Verification and Validation for Magnetic Fusion:
Moving Toward Predictive Capability

51th Annual Meeting of APS - DPP
Atlanta, 11/3/2009
M. Greenwald
Acknowledgements

- This talk grew out of work by a BPO/TTF Task Group – thanks to all for their contributions:
  - Paul Terry (Chair)
  - Martin Greenwald
  - Jean-Noel Leboeuf
  - George McKee
  - David Mikkelsen
  - Bill Nevins
  - David Newman
  - Daren Stotler

- Summary published as: Physics of Plasmas 15, 062503 (2008)

- Draws on extensive literature (especially CFD); see bibliography for sources
Note that the 4 talks in the following invited session are all interesting examples of comparisons between experimental measurements and simulations

KI3.00001: Simultaneous Measurement of Electron Temperature and Density Fluctuations in the Core of DIII-D Plasmas
A.E. White

KI3.00002: Comparison of experimental measurements and gyrokinetic turbulent electron transport models in Alcator C-Mod plasmas
Liang Lin

KI3.00003: Probing Plasma Turbulence by Modulating the Electron Temperature Gradient
J.C. DeBoo

KI3.00004: Fluctuation-Induced Particle Transport and Density Relaxation in a Stochastic Magnetic Field
David L. Brower
Motivation

- There have been dramatic advances in the scope and power of numerical simulation, driven by
  - Advances in theory
  - Better algorithms
  - More powerful computers
  - Improved comparisons with experiments, better measurements
- But “virtual reality” from simulations is nowhere in sight
- We want to improve our ability to predict
  - Demonstrate and embody fundamental understanding
  - Move beyond purely empirical approaches
  - Meet programmatic goals
- As a science program - How do we know what we know?
Plasma Theory Is, Of Course, A Formidable Challenge

- Problem is not tractable in closed form
  - disparate time and spatial scales
  - extreme anisotropy
  - complex geometry
  - essential non-linearity

- Direct integration of Boltzmann-Maxwell equations with full resolution is well out of reach.
Build “Models” Often for Subset of Temporal or Spatial Ranges – Domain Decomposition

- So our approach has been to develop models which...
  - obtain exact solutions to approximate equations or
  - approximate solutions to (somewhat less) approximate equations
- A model can be defined as “a representation of the essential aspects of some real thing, in an idealized or exemplary form, ignoring the inessential aspects” (Huber)
- The hard bit is identifying and demonstrating what is essential in each case
- We test the model to gain confidence that the approximations are “inessential” - this talk is about what that might mean
Theory/modeling and experiments as distinguishable but mutually dependent activities have an important place in the history of philosophy and science.

**Rationalism** – logical development of a model based on indisputable axioms – pure logic

(knowledge gained through the senses is always confused and impure)

**Empiricism** – requires that every axiom, deduction, assumption or outcome be empirically confirmed (only trust knowledge gained through the senses)

---

**Plato**

**Descartes**
Historical Note

Theory/modeling and experiments as distinguishable but mutually dependent activities have an important place in the history of philosophy and science.

Rationalism – logical development of a model based on indisputable axioms – pure logic (knowledge gained through the senses is always confused and impure)

Empiricism – requires that every axiom, deduction, assumption or outcome be empirically confirmed (only trust knowledge gained through the senses)
In Fact, Experiments and Computation Have Well-Known Complementary Roles and Strengths

**Theory/Simulation**
- provides predictive capability and fundamental understanding
- near perfect diagnostics
- high degree of flexibility computer “experiments”
- often cheaper and faster

**Experiments**
- test theory/models/codes
- extend performance (fusion)
- discovery

**with**
- “perfect” model (reality)

**but**
- highly incomplete and imperfect measurements
- lower degree of flexibility
Accuracy and Reliability of Codes Can Have Important Consequences

- Impacts
  - Safety
  - Environmental
  - Economic
  - Legal

- Tests of code reliability can be part of regulatory schemes
MFE Is Moving Into This High Consequence Domain

- ITER is a licensed nuclear facility
- At >$10 billion, ITER has global visibility
- Steps beyond ITER will have very specific requirements for plasma behavior
- How do we gain confidence in our predictions?
Introduction: Verification and Validation

- Since large simulations are increasingly our means for prediction, we must develop ways to assess their reliability.

- V&V are essentially confidence building activities – an accumulation of evidence that our codes are correct and useful.

- V&V can be a more methodical approach to activities already carried out at some level in our program.

- These activities are closely related to, but not identical to the scientific method.

- Experience (from our field and others) suggest that we need to make this process more systematic, quantitative, more rigorous and better documented.
A Standard Set of Terms and Definitions Have Emerged

**Verification** assesses the degree to which simulations correctly implement a physical (conceptual) model

**Validation** assesses the degree to which a physical model captures “reality” through comparison with experimental measurements

- These words are almost synonyms in ordinary usage, their use for the specific activities described here is arbitrary, but not unimportant
  - Model testing is a collective activity – standardization keeps us all on the same page
  - Using a common vocabulary allows us to interact more effectively with other fields
  - Lessons learned are more effectively shared
Various Authors Have Attempted to Show These Relationships Graphically

Schlesinger 1979
Experiments are the (Imperfect) Intermediary with Reality

- **Reality**
  - Experimental Design with Engineering Constraints
  - Measurements: Incomplete and with Uncertainties

- **Experiments**
  - Physical Insights
  - Code Validation

- **Model Development**
  - Theory/Computation (Code Verification)
Verification – “Solving the Equations Right”

“Verification: substantiation that a computerized model and its solution represent a conceptual model within specified limits of accuracy”

- Conceptual model is embodied as a set of equations, algorithms, approximations and assumptions
- Verification assesses computational model (code) and solutions
- Verification is essentially a problem in applied math and logic – though generally not one with a rigorous solution
- In contrast, Validation assesses the conceptual model, once a code and solution are verified
Verification – Assess Accuracy of Simulation

Two generic sources of error to identify and eliminate

- Problems with the code
  - errors in implementing the conceptual model
  - algorithms or numerics
  - coding errors, language or compiler bugs

- Problems with the solution
  - spatial or temporal gridding – discretization errors
  - convergence difficulties
  - numerical noise (e.g. accumulation of round-off errors)
  - problems with boundary or initial conditions
Verification - Methodologies

- Ideally, simulations are compared to highly accurate solutions – analytic or well-known numeric benchmarks
- These are rare – especially for realistic parameters and for coupled problems.

So supplement with

- Software quality assurance practices
- Formal discretization and convergence tests
- Conservation tests
- Symmetry tests
- Manufactured solutions (difficult for coupled, nonlinear systems)
  - Start with made-up solution to the set of equations
  - Modify equations by adding ad-hoc source terms which make that solution exact – then run simulation and compare
- Code to code comparisons (benchmarking)
Example: Convergence Studies for ICRF Spectral Solver (TORIC)

Simulated $^3$He minority absorption changes dramatically as mode-converted wave fields are converged!

- $N_m = 63$
- $N_m = 161$
- $N_m = 255$
- $N_m = 511$

$N_m = 15$

$N_m = 255$
Code to Code Comparisons: Benchmarking

- It is plausible that successful comparison between calculations solving the same set of equations, with the same set of initial and boundary conditions builds confidence in the codes (e.g. Cyclone Project, Dimits et al, 2000).
- It is argued that this is particularly true if the calculations implement very different methods.
- In a very well documented example (Nevins et al, PoP 2006) compares electron thermal diffusivity calculated with PIC and continuum codes.
- While valuable, these exercises must be built on systematic verification studies of individual codes.
- Disagreement demonstrates that at least one of the codes under consideration are wrong.
- **By itself, agreement does not prove that all are correct.**
Estimating Simulation Uncertainties – A Challenging Problem for Coupled Nonlinear Calculations

- Dealing with uncertainties from discretization
  - Grid refinement – convergence studies
  - Comparison of various grid geometries
  - Compare low and high-order solutions on same grid
  - Conservation and symmetry laws

- For evaluating uncertainties arising from boundary or initial conditions, or strong sensitivities - ensemble computing – carry out many runs in multi-dimensional space
  - But computational requirements can diverge
  - Challenge: How to reduce the computation of the full uncertainty space (by a huge factor) to something computationally tractable?

- Explore sensitivity in reduced spaces
- Sample in important parameters
Validation – “Solving the Right Equations”

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

- Validation is an essentially a physical problem
- No clearly defined end point – an ongoing activity not a one-time, up or down decision
  - Validation failures should not be seen as personal failures, but rather a natural part of the scientific process and an essential element in the development cycle.
  - Despite known deficiencies in a model, comparison with experiments at an early stage has a useful purpose – developers may be able to identify and concentrate on improving the most important or uncertain elements of the model
Note the Conditional Nature of V&V

- To avoid unbounded philosophical questions, V&V are best defined
  1. for a targeted class of problems and applications
  2. for a set of specified variables
  3. at a specified level of accuracy
- Technically, we validate a set of calculations then draw inferences about
  the validity of a model.
- Similarly, we verify a set of calculations, then draw inferences about the
  code.

Together, the goal of validation and verification is an assessment of the
extent to which a simulation represents true system behavior **sufficiently to be useful.** Both should be viewed as ongoing, open-ended processes.
Some Other Useful Terms

- **Qualification**: “Theoretical specification for the expected domain of applicability for a model”

- **Calibration**: “Adjustment of parameters in a computational model in order to improve agreement with data”
  - Calibration may be justified to account for physics beyond the scope of the model, however it should not be used to obscure essential errors in a model or its implementation.

- **Prediction**: “Use of a code outside its previously validated domain to foretell the state of a physical system”
  - Or more modestly “A simulation result for a specific case that is different from cases that have been validated”
  - Accuracy of prediction is not proven, but inferred from V&V activities
Prediction: Interpolation

- Application entirely within validation domain
- Prediction = Interpolation
- (For multi-dimensional parameter space, data may be sparse so this may not be trivial)
- Estimating prediction uncertainty from validation error is relatively straightforward

• Basis for sound engineering practice
Prediction: Strong Inference

- Application partially overlaps validation domain
- Prediction can require modest extrapolation
- Large errors may arise from transitions or bifurcations

- Physical coupling, complexity
- Physical parameters, geometry, boundary conditions

Validation Domain
Application Domain
Prediction: Extrapolation

- Prediction requires large extrapolation
- Very difficult to estimate prediction errors!
- Requires comprehensive understanding of underlying science!
- Proceed cautiously!

M. Greenwald, et al., APS-DPP November 2009
Validation: We Have to Confront the Significance of the Comparisons We Make

- What constitutes agreement?
- What inferences can we draw from the agreement?
- Uniqueness: Which measurements are important discriminators between models?
  - Very different models may predict essentially the same values for some quantities
- Sensitivity: Some measurable quantities vary strongly with certain parameters
  - For example, consider the dependence of energy flux on plasma gradients near marginal stability
  - Agreement can be extremely difficult for some quantities and too easy for others
Primacy Hierarchy
Not All Comparisons Are Equally Meaningful:

- We can try to distinguish between basic vs composite quantities
- Rank measured quantities in terms of the extent to which other effects combine
Example 1: Turbulence and Transport

1 (Autopower Spectra)  2 (Correlation Lengths)  3 (Heat Flux)

(C. Holland, et al.)

<table>
<thead>
<tr>
<th>D3D 128913</th>
<th>Expt</th>
<th>GYRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1300-1700 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r/a = 0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q_i (kW/m²)</td>
<td>32 ± 6</td>
<td>38 ± 7</td>
</tr>
<tr>
<td>Q_e (kW/m²)</td>
<td>26 ± 7</td>
<td>34 ± 7</td>
</tr>
<tr>
<td>RMS δn_e (%)</td>
<td>0.6 ± 0.1</td>
<td>0.56 ± 0.01</td>
</tr>
<tr>
<td>RMS δT_e (%)</td>
<td>0.4 ± 0.2</td>
<td>0.66 ± 0.02</td>
</tr>
</tbody>
</table>
Example 2: ICRF Heating

Primacy Level

1 (wave fields) 2 (velocity distribution) 3 (power deposition)

\[ \tilde{n} \]
\[ k \]
\[ \tilde{E} \]

Velocity space distribution \( f_i(v) \)

Heat Deposition

Profiles
Example 2: ICRF Heating - Results

Primacy Level

1 (wave fields)  2 (velocity distribution)  3 (power deposition)

Nelson-Melby PCI synthetic diagnostic for TORIC code
The Primacy Hierarchy Helps Address The Issue Of Discrimination

- Comparison at several levels in the hierarchy is best practice
- In general, discrimination between models is reduced as one goes up the primacy hierarchy
- It may be possible to identify ways in which physics cause uncertainties and errors to cancel
- The form of the hierarchy is not necessarily unique – the important thing is to come to grips with the issue
Advanced Diagnostics Will Be Critical For Validation of Fusion Codes

- For example, turbulence visualization diagnostics are providing unprecedented views into plasma dynamics

- How to use these capabilities for quantitative comparisons with codes?
Synthetic Diagnostics

• Validation requires comparison of identical quantities

• Diagnostics often can’t make local measurements of fundamental quantities

• Inverting the data may be impossible or may introduce artifacts

• To help with this problem, synthetic diagnostics have been developed as post-processors for many codes

• The synthetic diagnostic attempts to replicate, numerically, the physical processes and geometry along with any temporal or spectral averaging – essentially an exercise in phase-space geometry.

• Comparison between the synthetic diagnostic and data is direct (but at a cost - some power of discrimination may be lost)

• Thorough and careful characterization of diagnostic is required.

• The synthetic diagnostic code may be quite complex and must be carefully tested.
Synthetic Diagnostics Example

- Comparison of radial correlation of density fluctuations
- Proper treatment of diagnostic resolution brings simulation into reasonable agreement with experiment.
- From Holland et. al. PoP 2009

FIG. 12. (Color online) Comparison of density fluctuation radial correlation functions calculated for the unfiltered GYRO data (●), the synthetic BES data (■), and experimental data (♦) at (a) $\rho=0.5$ and (b) 0.75.
Powerful Time Series Analysis Methods May Provide Better Sensitivity and Discrimination

Harmonic analysis techniques:
- Short-time Fourier transform
- Fractional Fourier transform
- Bispectral analysis
- Continuous wavelet transform
- Chirplet transform

Chaotic analysis
- Fractal dimension (correlation dimension)
- Recurrence analysis, periodicity or cyclic analysis
- Lyapunov exponents

Principal components analysis

And many others
Perhaps These Tools Will Help Discriminate Between Models

- Physically, $k$ spectrum arises from drive, dissipation and nonlinear coupling
- Historically agreement is easier than for other quantities

Increasing model complexity, analysis sophistication $\Rightarrow$

- Higher order moments (e.g. bicoherence), though harder to measure, may provide better discrimination
Quantitative Analysis, Data Quality and Sources of Error

• Validation requires careful **quantitative** consideration of uncertainties and errors in both experiments and simulations
• Verification should precede validation to minimize problems with the codes and to assess simulation errors
• Sources of errors in experiments – systematic and random (reducible and irreducible)
  – Conceptual errors with measurement techniques
  – Differences arising from temporal or spatial averaging
  – Statistical or counting errors
  – Calibration errors
  – Electronic noise and data acquisition errors
  – Data reduction errors
Bringing It All Together
Assessment of Comparisons: Validation Metrics

- Metrics: Develop overall **quantitative** assessments
  - No unique and “correct” way to define a validation metric
  - Quantitative, but mathematical rigor is not possible
  - Confront disagreement in detail
  - Account for estimated errors in experiment and model

- Use theory to help identify quantities that discriminate between models

- Identify quantities that have low sensitivity to poorly measured parameters used by the models
Example Of A Simple Validation Metric

\[ V = 1 - \frac{1}{n} \sum_{i=1}^{n} \tanh \left| \frac{y_i - Y_i}{Y_i} \right| \]

Where \( y_i \) is a set of simulation results and \( Y_i \) are experimental measurements. (Errors in both are ignored)  Employs \( L_1 \) norm

\[ V = 1 - \frac{1}{n} \sum_{i=1}^{n} \tanh \left( \frac{y_i - \hat{Y}_i}{\hat{Y}_i} \right) \]

If ample experimental data is available, use mean of measurements

\[ V = 1 - \frac{1}{n} \sum_{i=1}^{n} \tanh \left( \left| \frac{y_i - \hat{Y}_i}{\hat{Y}_i} \right| + \frac{S_{\hat{Y}}}{\hat{Y}} \right) \]

Include normalized standard error, \( S_{\hat{Y}} \), in calculation of experimental mean
Other Possibilities for Validation Metrics

Chi squared
Well known statistic
Euclidean or $L_2$ norm

$$V = \chi^2_y = \frac{1}{N \text{ degrees}} \sum_{i=1}^{n} \left( \frac{y_i - Y_i}{\sigma_y + \sigma_Y} \right)^2$$

Confidence intervals with estimated error

$$\tilde{E} - t_{0.05, N_{DOF}} \frac{s}{\sqrt{n}} < E \text{ (true error)} < \tilde{E} + t_{0.05, N_{DOF}} \frac{s}{\sqrt{n}} \Rightarrow V = \frac{\tilde{E} + 2t_{0.05, N} \frac{s}{\sqrt{n}}}{Y}$$

Maximum error
$L_\infty$ norm

$$V = \left| \frac{\tilde{E}}{Y} \right| = \max \left| \frac{y_i - Y}{Y} \right|$$

Hypothesis testing – classical hypothesis testing is aimed at returning a binary result: could be applied, perhaps not particularly helpful for validation.

Composite metrics – Sum metrics for different quantities (higher for spanning primacy hierarchy), different experiments (higher for repetition, more points for independent tests), weighted for uncertainty and sensitivity
A Few Words About Graphical Methods

- We’ve stressed here quantitative techniques – the “vugraph norm” is often deprecated in discussion of validation.
- However, the power of good graphical techniques should not be underestimated – especially for data exploration.
- The best practice probably combines both approaches.
- Example:
Validation Should Build from Simple to Complex Systems

- It’s usually not wise to start with the most complex problems – there are too many ways to go wrong

- Consider field of aerodynamics
  - Start with laminar flow around sphere
  - Move on to more complex flow regimes, turbulence
  - Move to more complicated shapes
  - Analyze full airframe
  - Include operation of engines and control surfaces

- Often called the “Validation Hierarchy”
Validation Hierarchy

**Increasing**
- Realism in physics and geometry
- Coupling of physics
- Complexity

**Complete Systems**

**Subsystem Cases**

**Unit Problems**

**Decreasing**
- Number of code runs
- Number of Experiments
- Quality and quantity of data
- Info on initial and boundary conditions
Editorial Opinion: The Validation Hierarchy Points Out A Notable Gap In the MFE Program

- Typically most attention is lavished on the most complex (typically highest performance) devices at the top of the hierarchy
- But we have not validated many of the basic phenomena that the codes are predicting
- Many basic linear processes were originally observed on low temperature devices (linear and toroidal plasmas)
- We should think about taking the same approach for fundamental nonlinear processes
- This probably require configuring codes to work in simpler geometry and in unfamiliar parameter ranges
- Basic experiments must be well diagnosed ($$)
Some Efforts In This Area Are Underway

- “Simple” experiments dedicated to validation might better isolate critical physics
- In this example CSDX (UCSD) tests Hasegawa-Wakatani model
  - Move to testing modern gyro-kinetic models
- Other examples include CLM (Columbia), LAPD (UCLA), Helimak (UT)
  + notable efforts at several European Universities
Summary

- Despite dramatic advances in computational plasma physics, we are still far from solving the critical problems.

- Experiments and simulation are complementary rather than competitive approaches – science benefits from a continuous and ongoing collaboration between them.

- **Verification and Validation can provide a framework for carrying out the collaboration in a methodical and systematic way**

- The goal is to increase confidence in our predictive capability

- This will require new modes of interaction – openness about uncertainties, errors and limitations of methods is essential


Bibliography 2


