Approved for Unlimited Release
LA-UR-07-7711

Theory and Examples of a New Approach to Constructive Model Validation

Didier Sornette,[1] Anthony Davis,[2] Kayo Ide,[3] James R. Kamm[2]
dsornette@ethz.ch adavis@lanl.gov kayo@atmos.ucla.edu kammj@lanl.gov

[1]Eidgenössische Technische Hochschule (ETH) Zürich
[2]Los Alamos National Laboratory (LANL)
[3]University of California at Los Angeles (UCLA)

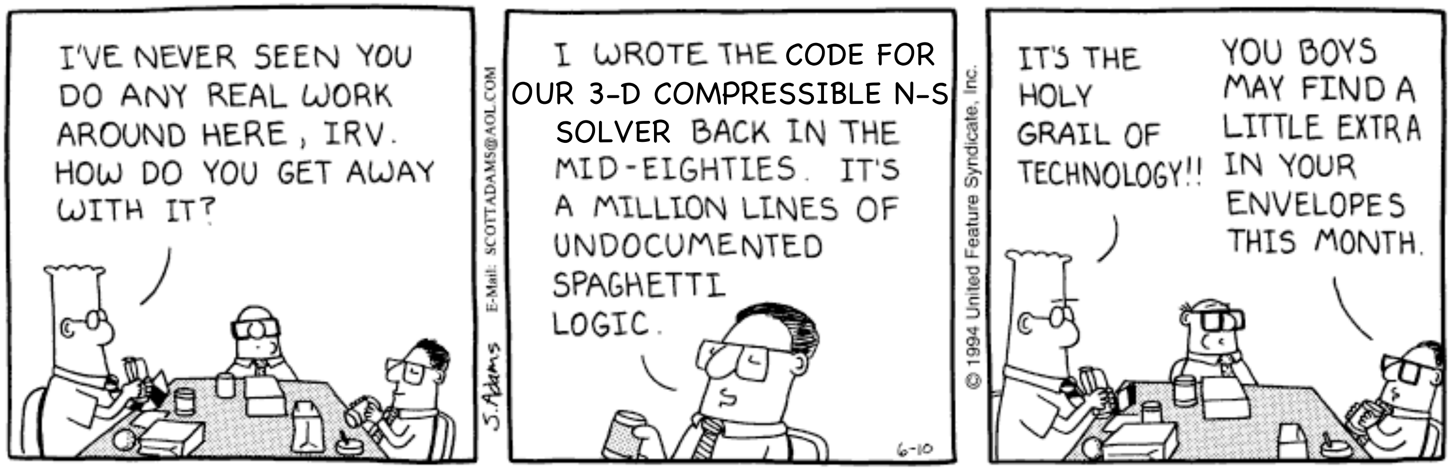
With thanks to:

Jerry Brock, François Hemez, Vladilen Pisarenko, Kathy Prestridge, Bill Rider, Chris Tomkins, Tim Trucano, Kevin Vixie

NATO RTA AVT-147 Symposium on Computational Uncertainty
Athens, Greece
3–6 December 2007

Motivation for V&V

Dilbert



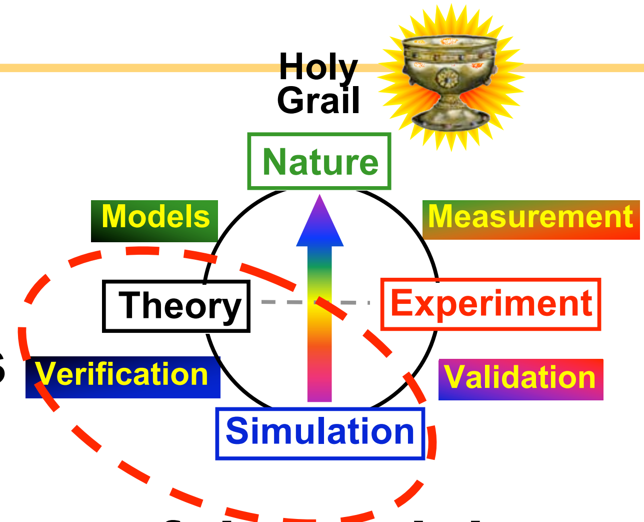
© Scott Adams, Inc./Dist. by UFS, Inc.

Outline

- **Modeling, Verification & Validation**
  - **Verification**
  - **Validation**
  - **"Impossibility" statements**
  - **Complex systems**
- Validation as a constructive, iterative process
- Properties of the proposed validation "multiplier"
- Two examples of the constructive validation process
- Summary
- Additional material

"A computer lets you make more mistakes faster than any invention in human history — with the possible exceptions of handguns and tequila." Mitch Ratliffe, Technology Review, April, 1992

A stable definition of "Verification" has evolved in CSE.

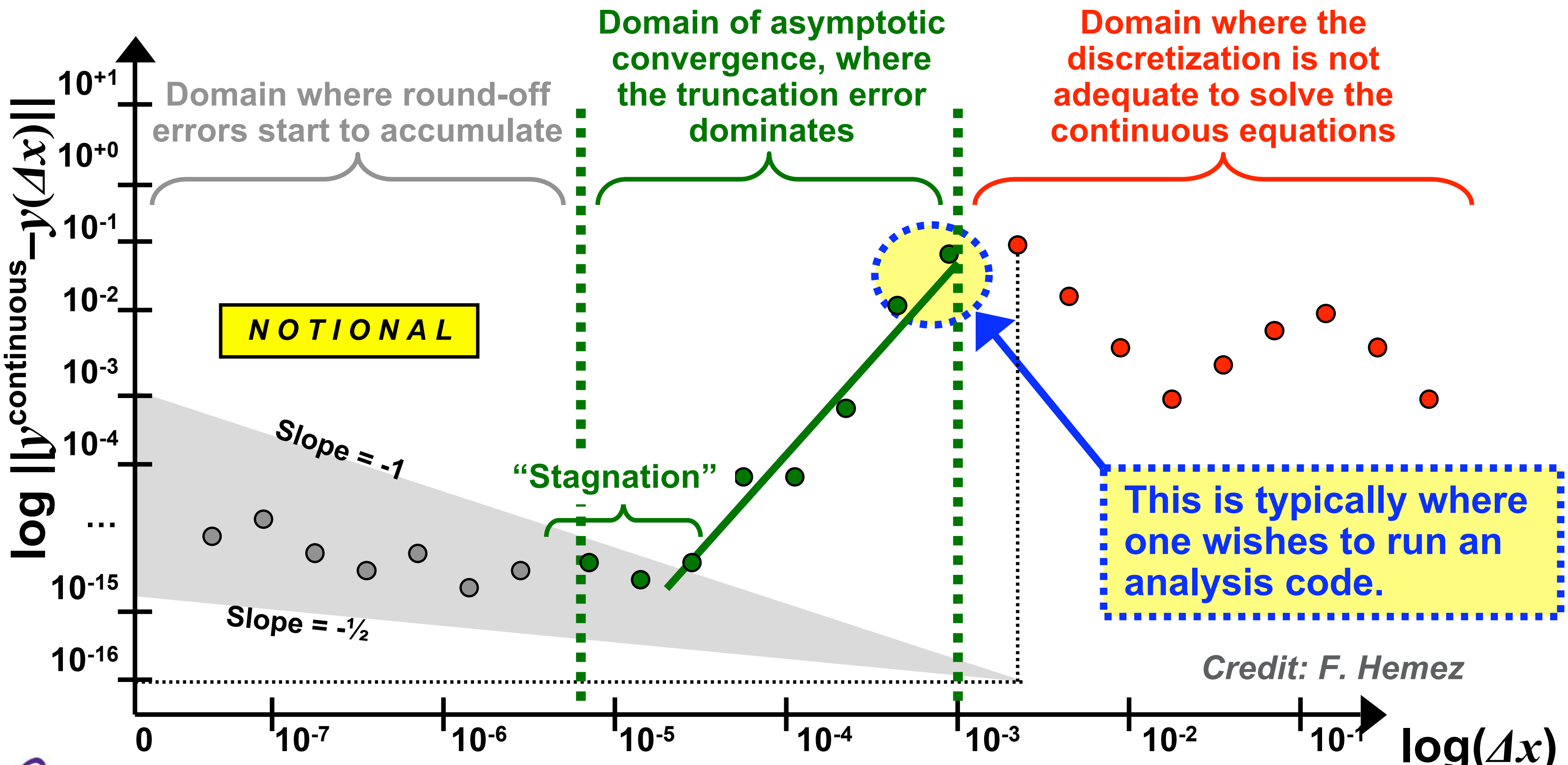


- ASC: Verification is the process of confirming that a computer code correctly implements the algorithms that were intended.
- AIAA/ASME: Verification is the process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model.
- P. Roache: Verification is demonstrating that one solves the equations correctly.

Verification is about mathematics

Asymptotic convergence is the fundamental concept behind verification analysis.

- PDEs are discretized in space Δx, time Δt, etc., for resolution using finite-digit arithmetic.



"As we refine the grid we hope to get better approximations to the true solution."
Randy LeVeque, Computational Methods for Astrophysical Fluid Flow

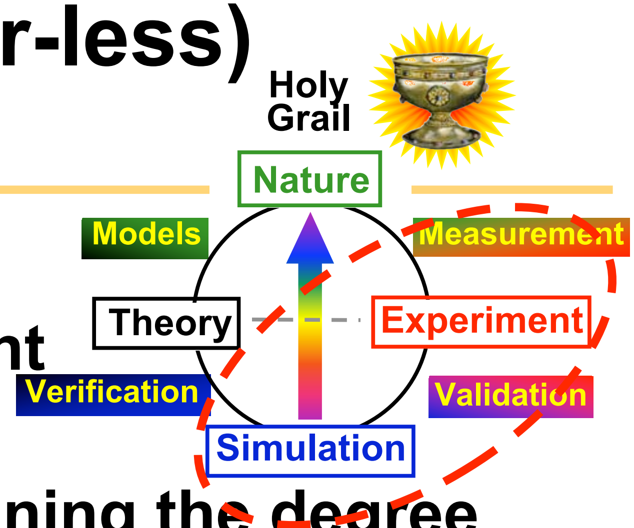
Is This "Verification"?

- **Simple refinement studies are the one approach to conducting meaningful code physics verification — but this must be done *mindfully*.**

> ***"We attain convergence to 1% with respect to increasing spatial and temporal resolution. This ensures that the results shown are converged to the eye, apart from the stress signal in Fig. 4: this shows slight quantitative, but not qualitative, changes."* (*Physics Review Letters*, 1996.)**

"Faced with the choice between changing one's mind and proving that there is no need to do so, almost everybody gets busy on the proof." John Kenneth Galbraith

"Validation" has a (more-or-less) consensus definition.



- <u>ASC</u>: The process of confirming that code predictions adequately represent measured physical phenomena.
- <u>AIAA/ASME</u>: The process of determining the degree to which a model is an accurate representation of the real world from the perspective of its intended uses.
- <u>Schlesinger</u> (1979): The substantiation that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended applications of the model.
- <u>P. Roache</u> (1998): Validation is demonstrating that one solves the correct equations.

Validation is about physics

Is This "Validation"?

- The *view-graph norm* remains entrenched as "the method of choice" by which many in the scientific community continue to approach model validation.

  "If the test data are shown in blue and the simulation data are shown in yellow, then all I want to see is green." **(heard at Los Alamos, October 2005)**

  "…and this movie shows that the simulation is validated … well, I mean, it shows that the model runs." **(heard at a presentation given at Los Alamos, August 2006)**

"People see what they want to see." Mahaffy's Fourth Law of Human Nature

"Impossibility Statements" claim that verification and validation are unattainable.

- **Oreskes et al.: "*Verification and validation of numerical models of natural systems is impossible. This is because natural systems are never closed and because model results are always non-unique*."**
- **Sterman: "*Any theory is underdetermined and thus unverifiable, whether it is embodied in a large-scale computer model or consists of the simplest equations.*"**
- **Similarly, most complex systems can be proved to be computationally irreducible: the only way to predict their evolution is to actually let them evolve in time.**
- **These claims beg the question, "Is V&V hopeless?"**

"Insofern sich die Sätze der Mathematik auf die Wirklichkeit beziehen, sind sie nicht sicher, und insofern sie sicher sind, beziehen sie sich nicht auf die Wirklichkeit." Albert Einstein, Geometrie und Erfahrung (1921)

The search for coarse-grained properties renders "impossibility" claims irrelevant.

- The answer to the preceding question is ***NO*** — the "impossibility statements" of the previous slide have **little practical value**.
- Why? Because in practical physics and engineering, one seeks to predict **coarse-grained properties**.
  - E.g., only by ignoring most molecular detail were laws of thermodynamics, fluid dynamics, chemistry, etc., developed.
- Physics "works" and is not hampered by computational irreducibility because we desire only **approximate answers** at some **coarse-grained level**.
  - The description of coarse-grained scales of practical interest requires "effective" laws generally based on finer scales.

"Predictive capability is about getting the right answer to the right question for the right reason."
S. Doebling, LANL

Outline

- Modeling, Verification & Validation
- **Validation as a constructive, iterative process**
  - **A validation "loop"**
  - **A validation "multiplier"**
- Properties of the proposed validation "multiplier"
- Two examples of the constructive validation process
- Summary
- Additional material

"Verification and validation is what distinguishes a physics code from a computer game."
F. Graziani, LLNL

Studies from a range of disciplines suggest that principled validation is necessary.

"The two most common biases are over-optimism and overconfidence. Overconfidence refers to a situation whereby people are surprised more often than they expect to be. Effectively, people are generally much too sure about their ability to predict. This tendency is particularly pronounced amongst experts. That is to say, experts are more overconfident than lay people. This is consistent with the illusion of knowledge driving overconfidence."

J. Montier, in *The Folly of Forecasting: Ignore All Economists, Strategists & Analysts*

"Nobody's perfect, and most people drastically underestimate their distance from that state."
Mahaffy's First Law of Human Nature

We propose a validation "loop" with four distinct steps.

1. **Start with a prior trust of the model's value, measured by the quantity V_{prior}.**
   - V_{prior} **is a gauge of accumulated trust or confidence.**
   - **On the first iteration of this loop, arbitrarily set $V_{\text{prior}} = 1$.**
   - **The *change* in V_{prior} is important, not its absolute value.**

2. **Conduct an experiment or observation, perform the corresponding simulation, and compare results.**
   - **Each of these three tasks presents its own challenges.**
     - **Which experiments?**
     - **How to calibrate a simulation?**
     - **How to compare experimental data and model results?**

"In science, if you know what you are doing, you should not be doing it."
Richard Hamming

A complete iteration in this validation process has well-defined characteristics.

3. **Assign a metric-based "grade" of the quality of the comparison between observations y_{obs} and model M.**
   - **This is ideally formulated as a statistical test of significance in which the hypothesis (i.e., the model results) is tested against the alternative, which is "all the rest."**
   - **This grade $p(M \mid y_{\text{obs}})$ quantifies the quality of the comparison compared against the reference likelihood q of "all the rest."**

4. **Update to obtain the posterior trust as:**

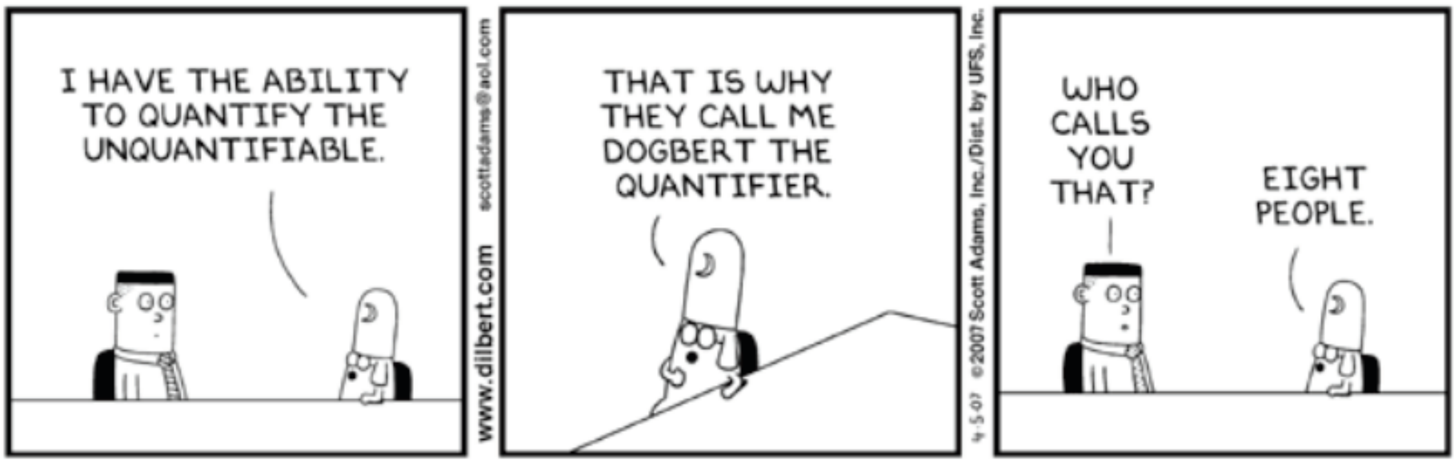
$$V_{\text{posterior}} \big/ V_{\text{prior}} = F[p(M \mid y_{\text{obs}}),\ q\ ;\ c_{\text{novel}}]$$

   - $V_{\text{posterior}} > V_{\text{prior}} \Rightarrow$ **trust/confidence has *increased*.**
   - $V_{\text{posterior}} < V_{\text{prior}} \Rightarrow$ **trust/confidence has *decreased.***
   - c_{novel} **measures the novelty or impact of the experiment.**

"Mathematics is an interesting intellectual sport but it should not be allowed to stand in the way of obtaining sensible information about physical processes." Richard Hamming

A caveat for quantitative validation…

Dilbert



The multiplier F is an attempt to quantify the value of new validation experiments and their corresponding simulations.

$\boxed{V \geq 0}$

Posterior potential utility of model/code

Probability of the prediction of model M passing the statistical acceptance test on data y_{obs}

$\boxed{0 \leq p \leq 1}$

Statistical confidence level

$\boxed{0 < q < 1}$

$$V_{\text{posterior}} = V_{\text{prior}} \times F[p(M \mid y_{\text{obs}}),\ q\ ;\ c_{\text{novel}}]$$

Prior potential utility of model/code

Novelty & relevance of additional experiment

$\boxed{0 < c_{\text{novel}} < \infty}$

- **The identical approach could be applied to *code verification* using exact solutions in place of experiments/observations.**

"A habit of basing convictions upon evidence, and of giving to them only that degree or certainty which the evidence warrants, would, if it became general, cure most of the ills from which the world suffers." Bertrand Russell, in G. Simmons, Calculus Gems.

A complete loop is repeated for several experiments/simulation comparisons.

- **These iterations compound for each experiment:**

$$V_{\text{prior}}^{(1)} \rightarrow V_{\text{posterior}}^{(1)} = V_{\text{prior}}^{(2)} \rightarrow V_{\text{posterior}}^{(2)} = V_{\text{prior}}^{(3)} \rightarrow \ldots \rightarrow V_{\text{posterior}}^{(n)}$$

- **Validation is said to be asymptotically satisfied when the number of steps n and final value $V_{\text{posterior}}^{(n)}$ are sufficiently high.**
- **One can develop increasing trust in a model by subjecting it to more tests that "do not reject it."**
- **Importantly, a single test is enough to reject a model.**
  - **The loss of "trust" can occur suddenly with one single failure and is difficult—if not impossible—to re-establish.**
  - **This encapsulates the common experience that reputation gain is a slow process requiring constancy and tenacity.**

"The plural of 'anecdote' is not 'evidence'." Alan Leshner, publisher of Science.

The PIRT process can be used to help select experiments of particular interest.

- **A codified approach to understanding the sources of uncertainty and lack-of-knowledge is to generate a *Phenomenon Identification and Ranking Table* (PIRT).**

Phenomenon	*Uncertainty*	*Sensitivity*	*Product*
Equation-of-state	Low	Medium	Medium
Artificial viscosity	Medium	Medium	High
Material opacity	Medium	Low	Medium
Energy source	Low	Low	Low

- **The logic of the PIRT is to identify phenomena that are not well-known (i.e.,high uncertainty) and that have significant influence (i.e., high sensitivity).**

"Forcing experts to give odds can be one of the best methods for exposing fundamental weaknesses in an argument." Gina Kolata, Flu

Outline

- Modeling, Verification & Validation
- Validation as a constructive, iterative process
- **Properties of the proposed validation "multiplier" F**
  - **Three important properties of F**
  - **Relation of parameters to aleatoric and epistemic uncertainties**
  - **Two examples of validation multipliers**
- Two examples of the constructive validation process
- Summary
- Additional material

1. If the (statistical) comparison test is passed, then the potential increases.

- **This expresses the notion that the potential trust is an increasing function of the measure p.**
  - **In particular, the better the comparison test is passed, the more the potential trust increases.**
- **This property can be expressed mathematically as:**

$$\boxed{F > 1 \;\; (\text{resp.} < 1) \quad \textbf{for} \quad p > q \; (\text{resp. } p < q)}$$

- **This constraint can be expressed succinctly as:**

$$\log F / \log(p/q) > 0$$

- ***Note*: In *scientific <u>exploration</u>*, a prediction may be wrong but still be useful.**

2. The larger the significance of the passed test, the larger the posterior potential.

- **If the match between a given experiment and model "A" is <u>better than</u> the match with model "B" — all else being equal — then the potential trust in model "A" is greater than the potential trust in "B".**
- **We express this property mathematically as:**

$$\boxed{(\partial F/\partial p)_q > 0}$$

- **There could be saturation of F for large p/q, so that:**
  - $F < \infty$ **as** $p/q \to \infty$ ***—or—***
  - **There is a concavity requirement for large p/q:** $\partial^2 F/\partial p^2 < 0$
  - **Either of these constraints imply that a quality-of-fit beyond a certain level is not useful.**

3. The more "novel" the experiment, the larger the level of the passed test.

- **If the match between a model and experiment "α" equals the match between the model and experiment "β", but with "α" deemed more novel than "β", then the gain in potential trust is greater for "α".**
- **We express this mathematically as:**

$$\textbf{If} \quad p > q \;\; (p \le q) \quad \textbf{then} \quad \partial F / \partial c_{\text{novel}} > 0 \;\; (\le 0)$$

- **The parameter c_{novel} is a judgment-based weighting.**
  - **Its value is assigned by subject matter experts**
  - **Differences among experts will have to be acknowledged and reconciled.**

"Apart from the question of whether the simulation is telling us about the true solution or not, we must consider how much of its behavior we are prepared to see. What we see in a simulation may be biased strongly by what we expect to see." Thomas P. Weissert, The Genesis of Simulation in Dynamics.

Aleatoric and epistemic uncertainties enter through the parameters of F.

- **Epistemic ≈ reducible**
  - E.g., degree to which a model is faithful to the physics
  - ➢ The nature of the model M
    - As knowledge grows, the model improves.
  - ➢ The value of c_{novel}
    - One should target "sensitive" parts of the system with "high novelty" experiments.

- **Aleatoric ≈ irreducible**
  - E.g., experimental variability
  - ➢ How we choose to evaluate the quality-of-fit measure p that estimates the matching between M and y_{obs}
    - Different ps imply different results: seek "optimal" p
  - ➢ q is the reference probability level that any other model can explain the data

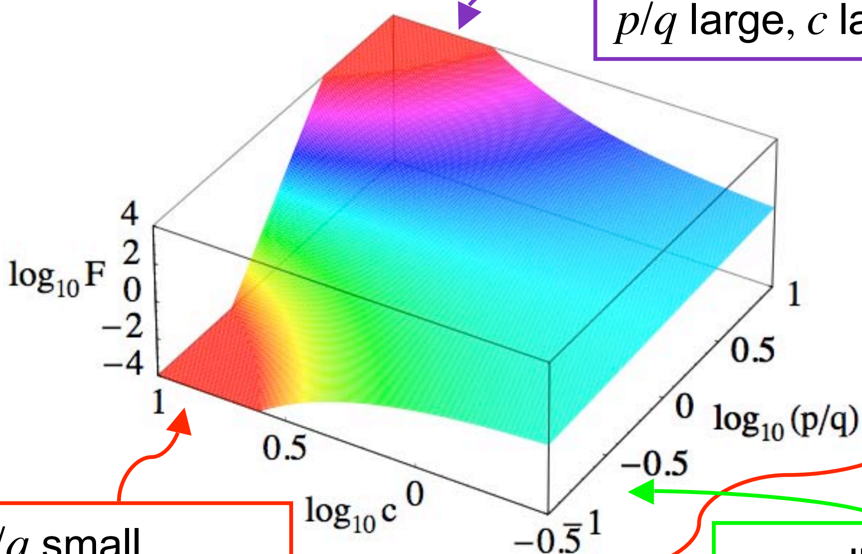
- **A rigorous use of experimental and computational uncertainties must be incorporated into this process.**
  - ➢ Some relevant ideas of F. Hemez are in the supplemental slides.

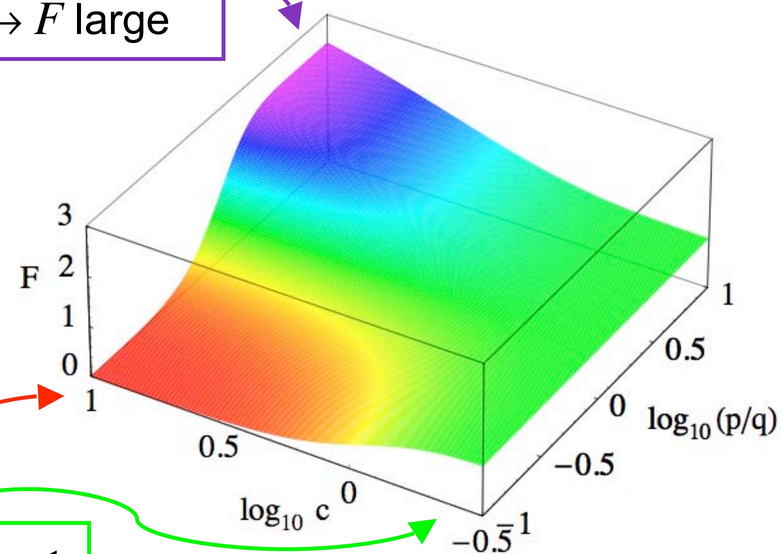
‹‹Le doute est un état mental désagréable, mais la certitude est ridicule.››
Voltaire, letter to Fréderic le Grand (1767)

Two simple functional forms exhibit the desired characteristics.

- **Key parameters:** p **= degree of match** c **= novelty of experiment**
  q **= statistical confidence level**

p/q large, c large $\rightarrow F$ large





p/q small,
c large $\rightarrow F \approx 0$

c small $\rightarrow F \approx 1$

$$F = (p/q)^{c_{\text{novel}}}$$

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

> **Many such forms are possible…**

Outline

- Modeling, Verification & Validation
- Validation as a constructive, iterative process
- Properties of the proposed validation “multiplier”
- **Two examples of the constructive validation process***
  - **Restriction of possible parameter values**
  - **Olami-Feder-Christensen model of seismicity**
  - **Compressible CFD code for Richtmyer-Meshkov instability**
- Summary
- Additional material

* For more examples, see http://arxiv.org/abs/physics/0511219

We make several simplifying assumptions in the following ad hoc examples.

- Use the tanh-based expression for F
- Restrict the possible c_{novel} values:

  $c_{\text{novel}} = 1$ ⇨ marginally useful new test 😐
  $c_{\text{novel}} = 10$ ⇨ substantially new test
  $c_{\text{novel}} = 100$ ⇨ important new test ☺

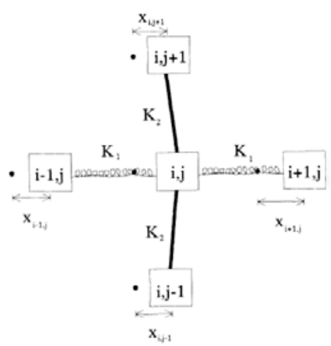
- Consider only the likelihood ratio p/q
- Restrict the possible p/q grades:

  $p/q = 0.1$ ⇨ poor fit ☹
  $p/q = 1$ ⇨ marginally good fit
  $p/q = 10$ ⇨ good fit ☺

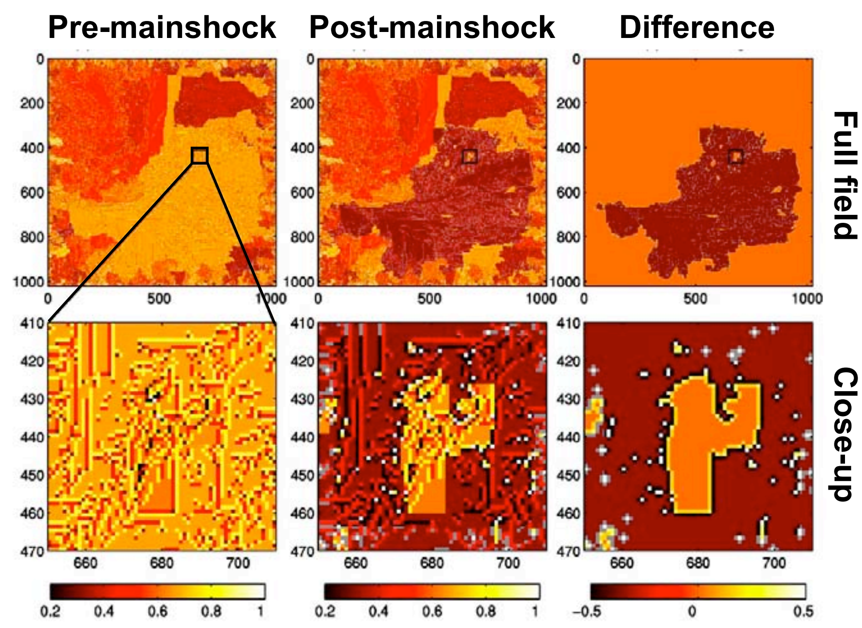
"Whatever you do will be insignificant, but it is very important that you do it."
Mahatma Gandhi

The Olami-Feder-Christensen (OFC) model exhibits real seismicity phemonenology.

- **The OFC model uses local interactions of discrete elements to capture aspects of real seismic behavior.**
  - **Based on a 2-D lattice of springs and blocks with simple, local interactions and threshold behavior in the inter-block forces.**



  - **Exhibits self-organized criticality (SOC): the convergence of dynamics to statistically stationary states with time-independent power law distributions.**
  - **1000s of simulations conducted with different ICs and cutoffs for foreshocks and aftershocks.**



* Z. Olami, H. Feder, K. Christensen, "Self-Organized Criticality in a Continuous, Nonconservative Cellular Automaton Modeling Earthquakes," *Phys. Rev. Lett.* 68, 1244–1247 (1992).

Validation assessment of the OFC model suggests a valuable—but flawed—approach.

i	Test	c_{novel}	p/q	$F^{(i)}$	F^{total}
1	Gives power-law distribution	10	∞	2.4	2.4
2a	Has foreshocks and aftershocks	100	10	2.9	7.0
2b	Omori law exponents	1	0.1	0.47	3.3
3	Scaling of number of aftershocks with main shock size	10	10	2.4	7.9
4	Scaling of number of foreshocks with main shock size	1	1	1	7.9
5	Nucleation of aftershocks at asperities on rupture plane	10	10	2.4	18.8
6	Clustering of earthquakes at faults	100	0.1	4×10^{-4}	7.5×10^{-3} ☹

This model faithfully captures many aspects of seismicity but is not a universally applicable approach.

A CFD code was used to investigate compressible hydrodynamic mixing.

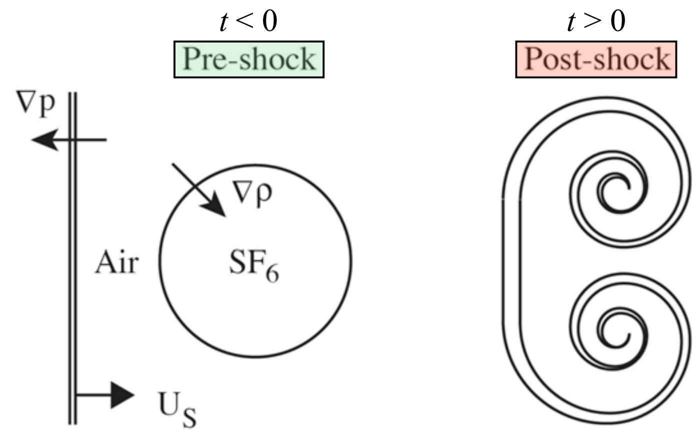
- Computational fluid dynamics (CFD) provides the sole approach to evaluating the complex phenomenology associated with shock-induced mixing.
- Eulerian-frame equations for compressible, inviscid, non-heat-conducting flow of ideal gas in 2-D.
  - Conservation of mass, momentum, and energy, plus EOS:

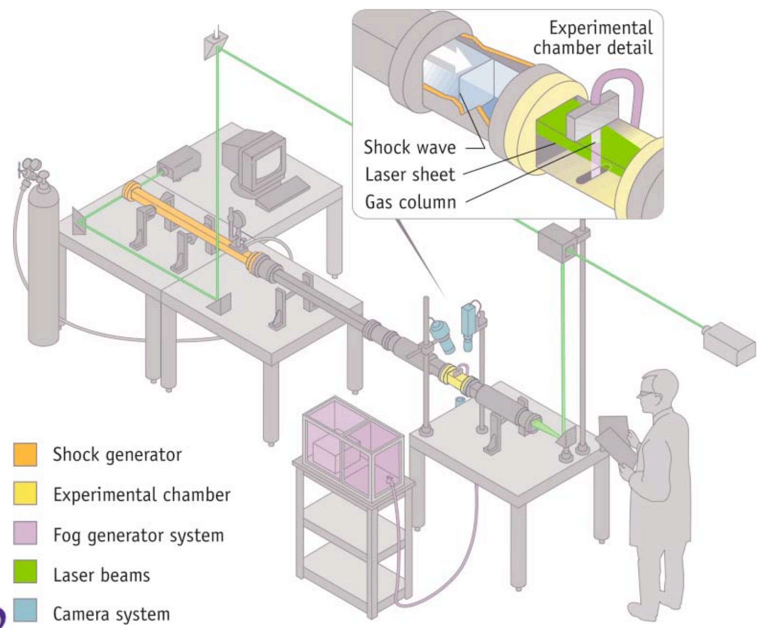
$$\frac{\partial U}{\partial t} + \nabla \cdot F(U) = 0 \ , \quad U \equiv [\rho, \rho u, \rho v, \rho E]^{\mathrm{T}}, \quad p \equiv P(\rho, e)$$

  - Uniform, finite volume discretization using high-resolution Godunov method.
  - Verification results computed for numerous idealized flows
    - E.g., Sod, Sedov, Noh, Cook-Cabot, Woodward-Colella, etc.
    - ~2nd order for smooth problems, ~1st order for problems with shocks.

Shock tube experiments capture idealized Richtmyer-Meshkov instability (RMI) growth.

- **RMI induced by interaction of a weak shockwave with a diffuse cylinder of SF_6 in air.**
  - ➢ **Vorticity deposition occurs due to the mismatch of density and pressure gradients.**





- **Experiments conducted at LANL Physics Div. labs.**
  - ➢ **Quantitative Planar Laser-Induced Flourescence (PLIF) gives quantitative concentration fields.**
  - ➢ **Particle Image Velocimetry (PIV) gives quantitative velocity vector fields.**

Credit: K. Prestridge & C. Tomkins

Validation assessment of the CFD model shows a useful but underwhelming approach.

i	*Test*	c_{novel}	p/q	$F^{(i)}$	F^{total}	
1	**Primary instability (break-up into two) and secondary instabilities**	10	10	2.4	2.4	
2	**Computational *prediction* of a material "bridge" between primary structures**	10	10	2.4	5.8	
3	**Exponential growth of power as a function of time**	10	10	2.4	13.8	
4	**Concentration power spectrum as a function of wavenumber**	1	0.1	0.47	6.5	😐

This model faithfully captures most of these aspects of RMI but must be subjected to more intense testing to be considered a reliable simulation tool.

Outline

- Modeling, Verification & Validation
- Validation as a constructive, iterative process
- Properties of the proposed validation “multiplier”
- Two examples of the constructive validation process
- **Summary**
- Additional material

Summary

- **A four-step approach for a quantitative validation step:**
  1. **Start with a prior "potential trust" of a model's value:** V_{prior} **.**
  2. **Conduct an experiment, use the model, compare results.**
  3. **Grade the comparison between data** y_{obs} **and model** M**.**
  4. **Update posterior "trust":** $V_{\text{posterior}} / V_{\text{prior}} = F[p(M \mid y_{\text{obs}}),\ q\ ;\ c_{\text{novel}}]$
     - **The multiplier** F **must satisfy certain (plausible) constraints.**
- **Iterate the validation process:**

$$V_{\text{prior}}^{(1)} \rightarrow V_{\text{posterior}}^{(1)} = V_{\text{prior}}^{(2)} \rightarrow V_{\text{posterior}}^{(2)} = V_{\text{prior}}^{(3)} \rightarrow \ \dots \ \rightarrow V_{\text{posterior}}^{(n)}$$

- **Two simplified examples—using discrete values of** p/q **and** c_{novel}**— illustrated the nature of this process.**
  - ➢ **These ad hoc examples demonstrate the utility of this approach.**
- **There remain aspects to be refined and worked out…**
  - ➢ **Better evaluation of** q **and algorithmic incorporation of uncertainty quantification must be addressed.**

"What can be asserted without evidence, can also be dismissed without evidence."
Christopher Hitchens, in Slate magazine

Outline

- Modeling, Verification & Validation
- Validation as a constructive, iterative process
- Properties of the proposed validation "multiplier"
- Two examples of the constructive validation process
- Summary
- **Additional Material**
  - **The seven "deadly sins" and seven "virtuous practices" of V&V**
  - **Codifying fidelity, robustness, and confidence of simulations**

"Seven Deadly Sins of V&V"

- Assume the code is correct.
- Only do a qualitative comparison (e.g., the viewgraph norm).
- Use problem-specific special methods or settings.
- Use only code-to-code comparisons.
- Use only one mesh.
- Only show the results that make the code look good, viz., the ones that appear correct.
- Don't differentiate between accuracy and robustness.

- Lust
- Gluttony
- Envy
- Wrath
- Sloth
- Pride
- Avarice

Traditional "7 Deadly Sins"

Hieronymus Bosch. 1485

Otto Dix, 1933

“Seven Virtuous V&V Practices”

- Assume the code has flaws, bugs, and errors then find them—and fix them!
- Be quantitative.
- Verify and Validate the same thing.
- Use analytic solutions & experimental data.
- Use systematic mesh refinement.
- Show all results—reveal the shortcomings.
- Assess accuracy and robustness separately.

- Prudence
- Temperance
- Faith
- Hope
- Fortitude
- Justice
- Charity

Traditional
“7 Cardinal Virtues”

Key objectives of simulations: *Fidelity*, *Robustness*, and *Confidence*.

- **Predictions must agree with the measurements available from experiments or observations.**

  → *High fidelity-to-data.*

- **Decisions based on predictive modeling must be robust to assumptions made, to lack-of-knowledge, and to other sources of modeling uncertainty.**

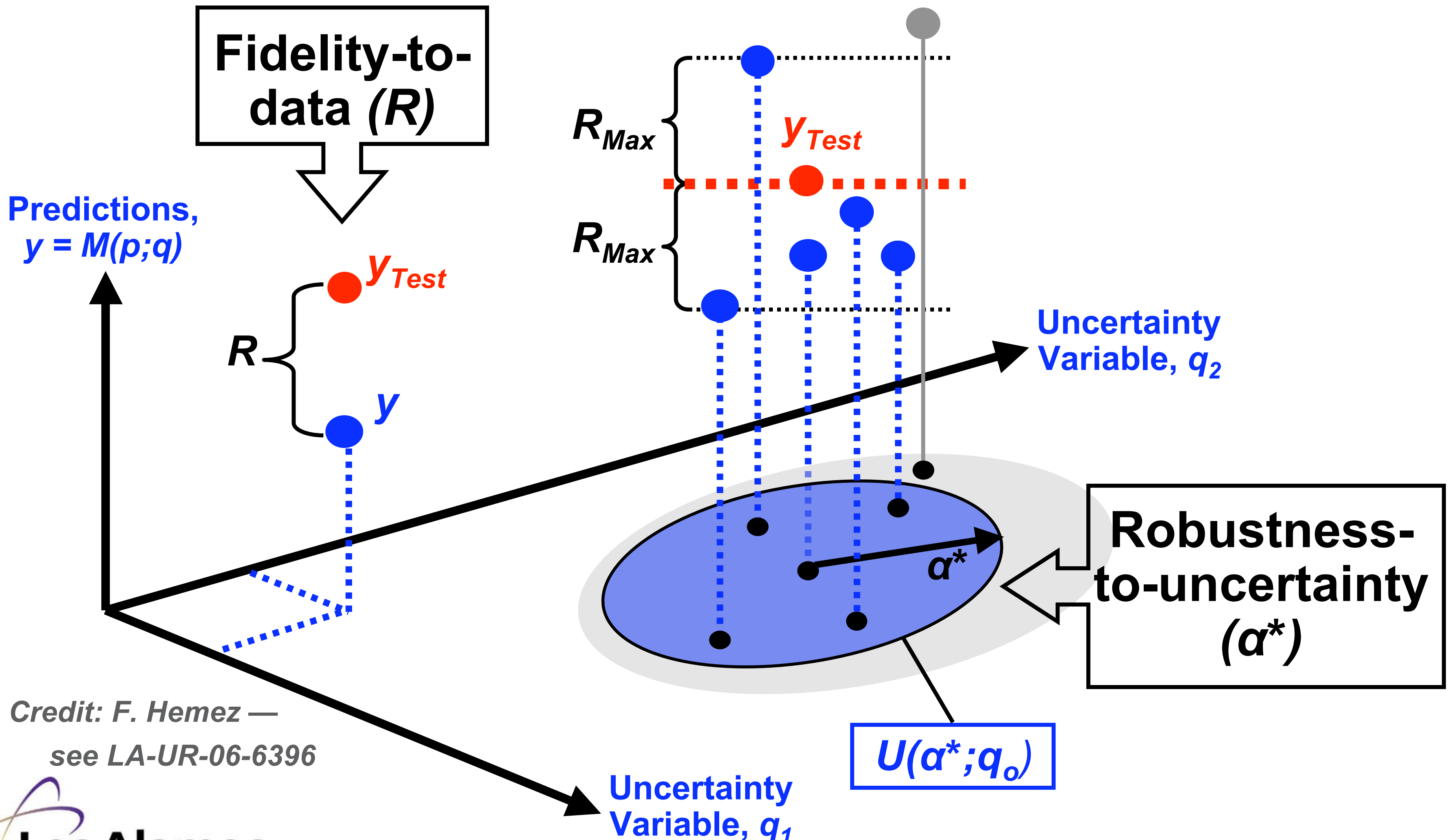
  → *High robustness-to-uncertainty.*

- **Predictions obtained from multiple models must provide a consistent body of evidence, from which *"confidence"* is derived.**

  → *High confidence-in-prediction.*

Credit: F. Hemez — see LA-UR-06-6396

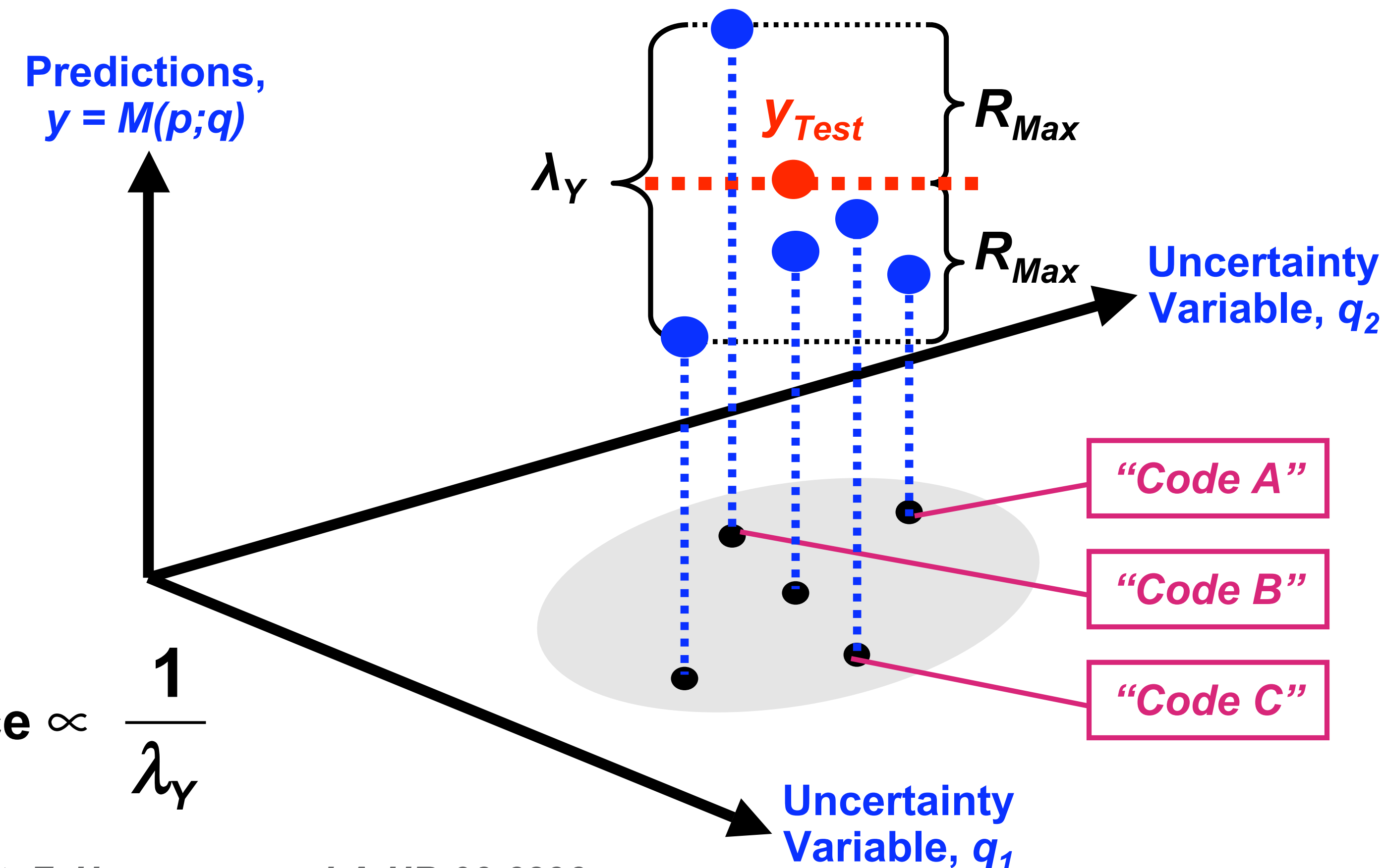
Clear, heuristic notions underlie both Fidelity (*R*) and Robustness (*α**).



Credit: F. Hemez — see LA-UR-06-6396

Confidence is (less intuitively) measured by "looseness" (λ_Y).

- We know of no formal definition of *confidence*, but "looseness" is one way to codify it.(*)



$$\text{Confidence} \propto \frac{1}{\lambda_Y}$$

Credit: F. Hemez — see LA-UR-06-6396

Rigorous definitions for these three quantities can be devised .

- ***Fidelity-to-data (R)***: **Degree of correlation between test data (y_{Test}) and simulation predictions (M).**

$$R^2 = \sum_{k=1}^{N_{Test}} \left(y_{Test}^{(k)} - M(p^{(k)};q) \right)^2$$

- ***Prediction looseness (λ_Y)***: **Range of predictions expected from a *family of equally-robust models*.**

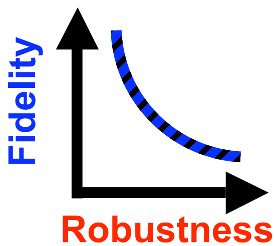
$$\lambda_Y = \max_{M \in U(\alpha^*;q_o)} M(p;q) - \min_{M \in U(\alpha^*;q_o)} M(p;q)$$

- ***Robustness-to-uncertainty (α*)***: **Maximum value of the horizon-of-uncertainty for which all models of the corresponding family *U(α;q_o)* meet a given fidelity requirement *R_{Max}*.**

$$\alpha^* = \max_{\alpha \geq 0} \left\{ R \leq R_{Max} \, , \, \forall \; M \in U(\alpha;q_o) \right\}$$

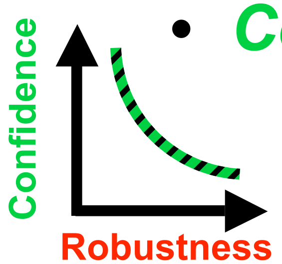
Credit: F. Hemez — see LA-UR-06-6396

These properties are *antagonistic*: one cannot avoid trade-offs between them.



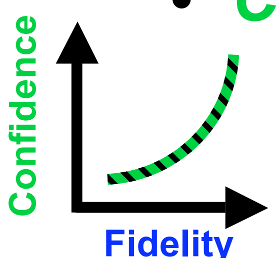
- ***Robustness decreases as fidelity improves.***

Models *calibrated to better reproduce* the available test data become *more vulnerable* to: (i) errors in modeling assumptions, (ii) errors in the functional form of the model, and (iii) uncertainty and variability in the model parameters.



- ***Confidence decreases as robustness improves.***

Models made *more immune to uncertainty and modeling errors* provide a wider range of predictions, and, thus are *less consistent* in their predictions (less predictive power).



- ***Confidence increases as fidelity improves.***

Models *calibrated to better reproduce* the available test data *provide more consistent forecasts*, leading to a *false sense of confidence* ("over-calibration" or "over-fitting").

Credit: F. Hemez — see LA-UR-06-6396

Abstract

Validation can be defined as the process of determining the degree to which a model provides an accurate representation of the real world from the perspective of its intended uses. Validation is crucial as the justification for decisions increasingly depend on simulations provided by computer models. In this talk, we formulate the validation of a given model/code as an iterative construction process that mimics the implicit process occurring in the minds of scientists. We offer a formal representation of the progressive build-up of trust in the model. We thereby replace static claims on the impossibility of validating a given model/code by a dynamic process of constructive approximation. Our procedure factors in the degree of redundancy versus novelty of the experiments used for validation as well as the degree to which the model predicts the observations.