The Validation issue in Modeling and Simulation

Marc Sancandi

CEA/CESTA-DSGA

Date

Advertisement

The first way to talk about validation is to deal only with a specific application
But there are as many domains than there are (types of) validation ...which makes tranposition to your problem sometimes challenging

---

Another way is to talk about validation in a broader, "neutral" way
I'm not sure that it will make transposition easier ... but I believe you'll gain by that way a better understanding of validation and that will preserve you from many traps.

That's the choice I made for this talk

What is this talk intended for ?

My first goal is to break with the naive idea that Validation is just Business Process, that is

> *"a collection of related, structured activities or tasks that produce a specific service or product . . .for a particular customer".*
> source http ://en.wikipedia.org/wiki/Business_process

So I will show you that

- validation fits into a long history which began well before that people thought in terms of business or magement : Validation cannot simply be seen as business process (*red tape*)
- a well-constructed validation strategy needs rigourous development methods,

My second goal is to refute the idea according to which validation could be a consensual well-adopted concept,

To avoid misunderstandings, let me clarify where my talk takes place

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

The world according to Post and Votta

Recently, Post and Votta advocated that computational science (CS) has to meet 3 challenges
Computational science demands a new paradigm, Physics Today, 35, January 2005

- the performance challenge : producing high-performance computers,
- the programming challenge : programming for complex computers,
- the prediction challenge : developing truly predictive complex application codes.

and claimed that the last one *"is now the most serious limiting factor for computational science."*

As a matter of fact, 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 ?

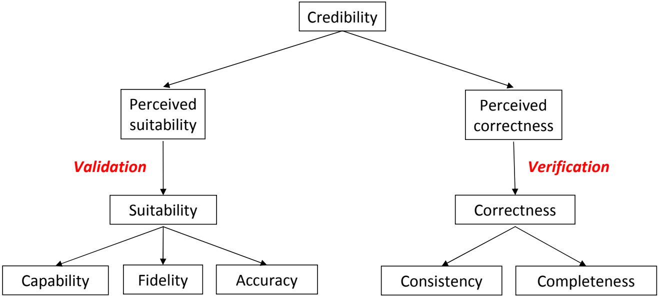
The answer is an essential component of any decision process . . .
. . .but not the only one (need to consider also strategic, politic, social, historic contexts)

In the beginning was the Credibility

Credibility is a high-level agregated notion, based on validation, verification, and human faith (accreditation)

For verification see François Hémez course in this Summer School

Credibility concept and its derived concepts



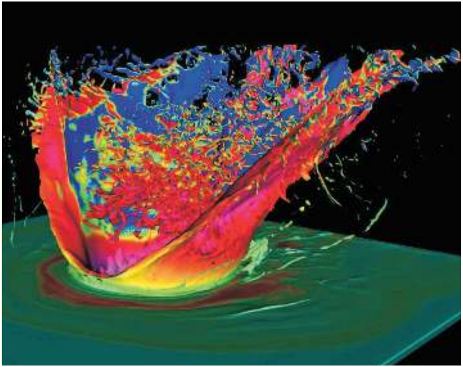
From Brade (p 14), *"A generalized process for the verification and validation of models and simulation results"* , PhD Universität der Bundeswehr München, 2003

Contexts differ from a field to another

	Geologist	Epidemiologist	Engineer
Purpose	Understand the Chicoxulub meteorit impact (Yucatan, 65 million years ago)	Contain the spreading of some virus (ie. H1N1 flue)	Conceive some industrial product
Intended Uses	Just understand	Minimize the number of infected people	Maximize performances while minimizing costs
Model-type	Purely Theoritical (from phys. principles and laws)	Particle models + empirical models (ev. stoch. models)	Theoritical models + empirical ones
Empirical grounds	Only one observation ; deduction, hypothetizing	Real time observations of spreading evolution	Many dedicated experiments
Val^ion expe.	I hope not ! ! !	The end of the infection	Many experiments
"Rewinding"	No of course	No, unfortunatly	Yes of course
Credibility enhanced	If new discoveries confirm theoritical computations	? ? ?	If experiments agree with numerical simulations
External constraints	No one	Public opinion, Media, Crisis management	Markets ?

The famous Chicoxulub meteorit impact

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



Reported here because of its beauty . . . but was it reality ?

What do you really want to validate (decision-maker point of view) ?

A Material Concept (MC) (i.e. 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

An Intellectual Concept (IC) (natural system, non-manufactured)

- Ozone layer damage caused by CFC, Chicoxulub impact, Supernova blast
  Here the system is neither reachable nor man-controllable
  Here no tests can be done
- Also (up to some extents) : portfolio management, infection spreading

When is "code validation" really necessary ?

When no other possibility exists (IC or MC one if we can't explore it only by material experiment)

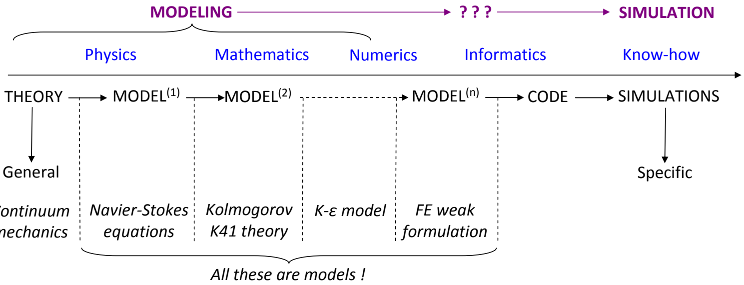
Then, we need to assess if code $\mathbb{C}$ is able to predict correctly observed situations (IC) or (MC) new "validation" tests (hereafter ValExp)

For this last situation :

- ValExp must be close to intended use configurations of MC ... And, ideally, we would like they be close to calibration experiments (CalExp)
  - Closer are ValExp and CalExp, HIGHER is the faith in the capability of $\mathbb{C}$ to correctly predict ValExp
  - But at the same time, LOWER is the faith in its ability to predict really new situations.
- Having supposed that material investigation alone wasn't possible, this implies that CalExp is necessarily far from ValExp

Are you sure that "code validation" is a clear expression ?

Point of view	Modeling	Simulation
Validation target(s)	Model . . .but what model ?	Simulations !



Fundamental questions are :

- ✓ What are models and simulations ?
- ✓ What are the links between them ?
- ✓ What are models concerned by validation ?
- ✓ What role does the code play here ?

What is a Model?

Do everyone agree on a common definition?
"The term 'model' is surely one of the most contested in all of philosophy of science", Godfrey-Smith, *The strategy of model-based science*, Biol. Philos., 21, (2006)

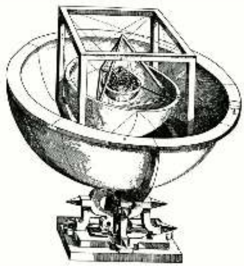
There are roughly three types of models
Frigg, *Models in science*, Stanford Encyclopedia of Philosophy, URI http ://plato.stanford.edu/entries/models-science

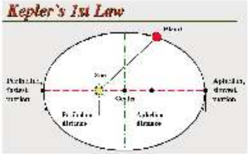
- Models of phenomena : scale models, idealized models (*ie.* point masses moving on frictionless planes), analogical models (*ie.* the billiard ball model of a gas), phenomenological models (*ie.* the nucleus as a liquid drop) ;(Do rats model human beings?)
- Models of data : typically all statistical models!!!...
- Models of theory : come from very general physical laws and principles, through rigourous mathematical derivations

... leading to three different meanings of validation

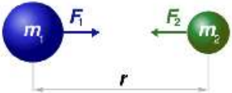
Ptolémée

Kepler





Newton



$$F_1 = F_2 = G\frac{m_1 \times m_2}{r^2}$$

Models of Data

Models of Theory

What is Simulation ?

Simulation : *"A process which mimics the relevant features of a target process"*

Hartmann, *The World as a process*, in Heselman, Müller & Troitzsch, *Modelling and simulation in the social sciences from the philosophy of science point of view*, 1996

Or simply a "numerical experiment" done with a computer (?)

Morrison, *Models, measurement and computer simulation : the changing face of experimentation*, Philosophical Studies (2009) 143,

What kind of simulations do we speak about ?

- A complete surrogate to unreachable "material" experiments :
  - ✓ Understand : Astrophysics, Genetics, . . .
  - ✓ Mimic : Flight simulator, *Artificial Life*, . . .
- A partial surrogate to expensive material experiments :
  - ✓ Numerical Design, assessment to health policy, . . .

> *"Ideally, one wishes to prove that a new industrial design satisfies given regulatory requirements only upon a computational analysis. Can we trust the results ?"*
> Tempone, *A verification and validation framework for computational science*, SANDIA CSRI Workshop on Math. Meth. for V&V, 2007

A zoology of Simulation

- **Continuous models**

  Simulation of systems governed by physical principles/laws (ODE, PDE, ...)
  - ✓ CFD, hydrodynamics, Electromagnetism, ...
  - ✓ Climate, Wind tunnel, Crash-test, Tsunami, ...

- **Discrete event models**

  Simulation of systems governed by by elementary transitions
  - ✓ Ising-like models, Cellular automata, Lattice gas, *Agent-Based Modeling*
  - ✓ Queuing, Urban traffic, Populations dynamics, Economy, Process management
  - ✓ Forrest fires, vitrification, coalescence, molecular dynamics, ...

- **Simulation** or **Simulacrum** ?

  Farge, *Numerical experimentation : a third way to study Nature* in Frontiers of Computational Sciences Kaneda, Kawamura & Sasai (Eds.), Springer, 2007
  - ✓ Simulacrum : Game Simulations !

Where M&S validation stances disagree

M&S [3]	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.

Hodges & Dewar defined 4 criteria the modeled situation must satisfy to hope to validate the model (simulation of weapons systems in battlefield scenarios)

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

1. *It must be possible to observe and measure the situation being modeled.*
2. *The situation being modeled must exhibit a constancy of structure in time.*
3. *" " " a constancy across variations in conditions not specified in the model.*
4. *it must be possible to collect ample data whith which to make predictive tests of the model.*

Some years later Oreskes claimed that *"Models in social and policy sciences generally fail to satisfy these criteria and therefore cannot be validated"*

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

Oreskes, Shrader-Frechette, Belitz : *" Verification and validation of numerical models of natural systems is impossible. This is because natural systems are never* [logicaly] *closed and because model results are always non unique."*
Verification, validation, and confirmation of numerical models in the Earth sciences, Science, Vol. 263, 1994

As many validations as there are *"validators"*

The key point is that there is no unique answer to what validation means !

> *"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 given below.

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

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*"

Verification is assessment of Truth ; while Validation asks *"The model works enough for me"*

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

Last example : Validation in CS&E

I am partial to this one :

> [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 ...

First encounter with VV&A (CS&E field)

VV&A : Verification, Validation and Accreditation

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(ting), hard./soft, languages, software engineering, ...
- Verification and Validation of codes, models and simulations, under Uncertainties.

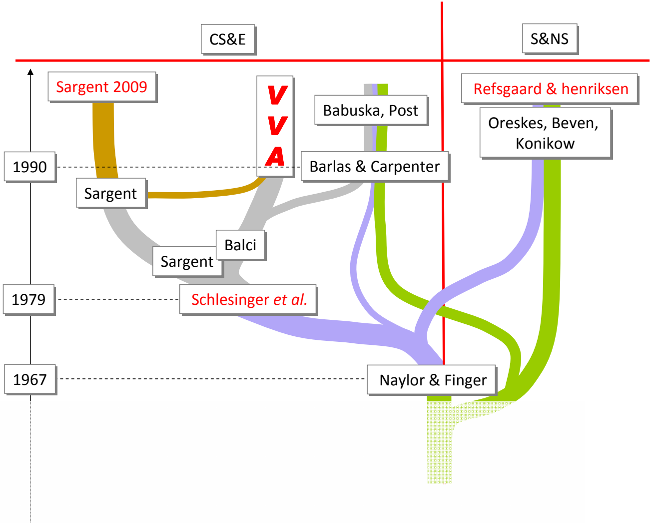
  in other words, the beginnings of VV&A

VV&A today

- 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)*,. . .
- The major world sponsor of these themes in CS&E, more generally in M&S (*Modeling & Simulation*)

  No one actor in this domain would have to ignore this program !

A simplified (and subjective) genealogy of main trends in Validation



A casual remark

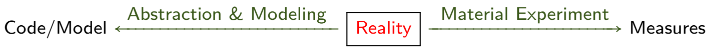
Note the surprising disymmetry of all the previous definitions :

- ✓ Why did they claim the *predictor* is "valid" if it agrees with experimental observations ?
- ✓ Couldn't we inverse the defs. and decide that measure is "valid" if it agrees with predictions ?
- ✓ If we are more confident in the *predictor* than in the meas. devices, that's probably a better attitude !

What I advocate is to consider symmetrically the *predictor* and the measure

- ✓ Anyway, aren't predictions considered as "numerical" experiment ?
- ✓ A comparable argument may be advanced in bayesian calibration.

Each kind of experiment brings on Reality its own information and then, if "validity" occurs, they must be compatible.



This suggests two main changes in our way to tackle validation :

1. Replace the concept of validity by another one like *consistency*, or *compatibility*, . . .
2. Take into account that the true target of validity is the couple *predictor*+measure

A casual remark (1/2)

1. The Simulation, the Experimental and the Real Worlds make up a Whole
2. There is no way to validate simulations without accounting for this Whole.
3. So, never forget that the Code is not an isolated object (*ie a closed system*)

I would like you keep this in mind all along this talk.

Finaly : any validation program develops in three directions

1. Organizational *(not evoked here)*
   - ✓ Identification of tasks, teams, objectives, traceability, . . .
   - ✓ A "paperwork" job . . .but a fundamental one !
2. Technical (by opposition to the previous one)
   - ✓ Here we are interested in answering theoretical questions
   - ✓ State the problem : Conceptualization + Specification, see below
   - ✓ 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, . . .

NOTICE

We'll see later that words validation, or validates are unsuitable to convey the idea that a test *"validates"* a code

To avoid any misunderstanding, and although they have strict meannings (I will precise later) I will use in this section the words

CONFIRMS instead of VALIDATES

INFIRMS instead of DOESN'T VALIDATE

- The Code/Model $\mathbb{C}$ predicts that test $\mathbb{T}_1$ WOULD GIVE the value T_1
- Doing $\mathbb{T}_1$ GIVES IN FACT observation $\widehat{T}_1$
- Choose a "metric" $\mathfrak{C}(T_1, \widehat{T}_1, \mathbb{K})$ to "decide for the validity" of $\mathbb{C}$ from the agreement between T_1 and $\widehat{T}_1$ according to the (prior) knowledge $\mathbb{K}$.
- Then (notional exemple) we can decide, for a given value of α that :
  - "The result of $\mathbb{T}_1$ CONFIRMS T_1 if $\mathfrak{C}(T_1, \widehat{T}_1, \mathbb{K}) > \alpha$
  - "The result of $\mathbb{T}_1$ INFIRMS T_1 if $\mathfrak{C}(T_1, \widehat{T}_1, \mathbb{K}) < \alpha$

The confidence in the capabilities of the code to properly predict the real behaviour increases as the number of $\boxed{\mathfrak{C}(T_n, \widehat{T}_n | \mathbb{K}) > \alpha}$ situations increases to.

Some constraints on “validation” tests

- Every $\mathbb{T}_n$ must be
  - measurable (observable) ! ! !
  - representative of the situation of interest, meaning that $\mathbb{T}_n$ is representative of the operational conditions the system is thought to encounter.
  - discriminant, in the sens that a result $\widehat{T}_n$ close to T_n must not be in the bag.

    The objective of any scientific researcher should be to honnestly try 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

The nature of the Uncertainties $\mathbb{U}$ (1/3)

p 376 from : Parker, *Computer simulation through an error-statistical lens*, Synthèse, 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

The nature of the Uncertainties $\mathbb{U}$ (2/3)

Formally any code may be seen as an application $\mathbb{C} : \mathscr{Z}_C \longmapsto \mathscr{Y}_C$

Let's see reality $\mathbb{R}$ in the same way $\mathbb{R} : \mathscr{Z}_R \longmapsto \mathscr{Y}_R$

To simplifiy I will consider that $\mathscr{Z}_F = \mathscr{Z}_R = \mathscr{Z}$ and $\mathscr{Y}_F = \mathscr{Y}_R = \mathscr{Y}$

Meaning by the fact that $\mathbb{C}$ could simulate $\mathbb{R}$ up to any arbitrary accuracy

Case 1 : The test $\mathbb{T}$ (with regulation "Take $\mathscr{Z} = Z$") is noisy :

- True test condition Z_{true} differs from prescribed Z ; only a measurement $\widehat{z}$ of Z_{true} is known

  $\widehat{z}$ may be thought as a sample of some R.V. $\widehat{Z}$

- $\mathbb{T}$ delivers $\mathscr{Y} = Y_{\text{true}}$ (here again Y_{true} is unknown)

  Measurement $\widehat{y}$ of Y_{true} may be thought again as a sample of some R.V. $\widehat{Y}$

- Let us remark that there is no reason to have $\widehat{y} = \mathbb{C}(\widehat{z})$, even if $\mathbb{C}$ perfectly mimics $\mathbb{R}$! ! !

The nature of the Uncertainties $\mathbb{U}$ (3/3)

Cas 2 : Now assume that $\mathbb{C}$ is imperfectly known

2a $\mathbb{C}$ may be a stochastic code (ray-tracing, neutronics, Molecular Dyn., ...)
In that case the output $Y = \mathbb{C}(Z)$ may be modelled as a R.V.

2b $\mathbb{C}$ is deterministic, but when splitting $\mathscr{Z}$ as $(\mathscr{X}, \mathscr{U})$ only $\mathscr{X}$ is assumed known.

Example : $\mathscr{U}$ represents the infered parameters in the calibration process of $\mathbb{C}$.

U may then be modelled as a R.V., so $Y = \mathbb{C}(X, U)$ becomes a R.V..

If we look at $\mathbb{C}(\mathscr{X}, \mathscr{U})$ as a family of codes $\mathbb{C}_{[\mathscr{U}]}(\mathscr{X})$ indexed by $\mathscr{U}$, the random character of U is now reported on $\mathbb{C}$!

So we returned in the situation "$\mathbb{C}$ is stochastic"

A more general situation mixes cases 1 and 2

About the choice of $\mathfrak{C}$ (1/2)

I write simply $\mathfrak{C}(T, \widehat{T} | \mathbb{K})$ under the form $\mathfrak{C}(\mathbb{C}, \widehat{T})$

A probabilist will naturaly think of the Likelihood and will choose $\mathfrak{C}(\mathbb{C}, \widehat{T}) = \mathbb{P}(\widehat{T} | \mathbb{C})$
Or some other posterior measure if he is a bayesian.

According to many researchers, a proper definition of confirmation must encompass the intuition that

> $\widehat{T}$ confirms the code $\mathbb{C}$ if we believe stronger in $\mathbb{C}$ AFTER $\widehat{T}$ is obtained, than we believed in it BEFORE the observation of $\widehat{T}$

This militates for an Relative (incremental) "metric" of Confirmation instead of an Absolute one
(think also to bayes factor and odds in bayesian model-selection strategies)

This comment raises the following question

> *Do probabilities give appropriate "metrics" for assessing confirmation ?*

We'll see below that the answer is NO

About the choice of $\mathfrak{C}$ (2/2)

Some common examples of incremental "metrics"

$$\mathfrak{C}(\mathbb{C},\, \widehat{T}) = \mathbb{P}(\mathbb{C}\,|\,\widehat{T}) - \mathbb{P}(\mathbb{C})$$

$$\mathfrak{C}(\mathbb{C},\, \widehat{T}) = \mathbb{P}(\mathbb{C}\,|\,\widehat{T}) \,/\, \mathbb{P}(\mathbb{C})$$

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

$$\ldots$$

Some of them remain highly notional and restricted to finite discrete probability spaces :

- How can we assess $\mathbb{P}(\mathbb{C})$ when a continuum of $\mathbb{C}$ exists ?
- And again, what's the meaning of $\mathbb{P}(\mathbb{C}\,|\,\neg\widehat{T})$?
- More fundamentaly, what's the meaning of $\mathbb{P}$ (cdf, pdf, ...) ?

We will return later on these "metrics". For the moment let's just say that their construction is really challenging

- $\mathfrak{C}$ must reflect common sense on confimation
- while measuring some "absolute faith" in $\mathbb{C}$
  (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 slight 10% increase from 0.1 to 0.11 be better ?)

A personal contribution to the debate

I evoked above the symmetry between Simulation and Material Experiment (Measure) ; this can be formalised in the following way :

- Simulation carries information $\mathbb{I}(\mathbb{R}\,|\,s,\,\mathbb{K}_S)$ on Reality $\mathbb{R}$; s is the simulation result and $\mathbb{K}_S$ knowledge used to form it
- Symmetricaly, Measure carries information $\mathbb{I}(\mathbb{R}\,|\,m,\,\mathbb{K}_M)$ on $\mathbb{R}$; m measure, $\mathbb{K}_M$ knowledge

> According to the orthodox position, the prediction s should be as *"valid"* as the consistency between $\mathbb{I}(\mathbb{R}\,|\,s,\,\mathbb{K}_S)$ and $\mathbb{I}(\mathbb{R}\,|\,m,\,\mathbb{K}_M)$ is large.

Why did we decide that the consistency of $\mathbb{I}(\mathbb{R}\,|\,s,\,\mathbb{K}_S)$ and $\mathbb{I}(\mathbb{R}\,|\,m,\,\mathbb{K}_M)$, gave us confidence in the value of s ?

Symmetry could suggest that :

> *"Consistency gives us confidence in the measure m but says nothing about s !"*

At last : a step through decision taking

How can we manage the situation "$\widehat{T}$ infirms $\mathbb{C}$" ?

First of all, it doesn't imply the rejection of the code :

- Just consider that the test doesn't increase our belief in the code (optimistic)
- Or if you are pessimistic that the test makes the code suspicious

So, there are three standards (and well discussed in ΦoS) positions we can adopt

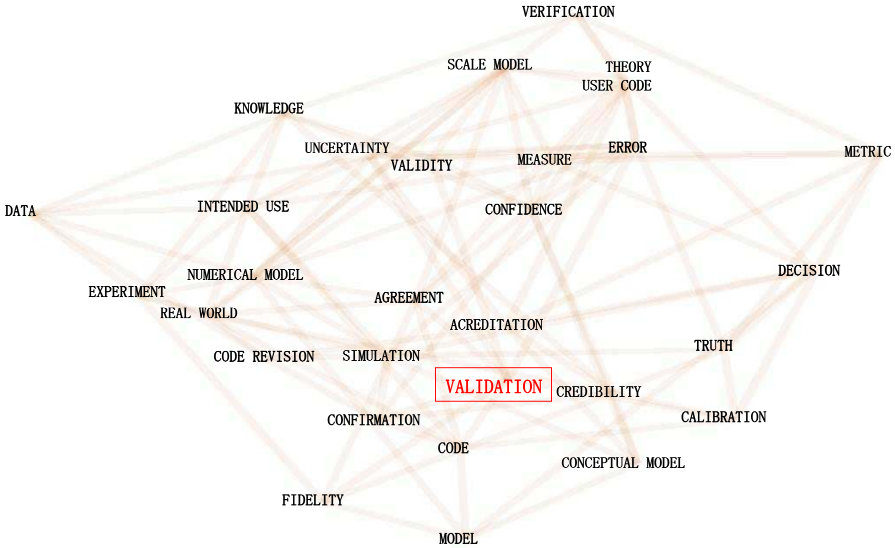
- Purist : we reject the code and start developping a new one (expensive of course)
- Pragmatic : we reject only if we have a better one. Try to invoke some protective argument lying on the possible falsity of some Auxiliary Hypotheses
- Kuhnian : anyway, this is not a technical question but a matter of social, historical, political conventions !
  From Thomas Kuhn, *Structure of scientific revolutions* Second Edition, Chicago University Press, (1970)

Any rational attempt to approach validation should :

- Explain what is validation : is it THE correct concept ?
- Draw up the inventory of all the entities that exist inside the concerned domain
  - ✓ "define" them in an unambiguous manner
  - ✓ exhibit their salient properties
- Establish the salient relationships between these entities

To sum up, to approach validation begins by organizing this mess

A random graph by Maple©



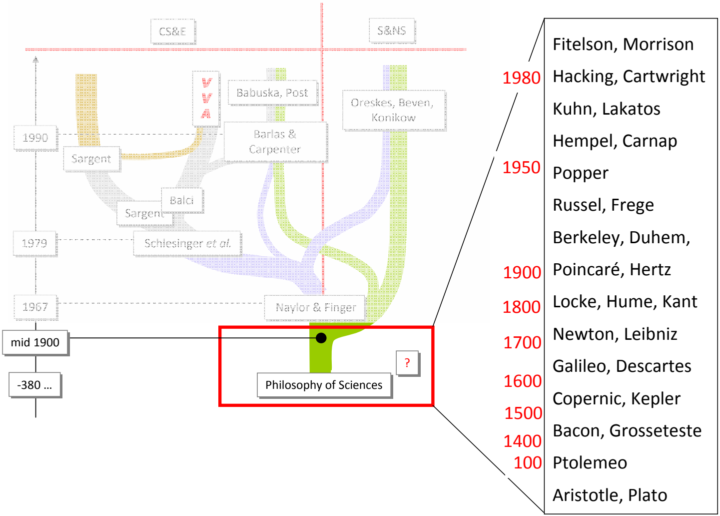
Any rational attempt to approach validation should :

- Explain what is validation : is it THE correct concept ?
- Draw up the inventory of all the entities that exist inside the concerned domain
  - ✓ “define” them in an unambiguous manner
  - ✓ exhibit their salient properties
- Establish the salient relationships between these entities

To realize this job need to find help in many different fields :

- ✓ Philosophy of Science, to highlight grounding ideas
- ✓ Linguistic, to precise the meanings of the words
- ✓ Mathematics to use formal methods for rigourous specifications
- ✓ And also Theoritical Informatics . . .

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

The logic of Induction and Falsification

Induction Principle (Aristotle ~ -350) : (originally "every man is mortal") , I prefer "a is a Raven ($R(a)$) and (&) it's Black ($B(a)$), "b is a Raven and it's Black", ... Induction : "every Raven is Black" :

$$[R(a)\,\&\,B(a)]\ \&\ [R(b)\,\&\,B(b)]\ \&\ \ldots\ \Rightarrow [\forall\ x\ :\ R(x) \vDash B(x)]$$

Falsification Principle (Grosseteste ~ 1200) : Only one non black raven falsifies the general rule

$$R(a)\,\&\,\neg B(a)\ \Rightarrow \neg[\forall\ x\ :\ R(x) \vDash B(x)]$$

Practical consequence

Accumulation of "positive" instances (*ie.* $[R(x)\,\&\,B(x)]$) never implies the truth of the theory

On the other and, only one "negative" instance ($[R(x)\,\&\,\neg B(x)]$) is enough to discredit it

Extension : Many predictions which agree experimental observations don't validate a code !.

"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

Falsification

Falsification Principle (2) (Popper 1937) : Model $\mathbb{M}$ can only be refuted by exp. observations, bever validated. As long as it is not refuted, $\mathbb{M}$ is said CORROBORATED

Corroborated ~ some kind of "provisional validity"

$\mathbb{M}(x)$: "Model $\mathbb{M}$ correctly predicts at point x"

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

→ Formally

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

~~The classical mistake~~

~~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 continued

The core of the Popper's thesis about validation is that strong corroboration follows from prediction of really new facts (*i.e.* facts unknown prior to the predictions)

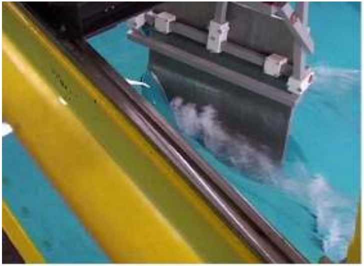
By contrast, trying to predict very well known facts only slightly increases the corroboration of the model.

Simulations that reveal unknown facts are a prominent step through validation

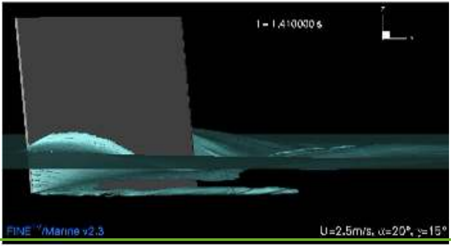
By courtesy of Michel Visonneau : Visonneau, Deng, Queutey & Wackers, *Anisotropic grid adaptation for RANS simulation of ship flow*, FAST 2011, 11th In. Conf. on Fast Sea Transportation Honolulu, USA, Sep. 2011 (to appear)

A breaking bow wave created by an inclined flat plane piercing the free-surface (leeward side).

Experiments : Ecole Centrale de Nantes



Computations done with the ISIS-CFD code (water-air limit indicated by the 0.5 air concentration surface ; note the air tube below the free surface)



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 !

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

Pratical consequence

A Theory (or Model, or Code) experimentally refuted will (allways ?) be able to be preserved by invoquing the falsity of some Auxiliary Hypotheses (*"Protective Belt"* concept by Lakatos)

Pragmatism : drop out the actual code only if we have another better !
(*i.e.* which successfully passes the test which falsified it).

Let's have fun : The Paradox of the Ravens (Hempel) . . .

. . .or *How to be an ornitologist 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 to enhance the confidence in* $\mathbb{T}$

What can we keep from this paradox ?

Just to say that many other funny paradoxes exist.

This paradox signed the death sentence of any attempt to approach validation from a strict logical point of view

At the same time, all attempt to search for a perfectly consistent corroboration (simply saying “validation”) strategy is fruitless

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

Fitelson, *Sudies in bayesian confirmation theory*, Phd, Univ. of Wisconsin Madison (2001)

What do we have to remember from this quick philosophical tour?

- That validation as a *quest for absolute truth* is an utopia

  Only disagreement between predictions and measures can be conclusive (falsification)
  Draw here a parallel with statistical tests of significance

  Use weaker concepts than validation Corroboration (Popper, ...), Confirmation (Carnap, ...)

- That Auxiliary Hypothesis may often be invoked to preserve desperate situations

  Calibration of code $F : (x, u) \to y$ leads to infer on u; let $\mathbb{P}(u|\mathbb{K})$ the posterior pdf according to knowledge $\mathbb{K}$. $\mathbb{K}$ is a collection of Auxiliary Hypothesis $\mathbb{A}_1, \mathbb{A}_2, \ldots$
  If experiment $O(x')$ disagrees with prediction $[F(x', u)\ ;\ u \sim \mathbb{P}(u|\mathbb{K})]$, that may be related to the falsity of some $\mathbb{A}_i$.

- That nothing can be developed without a serious effort of conceptualisation and formalization

  Meanning that we must specify nay thing from the Whole World that matters
  (Reality, Meaure, Code, Model, Simulation, ...)

The philosophical legacy in validation programs

- Broadly speaking there exist two kinds of "validation programs" :
  - ✓ the one which claim this legacy ....generaly encountered in S&NS
  - ✓ and the one which don't .... CS&E in general, and mainly VV&A
- Even if these attitudes can be explained, some natural questions arise
  - Has philosophy of science an interest in CS&E validation programs ?
  - If so, how can we reconcile sharp positions of the former with pragmatism of the latter ?
  - Could we imagine to take the best of CS&E and S&NS approaches in a single program ?

A personal opinion

We can't develop a validation program without knowing a bit of philosophy of science ; nevertheless don't let you trapped by it (impossibility statements are unproductive)

That's all very well, but what are the practical consequences ?

Let me cite Suppe reported by Irobi *et al.*

Irobi, Andersson, Wall *Correctness criteria for models' validation - A philosophical perspective*

- *"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 ?"*
  Hint : compare to Levins' claims (1966) cited above
- *"Don't take underdetermination, and assumption-ladenness of simulation models, too seriously."*

And Irobi *et al.* to conclude that the only question that matters are

- *"What level of certainty do we want for scientific knowledge ?"*
- *"Can simulation models provide that level of certainty ?"*.

George P.E. Box : *"Essentially, all models are wrong but some are usefull"*

To obviate any misunderstanding, I will write for now on the term *validation* between quotes

(unless omissions)

Structuring the mess around "validation" needs attention and rigour

Through its Theory of Explication, the great probabilist Rudolf Carnap laid down the foundations for this job. . .

Carnap, *Logical foundations of probability*, Chicago Univ. Press, chap 3, (1962)

In the few next slides I'm going to present you the two leading ideas developped by Carnap :

1. Clarify the vocabulary : from the notion to the concept
2. Formalize and specify : establish logical connexions among the concepts

Structuring the mess STEP 1 : from Explicandum to Explicatum

"[by] *the procedure of explication we mean the transformation of an inexact, prescientific concept, the explicandum, into a new exact concept, the explicatum. The explicandum may belong to everyday language or to a previous stage in the development of scientific language.*
The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logico mathematical or empirical concepts."

Fundamental talks here writen in purple

A naive example :

Explicandum for "Fish" : *"An underwater living animal"*.

Explicatum : set of all anatomical characteristics required for every zoologist be able to decide if any underwater living animal is a fish (at the explicandum level a whale is a fish !).

Hereafter we will use the analogies : Explicandum = Notion and Explicatum = Concept

A less naïve example than fishes and whales . . .

Let us start with the ASCI definition of validation :

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

1^{st} : Is each word of this sentence defined otherwise ?

2^{nd} : Is the meaning of the verb *to confirm* clearly established ?

3^{rd} : How this sentence must be intended if predictions match perfectly wrong measures ?

4^{th} : What is a *code* : a generic one or its restriction to the specific problem to which we must face up to ?

For exemple ; if I say that

> The code is the computer implementation of algorithms developed to facilitate the formulation and approximate solution of a class of models.
>
> Thacker *et al.*, *Concepts of model verification and validation*, Los Alamos National Laboratories Report, LA-14167-MS, (2004)

Do I really answer the 4^{th} question above ?

Do you think I really give an EXPLICATUM of the (notion of the) code ?

Structuring the mess STEP 2 : the need for 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, ontology is a way to describe formally the knowledge about the domain of interest

Some elements about the Carnapian logico-mathematical system (1/2)

First of all : let us remark that we have to deal with "things" of different types :

- Some are so-called Objets :
  - ✓ Reality, Code, Conceptual Model, Experiment, Measure, . . .
- Others are Functions :
  - ✓ Simulation, Validation, Modeling. . .
- Finally, some describe Properties :
  - ✓ Validity, Fidelity, Confidence. . .
  - ✓ We may think to a Property as the attribute of an object (Object :: Property)

A natural requirement is that any function F from domain $dom(F)$ onto codomain $cod(F)$ transports properties of any object $O \in dom(F)$ to its image Object $O' \in cod(F)$:

$$O' = F(O) \ \& \ O :: P \ \vDash O' :: P$$

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

> Only highly sophisticated mathematical tool may lead us towards a Carnapian logico-mathematical system

Some elements about the Carnapian logico-mathematical system (2/2)

A more common word today is that one of STRUCTURE (as a the mathematical object).

Examples of Structure(s) are :

- ONTOLOGY
  Soft. Eng., Knowledge Eng., Artificial Intelligence (AI), Discrete Systems, Agent-Bsed Systems, . . .
- FORMAL SYSTEM
  Mathematics (Logics), Meta-Mathematics, AI, Cognitive Sciences. . .

Examples of tools allowing for the construction ot these structures are :

- Graphs Theory (i.e. oriented/compositional/semantic), Nets, Networks, . . .
- Categories Theory, Sketches Theory, Diagrams, . . .

Just few words about ontologies (ref. given further)

Reality $\xrightarrow{\text{Observation}}$ Notion $\xrightarrow{\text{Conceptualization}}$ Concept $\xrightarrow{\text{Specification}}$ Ontology

Minimal properties required for any "good" ontology :

- Clarity of the definitions and the words used ; ambiguity must be precluded
- Consistency : we can't deduce from it any inference which disclaims it
  Original idea of Carnap of a Logical System SL ex : if SL $= \{ A \vDash B \ ; \ B \vDash \neg C \}$; then inference $C \vDash A$ disclaims SL because SL implies $C \vDash \neg A$
- Minimality (parcimony) : no plethoric glossary ! ! !

From an operational point of view, an Ontology is simply a diagram, made of symbols and connexions, caracterised by connection points and branches

UML class diagrams illustrate ontologies

What's the need for ontologies ?

- Outline the problem in an easily comprehensible way

  ie graphical ; Remember Pascal "A good drawing is better than a long discourse"

- Describe exhaustively all the knowledges used for solving it

  ie in bayesian calibration of some code against some experiments

- Share the problem solver among other researchers

  How can we do this if the description is equivocal ?

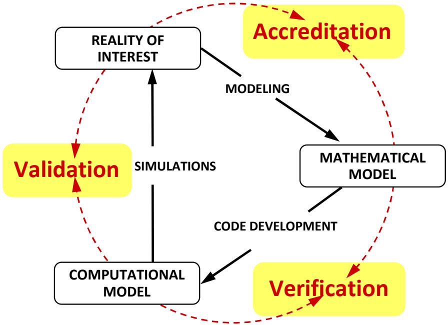
  See below different senses of the word "validation"

- Simplify improvements, modifications, traceability, maintenability. . .

Are there any ontologies in “validation” programs today ? (1/3)

Perhaps the most known diagram in the “validation” community ?

“The Sargent’s circle” (in fact due to Schlesinger *et al.* (1979))

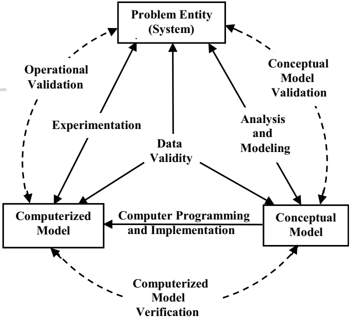


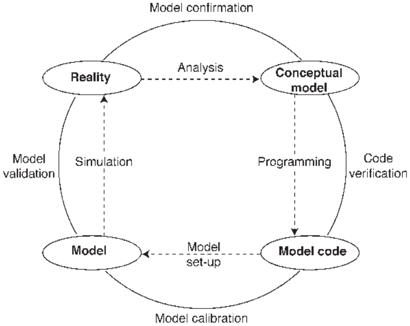
It’s very far from a true ontology (of course there is an associated glossary not presented here)

It’s only a first step . . .but it proved its usefulness !

Are there any ontologies in "validation" programs today ? (2/3)

Two refinements of the Sargent's circle (left in CS&E, right in S&NS)





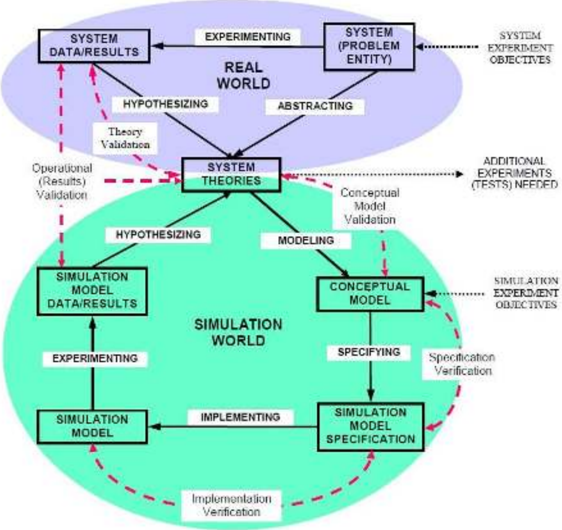
Left Sargent, *Verification and validation of simulations models*, Proceedings of the 2009 Winter Simulation Conference

Right Refsgaard & Henriksen, *Modelling guidelines - A theoretical framework*, in J.C. Refsgaard, Ed., *HarmoniQUA - State-if-the-Art Report on quality assurance in modelling related to river basin management*, (2002)

There is a slight progress, but we are still far from an ontology

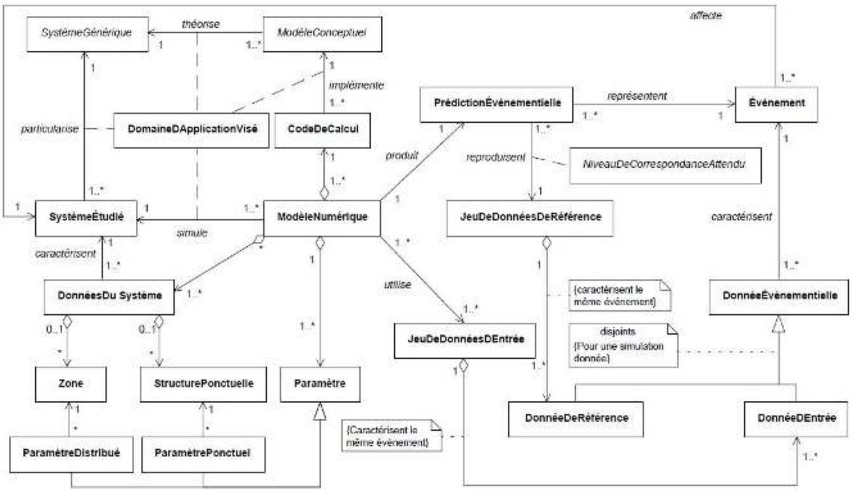
Are there any ontologies in "validation" programs today ? (3/3)

This one marks a real progress, but the representation rules remain still confused (Sargent, *Verification and validation of simulations models*. Proc. of the 2009 Winter Sim. Conf.)



Example of an ontology for code calibration

By courtesy of J-P Vidal, *"Assistance au calage de modèles numériques en hydraulique fluviale - Apports de l'intelligence articificielle"*, PhD Institut National Polytechnique de Toulouse (2005), fig 4.3, p 88



A primer on BCT (1/3)

Hempel, *Studies in the Logic of Confirmation*, Mind, 54, (1945)
Carnap, *Logical foundations of probability*, Chicago Univ. Press, (1962)

Theory of Confirmation lies on degrees of beliefs that :

1. may take continuous values (Gradualism) . . .not merely 0 and 1)
2. are represented by probabilities (Probabilism)
3. may be updated through conditionalization (Bayesianism)
   Hence the word Bayesian in BCT

As claimed above, only the incremental confirmation

(Carnap named *"Confirmation as Increase of Firmness"*)

is able to represent intuitive notions about confirmation.

But the price to pay for that is the lost of unicity in the conclusion we may draw (see delow)

A primer on BCT (2/3)

DEFINITION Fitelson, *Studies bayesian confirmation theory* , PhD, Univ. of Wisconsin, (2001)

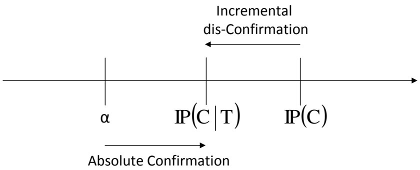
Two measures $\mathfrak{C}_1(\mathbb{C}, \widehat{T}|\mathbb{K})$ and $\mathfrak{C}_1(\mathbb{C}, \widehat{T}|\mathbb{K})$ of the degree to which $\widehat{T}$ confirms $\mathbb{C}$ relative to $\mathbb{K}$ ares said to be *ordinally equivalent* just in case, for all $\mathbb{C}$, $\widehat{T}$, $\mathbb{K}$, $\mathbb{C}'$, $\widehat{T}'$ and $\mathbb{K}'$:

$$\mathfrak{C}_1(\mathbb{C}, \widehat{T}|\mathbb{K}) \geq \mathfrak{C}_1(\mathbb{C}', \widehat{T}'|\mathbb{K}') \text{ iff } \mathfrak{C}_2(\mathbb{C}, \widehat{T}|\mathbb{K}) \geq \mathfrak{C}_2(\mathbb{C}', \widehat{T}'|\mathbb{K}')$$

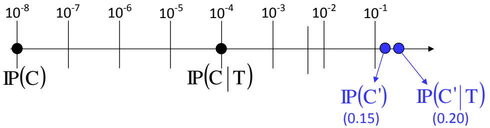
Fitelson, *ibid.* continued

"The surprising thing . . .is that this intuition couldn't be farther from the truth. It is rather surprising . . ., but most proposed relevance measures - although stemming from the very same qualitative notion - give rise to (radically) nonequivalent quantitative gauges of the degree to which $\widehat{T}$ *confirms* $\mathbb{C}$ *relative to* $\mathbb{K}$*."*

A primer on BCT (3/3)



Loss of unicity : all is a matter of confirmation “metric”



Which of C and C' is the most confirmed by T ?		
$\mathbb{P}(C \mid T) - \mathbb{P}(C) \approx 10^{-4}$	$\mathbb{P}(C' \mid T) - \mathbb{P}(C') = 0.05$	$\dfrac{\mathfrak{C}(C,T)}{\mathfrak{C}(C',T)} \approx \dfrac{1}{200}$
$\dfrac{\mathbb{P}(C \mid T)}{\mathbb{P}(C)} \approx 10^{4}$	$\dfrac{\mathbb{P}(C' \mid T)}{\mathbb{P}(C')} \approx 1.33$	$\dfrac{\mathfrak{C}(C,T)}{\mathfrak{C}(C',T)} \approx 8000$

Is BCT useful for "our validation" quest ? (1/2)

A well known BCT "metric" is $\mathfrak{C}(\mathbb{C},\, \widehat{T}) = \mathbb{P}(\mathbb{C} \,|\, \widehat{T}) - \mathbb{P}(\mathbb{C})$. (by def. :confirmation is $\mathfrak{C} > 0$)

By Bayes' theorem we have : $\mathfrak{C}(\mathbb{C},\, \widehat{T}) = \dfrac{\mathbb{P}(\widehat{T} \,|\, \mathbb{C})\, \mathbb{P}(\mathbb{C})}{\mathbb{P}(\widehat{T})} - \mathbb{P}(\mathbb{C}) = \left(\dfrac{\mathbb{P}(\widehat{T} \,|\, \mathbb{C})}{\mathbb{P}(\widehat{T})} - 1 \right) \mathbb{P}(\mathbb{C})$

The term between parenthesis is the Bayes Multiplier $\beta(\mathbb{C},\, \widehat{T})$:

- $\beta(\mathbb{C},\, \widehat{T})$ increases as the agreement between $\widehat{T}$ and (the prediction made by) $\mathbb{C}$ increases (likelihood)
- and/or as the prior probability of observing $\widehat{T}$ ($\mathbb{P}(\widehat{T})$) decreases (aka Surprise Effect)

Example : drawing of a card from a 52 cards deck.

$\mathbb{C}$ predicts "The card is a spade" and result $\widehat{T}$ is "The drawn card is black".

Here $\mathbb{P}(\widehat{T} \,|\, \mathbb{C}) = 1$, $\mathbb{P}(\widehat{T}) = \frac{1}{2}$ and $\mathbb{P}(\mathbb{C}) = \frac{1}{4}$, leading to $\mathfrak{C}(\mathbb{C},\, \widehat{T}) = \frac{1}{4}$.

Because $\mathfrak{C} > 0$, we conclude that $\widehat{T}$ confirms $\mathbb{C}$ because $\mathfrak{C}$ is > 0.

Note that $\mathbb{P}(\widehat{T} \,|\, \mathbb{C}) = 1$ and $\mathbb{P}(\widehat{T}) = \frac{1}{2}$ give the same conclusion !

Elementary my dear Watson !

Is BCT useful in "our validation" quest ? (2/2)

Is it really so elementary ?

> When we are concerned with a TRUE problem (no longer a toy one), how can we assess the values of $\mathbb{P}(\mathbb{C})$ and/or $\mathbb{P}(\widehat{T})$?

A brilliant idea could be to use a multiple-target approach . . .

Here is a very simple illustration

- As stated above, the bayesian calibration of $\mathbb{C}$ can be thought of the construction of an (often infinite) family of instances $\mathbb{C}_{[\mathscr{U}=u]}$ of $\mathbb{C}$ for different values u of $\mathscr{U}$ (each instance having its own faith)
- So, by marginalization we have $\mathbb{P}(\widehat{T}) = \sum_{(u)} \mathbb{P}\big(\widehat{T} \,|\, \mathbb{C}_{[\mathscr{U}=u]}\big)\,\mathbb{P}\big(\mathbb{C}_{[\mathscr{U}=u]}\big)$
- Unfortunately it is obvious to see that the Bayes multiplier is then always < 0, meaning that we can never confirm the code $\mathbb{C}$! ! !

  The reason is that the value needed for $\mathbb{P}(\widehat{T})$ must reflect the degree of belief we have in the event $\widehat{T}$, INDEPENDANTLY of any $\mathbb{C}$-based explanation such that $\mathbb{P}\big(\widehat{T} \,|\, \mathbb{C}_{[\mathscr{U}=u]}\big)$

My conclusion about BCT

- From a theoritical point of view, BCT is a very powerful and charming theory with some important leading ideas :
  - ✓ Pre-eminence of incremental confirmation. over absolute confirmation
  - ✓ Treatment of repeated experiments and of facts known before experiments (aka "old evidence")
  - ✓ Notion of the "surprise effect" of an experiment (see below)

- In the framework of "M&S-validation" BCT is quite inopportune
  Mainly because "absolute probabilities" ($\mathbb{P}(\widehat{T})$ above) may not be assessed

- Nevertheless some of its elements should be kept in mind as a guidance for developping ad'hoc "confirmation" strategies in M&S
  One of the most relevant illustration will be presented below

Advertisment

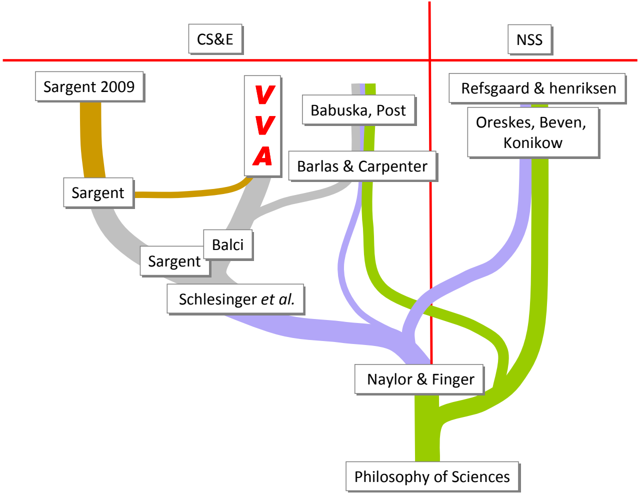
One of the typical french sport, when we're faced do some leader, is to try to dig up its weaknesses instead of blindly admire it !

So, to sacrify to this custom, and even if I recognize, without any restriction, VV&A as the world major program in validation affairs, the following will be closer to the Criticism than to the Promotion

I hope François Hémez will not hold this position against me !
"Nothing personnal Jake, it's just good business", Pirates of the Caribean 3

A list of selected references is given in the last slides to help you to make up you own mind

Remember where we are



Homeopathic doses of Philosophy of Science in VV&A !

VV&A is a PARADIGM for validation in CS&E.

As promoted by Th. Kuhn, a Paradigm describes 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 Kolmogorov (K41) theory

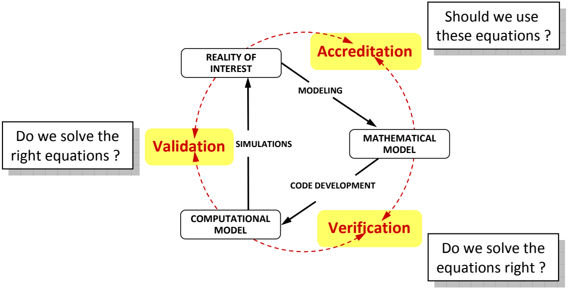
Paradigms imply :

- Normativity and Orthodoxy (some kind of *Tables of the Law*)
  only slight variants may exists (see AIAA and DoE defs.)
  I consider Sargent (a prominent actor of validation in CS&E) as outside VV&A community ; that seemed important to be said because of his very personal views
- 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 the counterpart of withdrawal into oneself and few openings towards other thoughtstyles

From a khunian perspective, VV&A can be considered as a paradigm for "validation"

The three majors axes along which VV&A develops

1. Semantics where notions are translated into concepts
   - ✓ Generally this giges rise to Glossaries (an example is shown below)
2. Syntactics where relations between concepts are established
   - ✓ I talked above of Carnap's logical system and Ontologies
   - ✓ In VV&A syntactics reduces to some "circle" or any sophisticated variant (see above)



3. Pragmatics which (here) concerns the relations between concepts and "validation" actors.
   - ✓ All the guidelines, documentation management, technical reviews, come under pragmatics

The analysis matrix of VV&A

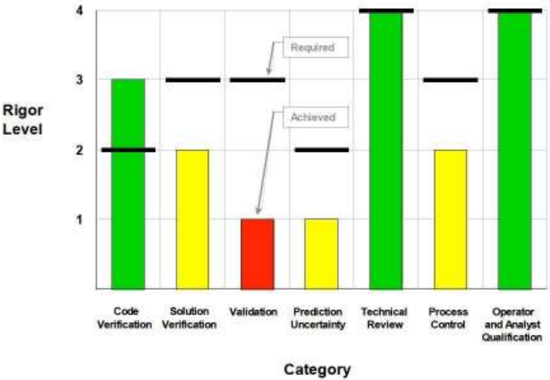
This matrix reflects my feeling about VV&A ...but, of course, VV&A's members would probably emit other opinions !

opinions in cells are established from carnapian ideal of well-constructed systems and ontological needs

	Semantics	Syntactics	Pragmatics
Verification	Sufficient	Sufficient	Sufficient
Validation	Perfectible	Insufficient	Sufficient
Accreditation	Not formalized	Not Fomalized	The poor relation

This opinion agrees roughly with those reported on the next slide

Maturity of the Validation process in CS&E



From Blattnig *et. al.*, *Towards a credibility assessment of models and simulations*, AIAA 2008-2156, April, 2008, [table 3 p10]

Established after the CAIB (*Columbia Accident Investigation Board*) conclusion reports ; see URL http ://caib.nasa.gov ; mainly from analysis of VV&A state-of-the-art

The "validation-row" of the analysis matrix

	Semantics	Syntactics	Pragmatics
Validation	Perfectible	Insufficient	Sufficient

1. Semantics : Take for exemple the very representative glossary given by Thacker *et al.*
   Thacker, Doebling, Hemez, Anderson, Pepin, Rodriguez *Concepts of model verification and validation*, Los Alamos National Laboratories Report, LA-14167-MS (2004)

   It is sizeable (40 items), detailed, but not free from ambiguities (notion stage instead of concept stage)

   I believe that a sound work still remains to finalize the semantic aspect.

2. Syntactics : even the most sophisticated diagrams ("circles") are far from ontologies ; objects, functions and properties are not clearly defined.

   Here again more formalisation seems desirable

3. Pragmatics is on the other hand the strong point of VV&A

   Perhaps it even concentrated too much attention relatively to semantics and syntactics ?

Even if they are not the real subject of this talk, let us give some comments about :

1. Verification

   Very well documented, plethoric list of verification methods (either formal or not).

   As pointed out by Blattnig (above) Code Verification (inherited strategies from software and system engineering) is more advanced than Calculation Verification
   See the course by François Hémez in this Summer School

2. Accreditation

   Sometimes called Certification, often mentionned and rarely discussed
   Accreditation is a very informal process, largely invoking experts elicitation, decision theory, politics, communications management, ...that maybe explains it

How AIAA explains "validation" !

> *"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."*
>
> *Guide for the verification and validation of computational fluid dynamics simulations*, AIAA, Rep. AIAA-G-077-1998

And how ASCI does to !

> *"Validation is the process of confirming that the predictions of a code adequately represent measured physical phenomena."*
>
> *US DoE Advanced Simulation and Computing Program Plan*, Sandia Nat. Lab, Rep. SAND 2004-4607PP

Comparison of the terms used in these two definitions. The most relevant feature ASCI ignores is probably the reference ot *"intended use"*.

AIAA	ASCI
Intended use	______
Model	Code
Representation	Prediction
Reality	Measure
Accurate	Adequate

Watch out !

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."*

None of these definitions entail that the model has a physical (viz theoritical) support.

They might just as well concern Models of data (viz statistical models . . .)

This seems inconsistent with the fact that "validation" must reflect some underlying reality . . .

. . .thus the need to precisely define what is a model !

A focus on Model entries. . .

Thacker et al., see complete ref. above

Conceptual Model	collection of assumptions, algorithms, relationships and data that describe the reality of interest from which the Mathematical Model and validation can be constructed.
Mathematical Model	the mathematical equations, boundary values, initial conditions and modeling data needed to describe the Conceptual Model.
Model	Conceptual/Mathematical/numerical description of a physical scenario, including geometrical, material, initial and boundary data.

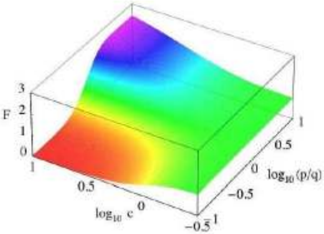
"Validation" according to SKID

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

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

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



- ✓ V_{prior} and V_{post} are values of "validity" respectively before and after observation $\widehat{T}$
- ✓ q is likelihood treshold used for decision making
- ✓ c_{novel} is an expert-like estimation of the *novelty* of the experiment

Even if expression of F is derived from mathematical considerations only, I'm going to show how this choice relates to Bayesian Confirmation Theory

Of BCT and SKID

- As in incremental BCT, SKID focuses mainly on the ratio $V_{\text{post}} / V_{\text{prior}}$

  Accounting for treshold q is an attempt to mix incremental and absolute confirmation

  This is motivated by the decision theoritic approach considered by SKID

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

- SKID try to account for the confirmatory potential of the test $\widehat{T}$ (typically BCT's) through the "novelty coeff." c_{novel} : higher is its values and higher the ratio p/q reveals confirmation or disconfirmation of $\mathbb{C}$ (see fig. above)

  "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

- Finally, SKID analyses repetition of an experiment $\widehat{T}$ in the same way BCT does

  Let us assume a test $\mathbb{T}$ is repeated twice giving meas. $\widehat{T}_1$ and next $\widehat{T}_2$.

  For 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

Other important claims in SKID are

- If SKID recognize that *"tests can only determine for certain that a code is not working properly"*, they also refute the position of Oreskes *el.* that *"verification and validation of numerical models of natural systems is impossible"*, qualifying it as *"newsworthy"*.
- Agreeing with Trucano, Pilch & Oberkampf,
  On the role of code-comparisons in verification and validation, Sandiat Nat. Labs Rep. SAND2003-2752
  SKID say that "*code comparison is not sufficient for validation since validation requires comparisons withs* [material] *experiments . . .*"
- And further *"For validation to remain independent of calibration, it is imperative that these data sets be diskoints"*
  This conclusion is hardly defended by many other authors
  Trucano, Swiler, Igusa, Oberkampf, Pilch : complete ref. in Bib.

> SKID's approach appears as an attempt to bring the gap between S&NS (Sornette, Ide) and CS&E (Davis, Kamm)

The very common mistakes

- The code is validated because measure agrees with prediction

  An experimental result close to a prediction never validates the code neither more nor less than 10, 100 or 1000 conforming results do

- The code is validated because I succeeded in tuning it to represent old experiments

  Calibration is not validation ! (Often claimed but rarely understood)

  "The inferential ideas behind the methodology were given by Kennedy and O'Hagan (2001) but are still not well understood.",
  Bayarri *et al.*, *A framework for validation of computer models*, Technometrics, **49**, num 2, (2007)

- I try to validate the code before trying to verify it (as far as I do this)

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

- When validating a code, only the code matters

  I asserted above the symetry between numerical and material experiments : we can't validate a code by ignoring the world in which it is embedded.

- The best till last : I can validate a code without accounting for uncertainties

  Is it necessary to spend time for demonstrating that it's pure nonsense ?

Did my talk let you confused ?

Maybe you think that Bayesian Calibration is a mathematical activity ?
see Rui Paulo’s course in this Summer School
. . .while Validation is just a cerebral one (in the worst meaning) ?

If you think that you’re wrong :

- About Bayesian Calibration

  Many people believe they calibrate their code : nothing is less true, as for validation they calibrate the Ontology in which their code is plugged ! ! !

  Thus the requirements for rigor and *well-constructedness* are roughly the same, and so are the mathematical tools to asses them !

- About “Validation”

  Do you really believe that Graph Theory, Ontology, Formal logic . . .are only intellectualism and that they should be confined to idle chatter ?

  The fact is, they are at least as mathematical as bayesian calibration is.

So, don’t keep your eyes wide shut

In tribute to S. K. of course

Thoughts

The assessment of the faith we may have in simulation results raises many difficult questions. Among them is those of "validation"

I tried to show you that no universal answer is opportune

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

Thus forget the comfortable idea that there could be a ready-made answer you could deal with. Be suspicious to simple opinions, do not believe those who say that it's easy to validate a code.

Guidelines exist to help you, but YOU have to develop YOUR own validation program for YOUR own needs.

Finaly, remember that Validity is not Truth, *The truth is elsewhere*, X-Files

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)

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, *Leçons d'introduction à 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)

- Küppers, 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	$\dfrac{V_{\text{post}}}{V_{\text{prior}}} = F(\mathbb{P}(\mathbb{C} \mid \widehat{T}),\ q,\ c_{\text{novel}})$	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. $\Rightarrow$ 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 $\mathbb{T}$

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

🟥 : not meet ; 🟩 : meet

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