

Let's start with the next program :

- First **(1)** Alice has predicted that empirical test $\mathbb{T}$ would give the "answer" T.
  Next **(2)** Bob does the test and obtains observation $\widehat{T}$
- One chooses a "metric" $\mathfrak{C}(\mathbb{C}\,|\,T,\ \widehat{T},\ \mathbb{K})$ to decide, given T, $\widehat{T}$ and some background knowledge $\mathbb{K}$, if empirical answer $\widehat{T}$ enhances our faith in $\mathbb{C}$ or falsifies it.

  At first sight one may consider $\mathfrak{C}$ as a binary application onto $\{$"Confirms", "Infirms"$\}$:

  - "The result of $\mathbb{T}$ CONFIRMS $\mathbb{C}$ if $\mathfrak{C}(\mathbb{C}\,|\,T,\ \widehat{T},\ \mathbb{K}) > \alpha$
  - "The result of $\mathbb{T}$ INFIRMS $\mathbb{C}$ if $\mathfrak{C}(\mathbb{C}\,|\,T,\ \widehat{T},\ \mathbb{K}) < \alpha$

Simple ideas about Confirmation measures : **(1)** Metrics (1/2)

Let's start with the next program :

- First **(1)** Alice has predicted that empirical test $\mathbb{T}$ would give the "answer" T.
  Next **(2)** Bob does the test and obtains observation $\widehat{T}$
- One chooses a "metric" $\mathfrak{C}(\mathbb{C} \,|\, T, \widehat{T}, \mathbb{K})$ to decide, given T, $\widehat{T}$ and some background knowledge $\mathbb{K}$, if empirical answer $\widehat{T}$ enhances our faith in $\mathbb{C}$ or falsifies it.

  At first sight one may consider $\mathfrak{C}$ as a binary application onto {"Confirms", "Infirms"} :
  - "The result of $\mathbb{T}$ CONFIRMS $\mathbb{C}$ if $\mathfrak{C}(\mathbb{C} \,|\, T, \widehat{T}, \mathbb{K}) > \alpha$
  - "The result of $\mathbb{T}$ INFIRMS $\mathbb{C}$ if $\mathfrak{C}(\mathbb{C} \,|\, T, \widehat{T}, \mathbb{K}) < \alpha$

Note here the order of the operations : predicting is the first, experimenting is the last

Within a Bayesian framework, answer T appears as "a fact", or "prior" , not an assumption (as in Bayesian calibration)

Simple ideas about Confirmation measures : **(1)** Metrics (2/2)

Intuitively, the idea of confirmation should reflect the opinion that

$\widehat{T}$ is declared to confirm the code $\mathbb{C}$

1. If we have more faith in $\mathbb{C}$ AFTER $\widehat{T}$ has been obtained, than we had BEFORE the observation of $\widehat{T}$
2. If we have more faith in $\mathbb{C}$ after $\widehat{T}$ has been obtained, than we would have had if $\widehat{T}$ hadn't been observed
3. If the value of $\mathfrak{C}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K})$ exceeds some treshold

What's the point of the faith in $\mathbb{C}$ raising by a factor 10000 if its prior value was 10^{-10} or less ; wouldn't a lower 100% increase from 0.1 to 0.2 be better ? ($\rightarrow$ Bayesian model-selection strategies : what is the model we believe in stronger ?)

Simple ideas about Confirmation measures : **(1)** Metrics (2/2)

Intuitively, the idea of confirmation should reflect the opinion that

$\widehat{T}$ is declared to confirm the code $\mathbb{C}$

1. If we have more faith in $\mathbb{C}$ AFTER $\widehat{T}$ has been obtained, than we had BEFORE the observation of $\widehat{T}$
2. If we have more faith in $\mathbb{C}$ after $\widehat{T}$ has been obtained, than we would have had if $\widehat{T}$ hadn't been observed
3. If the value of $\mathfrak{C}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K})$ exceeds some treshold

What's the point of the faith in $\mathbb{C}$ raising by a factor 10000 if its prior value was 10^{-10} or less ; wouldn't a lower 100% increase from 0.1 to 0.2 be better ? ($\rightarrow$ Bayesian model-selection strategies : what is the model we believe in stronger ?)

Classical examples of incremental "metrics" (not to say they are easily computable !) :
Other given in Fitelson PhD, *Studies in Bayesian confirmation theory*, 2001, Univ. of Madison

$$\mathfrak{C}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K}) = \text{Prob}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K}) - \text{Prob}(\mathbb{C} \,|\, \mathbb{K})$$

$$\mathfrak{C}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K}) = \text{Prob}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K}) - \text{Prob}(\mathbb{C} \,|\, T,\, \neg\widehat{T},\, \mathbb{K})$$

Simple ideas about Confirmation measures : **(2)** Test choice

Each test $\mathbb{T}_n$ must be

- measurable (observable) ! ! !
- representative of the operational conditions the system (CS&E) is thought to encounter.
- discriminant, in the sense that a result $\widehat{T}_n$ close to T_n must not be in the bag

  Remember Popper

  The objective of any scientific researcher should be to try honnestly to falsify its model (to force it to the wall)

  Popper, *Logic of scientific discovery, First English Edition (1959)*, Routledge Classics, Taylor & Francis (reprinted 2009)

Part of knowledge $\mathbb{K}$ is made of Uncertainties $\mathbb{U}$: they need to be controled during the tests

Some types of Uncertainties $\mathbb{U}$

p 376 from : Parker, *Computer simulation through an error-statistical lens*, Synthï£¡se, 163, 2008

Study Design Error

- Error due to limited number of simulation runs / trials
- Inadequate sampling method

Substantive Modeling Error

- Error in equations for modeled processes (form, parameter values)
- No representation of relevant processes
- Overly simplified/erroneous initial and/or boundary conditions

Data Processing Error

- Error introduced by processing of raw simulation results

Solution Algorithm Error

- Inapplicable solution algorithm
- Unstable solution algorithm

Numerical Error

- Discretization error
- Iterative convergence error
- Truncation error

Programming Error

- Inadequate/faulty program design
- Coding typo/mistake

Hardware-related Error

- Round-off error
- Internal malfunction
- External interference

Simple ideas about Confirmation measures : **(3)** Decision

When "$\widehat{T}$ infirms $\mathbb{C}$" three standards positions may be taken

1. Purist (aka "the naï£¡ve falsificationist")

   Following Popper's claims we reject the code and start developping a new one (expensive of course)

   K. Popper, *Logic of scientific discovery*, ibid.

2. Pragmatic (aka "the subtle falsificationist)

   Lakatos (for ex.) suggests that we reject $\mathbb{C}$ only if we have a better code to use.

   "Protective arguments", such as the falsity Auxiliary Hypotheses (part of $\mathbb{K}$) are invoked

   I. Lakatos, *Methodology of scientific research programs*, 1970 Cambridge Univ. Press

3. Relativist

   According to Kuhn : anyway, this is not a technical question but a matter of social, historical, political conventions !

   T. Kuhn, *Structure of scientific revolutions* Second Edition, Chicago University Press, 1970

Bayesian Confirmation Theory (BCT)

Carnap, *Logical foundations of probability*, Chicago Univ. Press, (1962)
Howson & Urbach, *Scientific reasoning, the Bayesian approach*, (1989)

Even it suffers some fundamental difficulties [a] BCT is the most popular way to assess the degree to which an empirical observation confirms our beliefs in a theory

Proponents of BCT claim :

- Firstly : that the correct framework to express ideas about Confirmation is the one offered by Probability Theory (a few alternate versions are based on plausibilities)
- Secondly : that Bayesian reasoning (aka *Bayesianism*) is the correct way to account for these probabilities
- Thirdly : that Confirmation metrics have to be incremental [b] , not absolute.

Nest slide will show that the price to pay for point (3) above, is the lost of unicity in the conclusion we may draw

a. (Norton, *Challenges to Bayesian confirmation theory*, 2009)
b. *"Confirmation as Increase of Firmness"*, Carnap ibid

A kind of *No Free Lunch Theorem*

Are all the incremental metrics equivalent ?

Definition

Let $\left\{\mathbb{T}_{n=1,\ldots,N}\right\}$ a countable set of tests with corresponding $\mathbb{C}$-predictions $\left\{T_{n=1,\ldots,N}\right\}$ and empirical observations $\left\{\widehat{T}_{n=1,\ldots,N}\right\}$.

Let $\mathfrak{C}_1$ an incremental metric ; we define the map σ from $\{1,\ldots,N\}$ onto itself by

$$\mathfrak{C}_1\left(\mathbb{C}\,|\,T_{\sigma(1)},\widehat{T}_{\sigma(1)},\mathbb{K}\right)\leq\ldots\leq\mathfrak{C}_1\left(\mathbb{C}\,|\,T_{\sigma(N)},\widehat{T}_{\sigma(N)},\mathbb{K}\right)$$

Thus, incremental metrics $\mathfrak{C}_2$ is said *ordinally equivalent* to $\mathfrak{C}_1$ iff

$$\mathfrak{C}_2\left(\mathbb{C}\,|\,T_{\sigma(1)},\widehat{T}_{\sigma(1)},\mathbb{K}\right)\leq\ldots\leq\mathfrak{C}_2\left(\mathbb{C}\,|\,T_{\sigma(N)},\widehat{T}_{\sigma(N)},\mathbb{K}\right)$$

Given two different incremental metrics $\mathfrak{C}_1$ and $\mathfrak{C}_2$, it has been proved that it exist at least a couple of test cases for which the previous requirement fails.
Fitelson, *Studies Bayesian confirmation theory* , PhD, Univ. of Wisconsin, (2001)

So there is no good choice for $\mathfrak{C}$, and different choices may lead to different conclusions !

How BCT interpretes Bayes' formula ?

I consider here the absolute metric $\mathfrak{C}(\mathbb{C}\,|\,T,\,\widehat{T},\,\mathbb{K}) = \text{Prob}(\,T\,|\,\widehat{T},\,\mathbb{K})$

By Bayes' formula we have $\mathfrak{C}(\mathbb{C}\,|\,T,\,\widehat{T},\,\mathbb{K}) = \dfrac{\text{Prob}(\,\widehat{T}\,|\,T,\,\mathbb{K}) \times \text{Prob}(\,T\,|\,\mathbb{K})}{\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})}$

All the *rhs* terms, but $\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})$, have classical Bayesian interpretations.

$\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})$ quantifies the so-called "Surprise Effect" :

The more unusual $\widehat{T}$ is, the lower $\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})$ is, and the higher $\mathfrak{C}(\mathbb{C}\,|\,T,\,\widehat{T},\,\mathbb{K})$ is

How BCT interpretes Bayes' formula ?

I consider here the absolute metric $\mathfrak{C}(\mathbb{C}\,|\,T,\,\widehat{T},\,\mathbb{K}) = \text{Prob}(\,T\,|\,\widehat{T},\,\mathbb{K})$

By Bayes' formula we have $\mathfrak{C}(\mathbb{C}\,|\,T,\,\widehat{T},\,\mathbb{K}) = \dfrac{\text{Prob}(\,\widehat{T}\,|\,T,\,\mathbb{K}) \times \text{Prob}(\,T\,|\,\mathbb{K})}{\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})}$

All the *rhs* terms, but $\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})$, have classical Bayesian interpretations.

$\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})$ quantifies the so-called "Surprise Effect" :

The more unusual $\widehat{T}$ is, the lower $\text{Prob}(\,\widehat{T}\,|\,\mathbb{K})$ is, and the higher $\mathfrak{C}(\mathbb{C}\,|\,T,\,\widehat{T},\,\mathbb{K})$ is

Two hard points

(1) Whatever its name be (Evidence in Bayesian model selection or Surprise Effect in BCT) assessing the denominator of Bayes's formula is a challenging problem (ex given further).

(2) What we put in $\mathbb{K}$ is the second point. We must be very careful to let it free of evidences in favor of the theory we attempt to confirm (aka "old evidence"), see Fitelson ibid

Surprise Effect is a core term in incremental BCT metric

For example, if one says that observation $\widehat{T}$ confirms $\mathbb{C}$ if

$$\mathfrak{C}(\mathbb{C} \,|\, T,\, \widehat{T},\, \mathbb{K}) = \text{Prob}(\,T \,|\, \widehat{T},\, \mathbb{K}) - \text{Prob}(\,T \,|\, \mathbb{K}) > 0$$

(meaning that our faith in $\mathbb{C}$ is increased by the observation of $\widehat{T}$)

it is obvious to show that confirmations arises if

$$\frac{\text{Prob}(\,T \,|\, \widehat{T},\, \mathbb{K})}{\text{Prob}(\,\widehat{T} \,|\, \mathbb{K})} > 1$$

Thus we need to assess the "evidence of $\widehat{T}$ given $\mathbb{K}$

This is a general result that for any incremental metric we have to asses an "evidence-like term" of the form $\text{Prob}(\,\widehat{T} \,|\, \mathbb{K})$ or $\text{Prob}(\,T \,|\, \mathbb{K})$

Accounting for operators subjectiveness in SGA process

Here $\mathbb{C}$ refers to the "generic code" (the code before its complete instanciation).

Any attempt to assess the level to which an empirical observation $\widehat{T}$ confirms (or not) $\mathbb{C}$ from its prediction T, is in fact an attempt to confirm the user-code which has given this prediction.

To emphasize the subjective role of the operator "op" who instanciated $\mathbb{C}$, it's better to write $\mathfrak{C}(\mathbb{C}\,|\,\mathsf{op},\ T,\ \widehat{T},\ \mathbb{K})$, instead of $\mathfrak{C}(\mathbb{C}\,|\ T,\ \widehat{T},\ \mathbb{K})$

Let us suppose that N operators do, independently the ones of the others, the same prediction job. Denoting by $\mathsf{Bel}\,(\mathsf{op}_i)$ our believe in operator op_i, one can build the following op-independant estimation for $\mathfrak{C}(\mathbb{C}\,|\ T,\ \widehat{T},\ \mathbb{K})$

$$\mathfrak{C}(\mathbb{C}\,|\ T,\ \widehat{T},\ \mathbb{K}) = \frac{\sum_{i=1}^{i=N} \mathfrak{C}(\mathbb{C}\,|\,\mathsf{op}_i,\ T,\ \widehat{T},\ \mathbb{K}) \times \mathsf{Bel}\,(\mathsf{op}_i)}{\sum_{i=1}^{i=N} \mathsf{Bel}\,(\mathsf{op}_i)}$$

We saw in section *"First steps in validation"* that questions arising in Simulation use several terms, sometimes in quite different meanings, dependening themselves on some other words, the meaning of which . . .and so on

At least in CS&E, some attempt has been made to merge these terms in specific glossaries

We're going here to discuss some points concerning semantic and syntactic aspects of these glossaries and of more sophisticated formal representations

A starting example

From Thacker *et al.*, *Concepts of model verification and validation*, (Glossary), Los Alamos National Laboratories Report, LA-14167-MS, (2004)

> Validation is the process of confirming that the predictions of a code adequatly represent measured physical phenomena.

Validation is defined here to as a confirmation process : is confirmation defined elsewhere in this Glossary ? the answer is no

A starting example

From Thacker *et al.*, *Concepts of model verification and validation*, (Glossary), Los Alamos National Laboratories Report, LA-14167-MS, (2004)

> Validation is the process of confirming that the predictions of a code adequatly represent measured physical phenomena.

Validation is defined here to as a confirmation process : is confirmation defined elsewhere in this Glossary ? the answer is no

This definition depends on several other premisses (prediction, code, measure[d], phenomena)

- Are they already specified ?
- Does circularity in definitions avoided ?
- Excepted *atomic* terms (*of, the, many, exist, . . .*) do these definitions form a self-documented part of the Glossary ?

here again the answer is no

What a glossary is made of ?

We can separate its entries in three differents classes (sets) named

- Object

  ex : Reality, Code, Conceptual Model, Experiment, Measure, ...

- Property :

  ex : Validity (or "is valid"), Fidelity, Confidence...

  properties caracterize objects

- Function

  ex : Validating, Predicting, Modeling, Measuring...

  functions map an object to another

A natural requirement is that any function from an objet O to an object O' transport each property of O onto O'

Ex. : Function [Compile] must transport Property [Is_Verified] from Object [Source_Code] to Object [Exec_Code]

A lot of mathematical (ex. categories, sketches) or software engineering (ex. ontologies) tools may be used to formalize correctly these few lines

What is an Ontology ?

One of the most important characteristics of today's society is that a huge amount of information is shared by many participants . . . This information must be characterized by a uniformity of terms. This means that, in similar contexts, everyone should understand the same meaning when reading or hearing the same word and everyone should use the same word to refer to the same concept. In different Computer Science disciplines one of the methods that satisfies this need for "common understanding" of concepts is the creation of ontologies.

Sánchez *et. al*, *On models and ontologies*, in. Ontologies, chap 1 ; Integrated Series in Information Systems, vol. 14, Springer US

An ontology is an inventory of the kinds of entities that exist in a domain, their salient properties, and the salient relationships that can hold between them.

Benjamin *et. al.* ; , *Towards a method for acquiring CIM ontologies*, Int. J. of Comp. Integrated Manufacturing, 8 (3)

What is an Ontology ?

> *One of the most important characteristics of today's society is that a huge amount of information is shared by many participants ...This information must be characterized by a uniformity of terms. This means that, in similar contexts, everyone should understand the same meaning when reading or hearing the same word and everyone should use the same word to refer to the same concept. In different Computer Science disciplines one of the methods that satisfies this need for "common understanding" of concepts is the creation of ontologies.*
>
> Sánchez *et. al*, *On models and ontologies*, in. Ontologies, chap 1 ; Integrated Series in Information Systems, vol. 14, Springer US

> *An ontology is an inventory of the kinds of entities that exist in a domain, their salient properties, and the salient relationships that can hold between them.*
>
> Benjamin *et. al.* ; , *Towards a method for acquiring CIM ontologies*, Int. J. of Comp. Integrated Manufacturing, 8 (3)

Simply speaking, the Ontology of Simulation (or SGA) is just a formal, exhaustive, unambiguous and self-defined description of the whole knowledge this Simulation (SGA) is made of (from WoI to user-code and experiments)

A few requirements an ontology must met

- Clarity of the definitions and the words used ; ambiguity must be precluded
- Decidability : every logical proposition one may drawn from its containt is decidable (either true or false) ; concept of a "logically closed system"
- Consistency : it contains no inference which disclaims it.
  Ontology as a Logical System in Carnap's sense
- Minimality (parcimony) → no plethoric glossary
  Ontology is more than a Glossary : a Glossary contains vocabulary, Ontology contains also Semantic and Syntax

From a mathematical point of view, an Ontology is a Diagram.

Informatical translations of Ontologies are UML class diagrams

The Life-Cycle of SGA

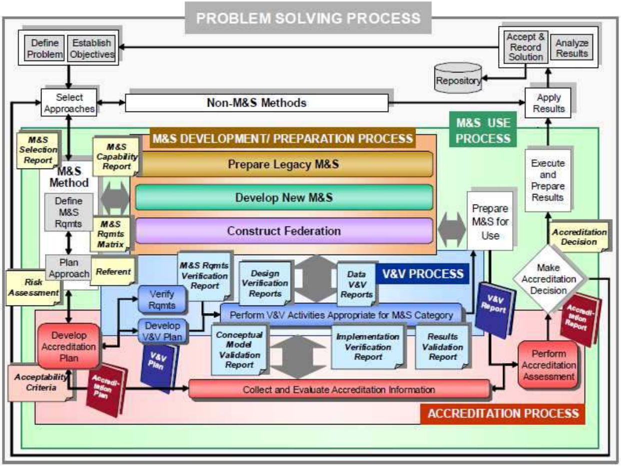
SGA must not begin once the Simulation process is ended ; these two processes must be planed simultaneously

- Important requirement in CS&E where an overall development plan is neccessary
  Ideally tests should be defined in parallel with code specifications
- Software Engineering proposes many strategies and tools to address this question
  - Life Cycle Management (ex. V-Cycle)
  - Continuous Integration (systematic verification and non-regression tests)
  - MDE (Model Driven Engineering) . . .

All these *"matter-of-fact"* questions are very studied and codified in CS&E field

- Tasks scheduling (definition, planning, objectives)
- Definition of roles and responsabilities ("key-players")
- Identification of milestones ; priority ranking tables ; pivotal requirements
- Establishment of advancing reports, traceability, . . .

An exemple of such an overall plan in CS&E



VV&A reports in the overall problem solving process ($\simeq$ Simulation+SGA)

(from MCSO [*Modeling & Simulation Coordination Office*], http ://www.msco.mil/VVA_RPG.html)

My subjective overview

Inheritance from	S&NS	CS&E
Philosophy of Sciences	Many works contain references to PhSci. Some of them use recommandations from it ☺☺	Practically no such thing . . .excepted to belittle PhSci. Most CS&E actors have willingly decided to ignore PhSci ☹☹☹
Formal methods	A few marginal attempts (see above), That field remains largely to reclaim ☹	Widely used in AI, agent-nased modeling Seems to emerge in Simulation ☺/☹
Probability (in the sense of BCT)	Begins to penetrate SN&S (example given next slides) ☺	Likelihood principle prevails Bayesianism is sometimes used but nothing similar to BCT ☹
Pragmatism (codification, regulation, guidelines traceability, . . .)	Doesn't seem to be the main concern (?) Exceptions are in health science ☺/☹	It belongs to the DNA of CS&E ! ! ! ☺☺☺

An illustration of the use of BCT concepts

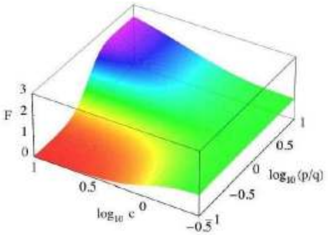
[SKID] Sornette, Kamm, Ide, Davis, *Theory and examples of a new approach to constructive model validation*, in Comp. Uncertainty in Military Vehicule Design, Meeting Proc. RTO-MP-AVT-147, paper 59 (2007)

SKID (same title) NATO RTA AVT-147 Symposium on Computational Uncertainty, Athens 2007

Their "metric" F to assess the increase of "validation" of code $\mathbb{C}$ due to experimental data $\widehat{T}$

with $p = \mathbb{P}(\mathbb{C} \mid \widehat{T})$: $\dfrac{V_{\text{post}}}{V_{\text{prior}}} = F(p,\ q,\ c_{\text{novel}})$

$$F(p,\ q,\ c_{\text{novel}}) = \left[\frac{\tanh\left(p/q + 1/c_{\text{novel}}\right)}{\tanh\left(1 + 1/c_{\text{novel}}\right)} \right]^4$$



- ✓ V_{prior} and V_{post} are "potential utilities" of $\mathbb{C}$ (= provisional validity) before and after observation $\widehat{T}$
- ✓ q : statistical confidence level (p/q scale inspired by that of Bayes Factor [Jeffreys' scale])
- ✓ c_{novel} is an expert-like estimation of the *novelty* of the experiment

Why I say that SKID do BCT (conciously or not) ?

- SKID approach mixes Incremental $\left(V_{\text{post}} \,/\, V_{\text{prior}} \right)$ and Absolute (q) Confirmation

  Absoluteness is needed in decision-making (operating specifications are absolute !)

  "validation problem is fundamentally one of decision theory and not fully objective probabilities alone", SKID, p 59-9

- They account for the confirmatory potential of test $\widehat{T}$ through Novelty coefficient c_{novel}

  Incremental Confirmation (disconfirmation) of $\mathbb{C}$ by $\widehat{T}$ is even greater than c_{novel} is large

  "We view $c_{\textit{novel}}$ as an estimate of the importance of the new data and the degree of 'surprise' it brings", SKID, p 59-11

- They consider repetition of an experiment $\mathbb{T}$ in the same way BCT does

  Suppose $\mathbb{T}$ is repeated twice and gives measures $\widehat{T}_1$ and next $\widehat{T}_2$.

  Within BCT, the confirmatory capability of $\widehat{T}_2$ must be less than those of $\widehat{T}_1$ just because $\widehat{T}_1$ modifies the background knowledge, impliyng that $\widehat{T}_2$ becomes more reliable.

  "Repeating an experiment twice is a degenerated case", SKID, p 59-7

Some fundamental positions in SKID approach

- "*code comparison is not sufficient for validation since validation requires comparisons with* [material] *experiments . . .*"

  "*For validation to remain independent of calibration, it is imperative that these* [calibration and validation] *data sets be disjoints*"

  This conclusion is hardly defended by many other authors

  Trucano et al, *On the role of code-comparisons in verification and validation*, Sandiat Nat. Labs Rep. SAND2003-2752
  Trucano, Swiler, Igusa, Oberkampf, Pilch : complete ref. in BIB.

- If they recognize that "*tests can only determine for certain that a code is not working properly*", they refute the position of Oreskes *et al.* that "*verification and validation of numerical models of natural systems is impossible*", qualifying it as "*newsworthy*".

From more than 30 years, the literature concerning VV&A (US-DoD VV&A) hadn't ceased to grow, now becoming so huge that it is practically impossible to form a synthetic view of what VV&A is today.

I recommend to people who should go further from this talk, to read the book by Oberkampf & Roy (see Bibliography), or to refer to the following site

http ://www.msco.mil/VVA_RPG.html

Most of the important sites are those of institutions from defense sector (US-DoD, SISO, NATO, ATEC, . . .,)

VV&A : Origins

End of the 80's : launch of the ASCI program *(Accelerated Strategic Computing Initiative)* by the US dep. of Defense (DoD) ; two major axes :

- Supercomputers : new technologies (hardware & software)
- Verification, Validation and Accreditation of codes, models and simulations, within Uncertainties framework.

VV&A : Rise of the machine

- Merges DoD, DoE *(US Dep. of Energy)*, AIAA *(American Institute of Aeronautics and Astronautics)*, NASA, ANSI, ISO, ASME *(American Society of Mechanical Engineers)*, NNSA *(National Nuclear Security Administration)*,. . .
- Undoubtedly the main World program on the topic for CS&E activities. Should not be ignored by any actor of this domain . . .
  . . .but it's a bit of a mixed bag, serve with moderation

VV&A : Foundation

"We believe that V&V are specialized processes that respond to the need to use computational simulations in finite periods of time for key policy decisions, not simply to reflect the growth of scientific and analytical understanding of complex physical models."

in Trucano et al. (p-1355) , Calibration, Validation and Sensitvity Analysis, What's What ?

VV&A : Foundation

"We believe that V&V are specialized processes that respond to the need to use computational simulations in finite periods of time for key policy decisions, not simply to reflect the growth of scientific and analytical understanding of complex physical models."

in Trucano et al. (p-1355) , Calibration, Validation and Sensitvity Analysis, What's What ?

VV&A : Judgement Day (or VV&A against Philosophy of Science)

"This position [falsification, confirmation] *is valuable for philosophy of science, but is nearly useless for assessing the credibility of computational results in engineering and technology . . .* [which] *must deal with practical decision making, . . .*
During the last two decades a workable and constructive approach to the concepts, terminology, and methodology of V&V has been developped, but it was based on practical realities in business and government, not in the issue of absolute truth in philosophy of nature."

in Oberkampf & Roy (p-22) , Verification and Validation in Scientific Computing, Cambridge Univ. Press, 2010

Two noticeable inheritances are the awareness that "validation" relies on prediction of the future, not on alignment of past experiments ; and that induction principle is fruitless

VV&A : a Paradigm for SGA in CS&E field

Kuhn has theorized the concequespt of Paradigm to describe a state of *“normal science”* in which knowledge and representation are consensualy and conservatively shared by a community of people working on the same subject.
Kuhn, *The Structure of Scientific Revolutions*, Chicago Univ. Press, (1970)

An exemple : The paradigm of the turbulence initiated by the “K41” Kolmogorov theory

VV&A : a PARADIGM for SGA in CS&E field

Kuhn has theorized the concequespt of Paradigm to describe a state of *"normal science"* in which knowledge and representation are consensualy and conservatively shared by a community of people working on the same subject.
Kuhn, *The Structure of Scientific Revolutions*, Chicago Univ. Press, (1970)

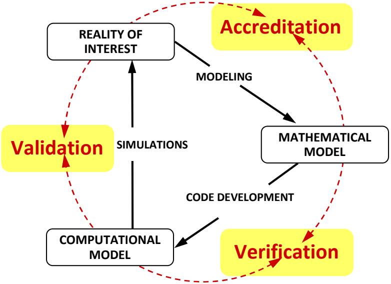
An exemple : The paradigm of the turbulence initiated by the "K41" Kolmogorov theory

Paradigms imply :

- Normativity and Orthodoxy (some kind of *Tables of the Law*)
  At first sight, only slight variants may exist (but look closely to AIAA and DoE defs.)
- Stabilization of thought and resilience under attacks from *exterior* people, low permeability to exterior opinions
- Comfort and safety offered by a well-codified framework
  With counterpart of withdrawal into oneself and few openings towards other thought-styles

THE ICON of VV&A community

(Cincerns Rational branch alone) :



From Schlesinger *et al.* (1979)

Known today as the "Sargent's Cycle"

More sophisticated pictures can be found in the references given at the back of this lecture

What does VV&A look like ?

Roughly speaking VV&A boils down to a set of *Recommended Practice Guides* (RPG)
(to get an idea, take a look to the MSCO site given above)

RPGs say

- What is the place and the role of each actor ; what he has to do and how he has to do it
  Who is responsible for what and to whom ?
- How to manage the process whether you develop a new code, use a ancient one, or a bought one
- What are the milestones, the many reports to write (and how to write them), what he has to do and how he has to do it
- What are the metrics and the technical tools to use, what are "good practices" and what are "bad ones"

Being reassuring, VV&A is somewhat castrating for it allows only for a slight margin of innovation
All in all you're just another brick in the wall . . .(references ?)

How to make your first steps in VV&A

Some difficulties may occur

- VV&A is verbose ; RPGs form a huge collection of informations
  Even if it structured, it remains hard to find your way within it
- Are you sure that VV&A is relevant for your own need in SGA ?
  VV&A will say you nothing about this
- VV&A is (very) costly and require many human resources
  Maybe a lighter specific approach could be sufficient (in this case VV&A will be only a guidance)
- If If you already have your own SGA program, adopting VV&A will force you to make a clean sweep of the past
  For this you will need motivation ...and receptive managers
  It's like Obelix : that's better if you fell into it when you were little [a]

a. french joke meaning : "you're been doing it all your life, it's not surprising you're good at it"

Analyzing VV&A

- Analysis of the sigification

  Does it resist to linguistic (semantic, syntactic and pragmatic) analysis ?

  Is the material consistent, sufficient, exhaustive ... ?

- Analysis of means and solutions offered (RPGs)

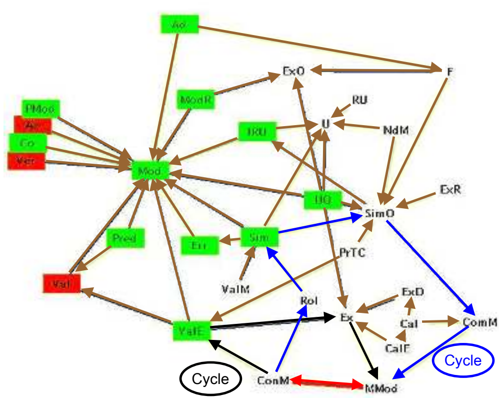
In light of what I said in section 2, I consider that VV&A suffers from some deficiences on the subject of signification (opinion otherwise shared).

> Heath & Hill, *Developping an agent-based modeling verification and validation approach ...*, 2012, URL http ://orsagouge.pbwiki.com/ABSVal
>
> Bair, *The missing link in modeling and simulation validation*, 2013
>
> Bair also quotes that despite 35 years of evolution, the differences between the first representations of VV&A (Sargent, 1979) and the most recent ones, the *"differences are minute"*. This is rightly a feature of khunian paradigms !

Concerning the means : VV&A is very well documented, although detailed solutions are rarely given (excepted for specific VV&A applications [ex. CFD] ... but generalization is not straightforward).

Example of signification deficiences : Graph analysis of a VV&A glossary

Thacker *et al.*, *Concepts of model verification and validation*, LANL Report LA-14167-MS, (2004)



A ⟶ B means "Def . of term A contains term B"

Ac	Accreditation
Ad	Adequacy
Cal	Calibration
CalE	Calibration Experiment
Co	Code
ComM	Computer Model
ConM	Conceptual Model
Err	Error
Ex	Experiment
ExD	Experimental Data
ExO	Experimental Outcomes
ExR	Experiment Revision
F	Fidelity
IrU	Irreducible Uncertainty
MMod	Mathematical Model
Mod	Model
ModR	Model Revision
NdM	Nondeterministic Method
PMod	Performance Model
Pred	Prediction
PrTC	PreTest Calculations
RoI	Reality of Interest
RU	Reducible Uncertainty
Sim	Simulation
SimO	Simulation Outcome
U	Uncertainty
UQ	Uncertainty Quantification
Val	Validation
ValE	Validation Experiment
ValM	Validation Metric
Ver	Verification

Observe the central role of the word "Model" (Mod).

The glossary has also 8 isolated entries (not linked to other entries)

A typical example of non-convergence of VV&A definitions

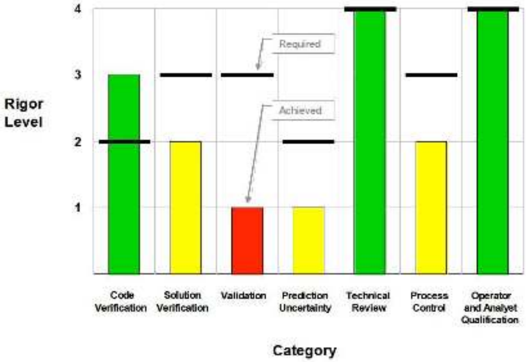
AIAA	*"Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model."*
ASCI	*"Validation is the process of confirming that the predictions of a code adequately represent measured physical phenomena."*
MSCO 1	*"The process of determining the degree to which a model and its associated data are an accurate representation of the real world from the perspective of the intended uses of the model"*
MSCO 2	*"The process of determining the fitness of a model or simulation and its associated data for a specific purpose.."*

A quick look to other concepts (ex. "Model"), would reveal similar non-convergences.

At the end I feel that VVA, while developing essentialy along the pragmatic axis, has forgoten to ensure its semantic consistency.

Of good usage of VV&A prescriptions : the road to "M&S Standards"

Although NASA has adopted VV&A very early, Columbia disaster has forced it to stand back. CAIB was set up to analyze the reasons of this accident ; one main conclusion was that (2003)-VV&A was not sufficient (or mature) to be used blindly as a "Standard".



From Blattnig *et. al.*, *Towards a credibility assessment of models and simulations*, AIAA 2008-2156, April, 2008
Established after the CAIB conclusion reports ; full conclusions accessible on URL http ://caib.nasa.gov

Some common mistakes

- I validated my code !

  Validation is an unreachable ideal, better would be to say "Up to now, my code is confirmed" (or not falsified)

- The code is validated because measure agrees with prediction

  No, it's just confirmation that the code is provisionnaly valid (= usable) !

- The code is confirmed for I was able to tune it to represent old experiments

  The most common mistake : prediction has to be done independently (previously) of the measurement

- One can do a Confirmation process without doing firstly the Verification one

  Confirmation should only been assessed once the code has been verified !

- When doing code confirmation, only the code matters

  No, the code can't be isolated from the context within which it is embedded.

- One can confirm a code without accounting for uncertainties

  Is it necessary to spend time to prove this is pure nonsense ?

Thoughts

My goal was to show you that are no easy answers to the question of validation

- for it depends on the field within which you operate
- for "validation" is essentially a multidisciplinary activity
  Where formal research takes place (mathematics, probability, logics, ...)
  And informal considerations too (social, historical, political, ...)

Forget the comfortable idea that there could be a ready-made answer you could deal with.
Be suspicious to simple opinions, especially such like "I know how to do ..."

Guidelines exist to help you, but YOU have to develop YOUR own "validation program" (aka SGA above) for YOUR own needs.

Thoughts

My goal was to show you that are no easy answers to the question of validation

- for it depends on the field within which you operate
- for "validation" is essentially a multidisciplinary activity
  Where formal research takes place (mathematics, probability, logics, . . .)
  And informal considerations too (social, historical, political, . . .)

Forget the comfortable idea that there could be a ready-made answer you could deal with.
Be suspicious to simple opinions, especially such like "I know how to do . . ."

Guidelines exist to help you, but YOU have to develop YOUR own "validation program" (aka SGA above) for YOUR own needs.

At the last, remember that Validity is not Truth, *The truth is elsewhere*

Step by step recommended bibliography

Level 1 : some basic texts from S&NS and CS&E

- Oreskes, Shrader-Frechette, Belitz, *Verification, validation and confirmation of numerical models in the earth sciences*, Science, 264, (1994)
- Barlas, Carpenter,*Philosophical roots of model validation : two paradigms*, System Dynamics Review, 6(2), (1990)
- Sornette, Davis, Ide, Kamm, *Theory and examples of a new approach to constructive model validation*, in Comp. Uncertainty in Military Vehicule Design, Meeting Proc. RTO-MP-AVT-147, paper 59, (2007)
- Post, Votta *Computational science demands a new paradigm* Physics Today, 58, (2005)
- Refsgaard, Henriksen, *Modelling guidelines - A theoretical framework*, in Refsgaard Ed., *HarmoniQUA - State-if-the-Art Report on quality assurance in modelling related to river basin management* (2002)
- Konikow, Bredenhoef, *Ground water models cannot be validated*, Advances in Water Ressources, 15, (1992)
- Stevenson, *A critical look at quality in large-scale simulations*, Computing in Science and Engrg, (1999)
- Pedersen, Emblemsvag, Bailley, Allen, Mistree, *Validating design methods & research : the validation square*, Proceedings of DETC'00 2000 ASME Design Engineering Technical Conferences September 10-14, (2000), Baltimore, Maryland

Step by step recommended bibliography

Level 2a : more advanced texts

- Babuska, Tempone, Nobile, *A systematic approach to model validation based on Bayesian updates and prediction related rejection criteria*, Comput. Methods Appl. Mech. Engrg., 197, (2008)
- Tempone, *A Verification and Validation (V&V) framework for Computational Science* SANDIA CSRI Workshop on Mathematical Methods for Validation and Verification, August 14, (2007)
- Easterling, Berger , *Statistical Foundations for theValidation of Computer Models*, Presented at Computer Model Verification and Validation in the 21st Century Workshop, Johns Hopkins University (2002)
- Mahadevan, Rebba, *Validation of Reliability Computational Models using Bayes Networks*, Reliaility Engrg and System Safety, 87, (2006)

Step by step recommended bibliography

Level 2b : texts from VV&A community

- Oberkampf, Trucano, *Validation methodology in computational fluid dynamics*, AIAA Technical Report 2000-2549, (2000)
- Trucano, Swiler, Igusa, Oberkampf, Pilch, *Calibration, validation, and sensitivity analysis : what's what ?*, Reliability Engrg. Systems Safety, 92, (2006)
- DoD, DoD Directive No. 5000.61, *Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)*, Defense Modeling and Simulation Office, www.dmso.mil/docslib
- Hills, Dowding, *Statistical validation of engineering and scientific Models : bounds, calibration, and extrapolation*, SANDIA rep., SAND2005-1826, (2005)
- Thacker, Doebling, Hemez, Anderson, Pepin, Rodriguez *Concepts of model verification and validation*, Los Alamos National Laboratories Report, LA-14167-MS (2004)
- Oberkampf, Roy, *Verification and Validation in Scientific Computing*, Cambridge University Press (2010)

Step by step recommended bibliography

Level 3 : Philosophical texts

- Fitelson, *Studies in bayesian confirmation theory*, PhD, University of Wisconsin, Madison, (2001)
- Morgan, *Experiments versus models : New phenomena, inference and surprise*, Journal of Economic Methodology, 12(2), (2005)
- Cozic, *Confirmation et induction*, IHPST, Cahier de Recherche DRI-2009-02, (2009)
- Clarke, *The necessity of being comparative. Theory cnfirmation in quantitative Political Science*, Comparative Political Studies, 40(7), (2007)
- Benjamin, Patki, Mayer, *Using ontologies for simulation modeling* Proceedings of the 2006 Winter Simulation Conference, (2006)
- Blais, *Toward a verification, validation, and accreditation (VV&A) ontology*, Naval Postgraduate School, Monterey, California, Rep. NPS-MV-08-003, (2008)
- Beven, *Towards a coherent philosophy for modelling the environment*, Proceedings of the Royal Society A, London, 458, (2002)
- Rosenbaum, *Leï£¡ons d'introduction ï£¡ la philosophie des sciences*, Les Presses de L'ENSTA, Paris, (2009)

Step by step recommended bibliography

LEVEL 3 : Philosophical texts (continuation)

- Kleindorfer, O'Neill, Ganeshan *Validation in simulation : various positions in the philosophy of science*, Management Science, 44(8), (1998)
- Klein, Herskowitz, *Philosophical foundations of computer simulation validations*, Simulation & Gaming, 36, (2005)
- Kï£¡ppers, Lenhard, *Validation of Simulation : Patterns in the Social and Natural Sciences*, Journal of Artificial Societies and Social Simulation, 8(4), (2005)
- Petersen, *Simulating Nature : a philosophical study of computer-simulation uncertainties and their role in climate science and policy advice*,Het Spinhuis Publishers, Amsterdam, (2006), http ://dare.ubvu.vu.nl/bitstream/1871/11385/1/5536.pdf
- Simpson, *Simulations are not models*, (2006) http ://philsci-archive.pitt.edu/archive/00002767/01/SimAreNotModelsRD7.pdf
- Stewart (2005), *Notes for the development of a philosophy of computational modelling*, Carleton University Cognitive Science, Tech. report 2005-04, (2005), http ://www.carleton.ca/ics/TechReports

Step by step recommended bibliography

Two (rather old) annoted bibliographies

- Gruhl, Gruhl, *Methods and examples of model validation : an annoted bibliography*, MIT Energy Lab. Working pPaper MIT-EL 78-022WP, (1978) Physics Today, 58, (2005)
- Hamilton, *Model validation : an annoted bibliography*, Communications in Statistical Theory Methods, 20(7), (1991)

Addendum 1 : detailed comparison of SKID and BCT (1/2)

BCT	SKID	Observations
$\mathfrak{C}(\mathbb{C}, \widehat{T})$ measures how $\widehat{T}$ increases (decreases) the firmness of $\mathbb{C}$; $\mathfrak{C}(\mathbb{C}, \widehat{T})$ comes from Kolmogorov axioms of probability theory	Prior to $\widehat{T}$, $\mathbb{C}$ has a potential trust value V_{prior} which, after observation of $\widehat{T}$ becomes V_{post}	For SKID neither V_{prior} nor V_{post} are defined as probabilities.
Ex : $\mathfrak{C}(\mathbb{C}, \widehat{T}) = \mathbb{P}((\mathbb{C} \mid \widehat{T}) - \mathbb{P}((\mathbb{C})$ Other examples given above	$\frac{V_{\text{post}}}{V_{\text{prior}}} = F(\mathbb{P}(\mathbb{C} \mid \widehat{T}),\ q,\ c_{\text{novel}})$	Likelihood $\mathbb{P}(\mathbb{C} \mid \widehat{T})$ here has a central role (c_{novel} discussed above) ; q denotes a reference likelihood (risk, above noted α). So SKID mixes incremental and absolute confirmation.
Again : confirmation is a matter of probability. Inherited approach of the logic of scientific explanation (induction, deduction, see also Popper)	*"validation problem is fundamentally one of decision theory and not fully objective probabilities alone"* SKID, p 59-9	Read Kelly, Glymour, *Why Bayesian confirmation does not capture the logic of scientific justification* tech. rep. CMU-PHIL-138 (2003)

Addendum 1 : detailed comparison of SKID and BCT (2/2)

BCT	SKID	Observations
BCT highlights the prior confirmatory power of test $\mathbb{T}$	c_{novel} is a subjective evaluation of the "novelty" of the test $\mathbb{T}$: *"how well the new observation explores novel 'dimensions' of the parameter/variable space"* SKID, p 59-7	A same need to quantify $\mathbb{P}(\widehat{T})$. Prior confirmatory power is highly related to the Bayesian solutions of the Paradoxes of the Ravens
BCT keeps reusing a test (or a replicate) in assessing confirmation. ⇒ Old Evidence Problem	*"Repeating an experiment twice is a degenerated case"* SKID, p 59-7 and farther *"Providing a value for c_{novel} ... remains a difficult and key step in validation process"*	Emphasis here on the following pb. : if test $\mathbb{T}$ gives obs. $\widehat{T}$ which confirms $\mathbb{C}$, does it make sense to replay $\mathbb{T}$ (thus obs. $\widehat{T}'$) to enhance the confirmation of $\mathbb{C}$? Cozic, *Le problème des données connues (old evidence)* séminaire IHPST, 1 février 2008
$\mathbb{P}(\mathbb{C}\,\vert\,\widehat{T}) = \frac{\mathbb{P}(\widehat{T}\,\vert\,\mathbb{C})}{\mathbb{P}(\widehat{T})}\mathbb{P}(\mathbb{C})$ may be large if $\mathbb{P}(\widehat{T})$ is highly unlikely to appear (surprise effect)	*"We view c_{novel} as an estimate of the importance of the new data and the degree of 'surprise' it brings"* SKID, p 59-11	BCT and SKID share a same concern to account for the selectivity of the test $\mathbb{T}$

Example from S&NS : Validation in Ecosystems studies

Source : Rykiel, *Testing ecological models : the meaning of validation*, Ecological Modelling, Vol. 90, 1996

- Levins (1966) *"Unlike the scientific hypothesis, a model is not verifiable directly by experiment. ...The validation of a model is not that it is 'true' but that it generates good testable hypotheses relevant to important problems."*

- Goodall (1972), validation tests the agreement between a model and the real system. Are predictions reliable ?

- Overton (1977), validation is an integral part of the (iterative) M&S process ; It must be a development constraint.

- Holling (1978), A model is a set of hypotheses that can only be refuted.

- Shugart (1984), "[validation is a set of] *procedures, in which a model is tested on its agreement with a set of observations that are independent of those observations used to structure the model and estimate its parameters*"

Example from S&NS : Validation in Hydrogeology

- Matalas *et al* (1982) : A model has no necessarily deductive capacity

- Beven *et al* (1989) : environmental models cannot be validated but only confirmed.
  Equifinality : many different models may have similar predictive capabilities.

- Konikow & Bredehoeft (1992), *"the terms validation and verification have little or no place in groundwater science ; these terms lead to a false impression of model capability"*

- Oreskes *et al* (1994) : "*verification* [= establish the Truth] *is only possible in closed systems*" ; "*In contrast ..., the term validation does not necessarily denote an establishment of truth ...Rather, it denotes the establishment of legitimacy*"

- Rykiel (1996) "*Validation is just one component of the larger task of model evaluation.*" ;
  Theoritical Validity is out of reach, only is Pragmatic Validity
  "*Validation describes ...a testing process on which to base an opinion of how well a model performs so that a user can decide whether the model is acceptable for its intended purpose*"

Last example : Validation in CS&E

- Shannon (1975) : "[Validation is] *the process of bringing to an acceptable level the user's confidence that any inference about a system derived from the simulation is correct*"

- AIAA (1998) : "*Validation is the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.*"

- Stevenson (2000) : "*Validation is the process by which we attempt to convince ourselves that the simulations correctly capture the model and have some relation to an observable world. But models are not reality !*"

- US DoE (2004) : "*Validation (ASC) is the process of confirming that the predictions of a code adequately represent measured physical phenomena.*"

- Revva/Thales (2004) : "*Validation answers the question of whether it is impossible to distinguish the model and the system in the experimental frame of interest*"
  Here : System denotes material (true) system, the behaviour of which the Model mimics.

continued next slide

Addendum 2 : Some famous technological failure and their analysis)

From Bahill & Henderson, *Requirements Development, verification, and validation exhibited in famous failures*, Systems Engineering, **8**, num 1, (2005)

RD : Requirements Development ; VER : Verification ; VAL : Validation

■ (red) : not met ; ■ (green) : met

System Name	Year	Putative cause of failure	RD	VER	VAL
HMS Titanic	1912	Poor quality control	not met	not met	met
Tacoma Narrows Bridge	1940	Scaling up an old design	met	met	not met
Ford Edsel	1958	Failure to discover customer needs	met	met	not met
Apollo-13	1970	Poor configuration management	met	not met	met
Concorde SST	1976-2003	It was not profitable	met	met	not met
IBM PCjr	1983	Failure to discover customer needs	not met	met	met
GE refrigerator	1986	Inadequate testing of new technology	met	not met	met
Space Shuttle Challenger	1986	Bureaucratic mismanagement	met	not met	not met
Chernobyl Nuclear Power Point	1986	Bad design and risk management	met	met	not met
A-12 Airplane	1980s	Failure to develop realistic requirements	not met	not met	not met
Hubble Space Telescope	1990	Lack of total system test	met	not met	met
Super Conducting Super Collider	1995	Cost overruns, lack of public support	met	met	not met
Ariane 5 Missile	1996	Incorrect reuse of soft., faulty scaling up	met	not met	not met
Lewis Spacecraft	1997	Design mistakes	met	met	not met
Motorola Iridium System	1999	Misjudged competition, mispredicted technol.	met	met	not met
Mars Climate Orbiter	1999	Use of different units	not met	not met	not met
Mars Polar Lander	2000	Failure of middle management	met	not met	met
Space Shuttle Columbia	2002	NASA corporate culture, lessons learned	met	not met	not met
Northeast power outage	2003	Lack of tree trimming	not met	met	met