Verification and Validation in Fusion
Toward Guidelines and Good Practices

Presented by P.W. Terry

for

M. Greenwald, J.-N. Leboeuf, G.R. McKee, D.R. Mikkelsen, W.M. Nevins, D.E. Newman, D.P. Stotler, and P.W. Terry

Verification and Validation Task Group, USBPO and TTF

Verification and validation in fusion: a brief history

- Pioneering efforts: Model/experiment comparisons
  Qualitative; limited assessment of uncertainty, sensitivity, error
  Outcome inconclusive; incomplete, unconvincing methodology
- Oberkampf (SLC TTF): Standardized procedures for testing models
  - Practiced in stockpile stewardship, fluid dynamics (engineering performance, software reliability)

  1) Verification: numerical algorithm faithfully solves mathematical model

  2) Validation: Mathematical model faithfully represents real world
- Fusion community: Mostly verification to date
  Orchestrated benchmarking exercises - GEM, CYCLONE
  Presentations at this meeting:
  ECC session on Friday, Edge 4 parallel session, Core 5 (Bravenec)

Verification efforts underway; focus here on collective task of validation

Verification and validation will be a central aspect of fusion science involving modeling, experiment and theory

US 10 year goal: “progress toward predictive understanding”

- Objective: demonstrably predictive models within tolerances

Process of getting there: validation under commonly understood standards for what constitutes agreement between models and experiment

Growing interest at DOE

PoP editorial statement encouraging submissions involving V&V

Significant challenges

- Resource limitations (budget, manpower)
- Complexity of modeling
  - Complexities of turbulence [multiple scales, nonlinearity, geometry (b.c.)]
  - Different regions - different physics, different models
- Difficulties with measurement
  - Limited access
  - Limited diagnostic capability
  - Plasma diagnostics involve significant modeling a priori

Draft editorial policy statement, Physics of Plasmas

Physics of Plasmas editorial board is currently reviewing a draft editorial policy statement proposed by editor Ron Davidson

Notes importance of verification and validation in quest for predictive capability

Excerpt:

. . .Since the ability of a theoretical model to predict plasma behavior is a key measure of the model's accuracy and its ability to advance scientific understanding, it is the policy of *Physics of Plasmas* to encourage the submission of manuscripts whose primary focus is the verification and/or validation of codes and analytical models aimed at predicting plasma behavior. . . . .

Verification and validation will be a central aspect of fusion science involving modeling, experiment and theory

US 10 year goal: “progress toward predictive understanding”

- Objective: demonstrably predictive models within tolerances

Process of getting there: validation under commonly understood standards for what constitutes agreement between models and experiment

Growing interest at DOE

PoP editorial statement encouraging submissions involving V&V

Significant challenges

- Resource limitations (budget, manpower)
- Complexity of modeling
  - Complexities of turbulence [multiple scales, nonlinearity, geometry (b.c.)]
  - Different regions - different physics, different models
- Difficulties with measurement
  - Limited access
  - Limited diagnostic capability
  - Plasma diagnostics involve significant modeling a priori

Fusion community is beginning to think seriously about validation

Setting out guidelines is evolving process - much still to be learned

Hope: validation becomes part of research culture

- We will learn as we go
- “good practices” become better as we learn

Different models will have different levels of validation, guidelines not rigid

- Details will be individualized
- Onus on researcher to make convincing case for validation
- Widely accepted guidelines will build confidence

Outline

- Key concepts
- Approaches to code validation
  - Useful starting points for experiment/model comparison
  - Sources of discrepancy between experiment and models
  - Primacy hierarchy of measured quantities
  - Landscape of model behavior
  - Validation metric
  - Working the primacy hierarchy and model landscape
- Changing the culture of modeling
- Where we go from here
- Questions for discussion

Validation as collective endeavor $\Rightarrow$ standardized concepts

From glossary, key concepts for validation

- Validation - process of determining degree to which model is accurate representation of real world, given intended uses
- Qualification - theoretical specification of expected domain of applicability of model
- Uncertainty - potential deficiency in modeling process due to lack of knowledge, either in model or in experimental data used for validation
- Sensitivity analysis - study of how output variation is apportioned to different sources of variation
- Prediction - use of code outside previously validated domain to foretell state of physical system
- Primacy hierarchy - ranking of measurable quantity in terms of extent to which other effects integrate to set value of quantity
- Validation metric - formula to objectively quantify comparison between experiment and model. Takes into account errors, uncertainties, primacy, sensitivities. Can be designed to assess accuracy, or discriminate between models

Obvious but not-to-be-forgotten points for experiment/model comparisons

Code validation is a joint enterprise between modeling, experiment, theory

Long term product of US fusion sciences: Validated predictive model or set of models for moving to DEMO, commercialization

- Use of common units
  e.g., SI units (including μ_0 and ε_0)
- Full disclosure of simple (easily overlooked) conventions
  e.g., $\sqrt{2}$ in v_{th}
- Common understanding of what quantities are measured or could be measured
  including limitations, effect of modeling in diagnostic
- Application of experimental resources (runtime) for validation work
  may not be the most interesting runs from physics or fusion perspective
- Application of qualified models appropriate to experimental conditions

Important to identify, understand and quantitatively assess sources of discrepancy between models and experiments

This task is central to several validation elements:

- Qualification

  Under what conditions would model deficiencies not be expected to affect a comparison, or to affect only within some tolerance?

- Error and Uncertainty

  What are a priori deficiencies in model or experimental measurement?

- Validation metric

  Assign confidence level to results of validation activity

  Confront disagreement in quantitative detail, figure out its source

  Can deficiencies be quantified?

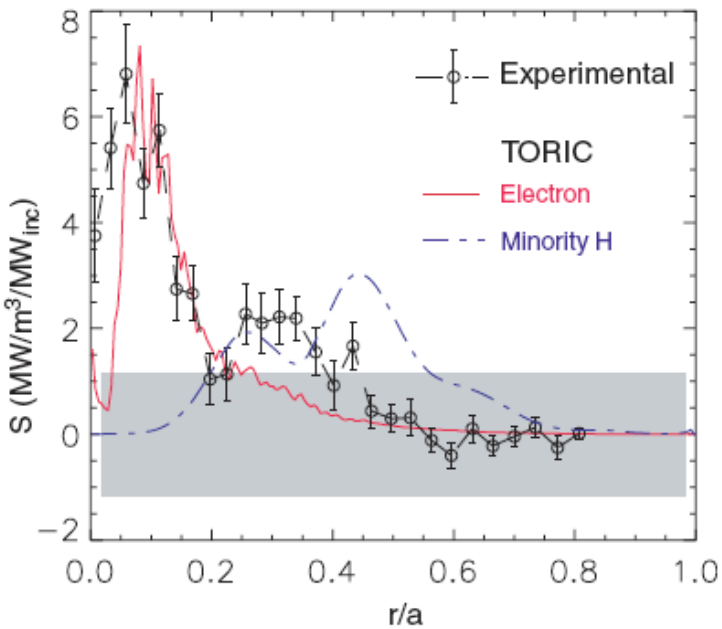
  Can differences in comparison results be reasonably attributed to deficiencies?

  Reasonable: Qualification of model (where and how deficiencies arise)

  Quantitative assessment of deficiencies (magnitude of effect)

  Are there refinements to comparison (in model or experiment) that could establish source of disagreement between model and experiment?

For validation, “generally in agreement” needs to followed up with quantitative analysis of features not in agreement



Mode converted electron heating profile from ICRF in C-Mod

Modeling from toroidal full-wave ICRF

- Agreement is generally good
- Qualitative discussion of
  - -Shift of peak near magnetic axis
  - -Second peak

Needed for validation

- Quantitative analysis - demonstrate sources of disagreement are identified
- Can systematic deviations be bounded?

Discrepancies include statistical error and systematic deficiencies in experiment

Statistical error

- Relatively easy to rate; often exclusive content of error bars
- Important to describe how error bars are arrived at
- Magnitude relies on statistical assumptions that may not be valid

  Large ensembles (Markov), sampling $\Rightarrow$ Gaussian

  Dynamical fluctuations need not obey Gaussian statistics

Uncertainty in experiment (mostly systematic error)

- Equilibrium solver
- Lack of precision in input to equilibrium solver
- Diagnostic sensitivity
- Diagnostic resolution
- Spatial deconvolutions
- Modeling is intrinsic to inferred meaning of diagnostic signals
- Additional processing and interpretation of diagnostic signals

Models and simulations often have numerous uncertainties

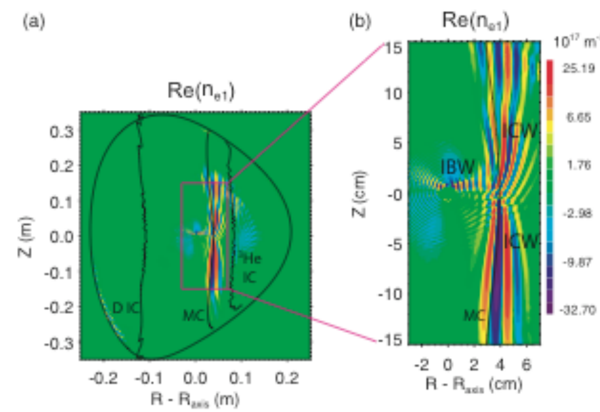
Practical considerations may dictate reduced models even if models with fewer limitations exist ⇒ assessing uncertainties unavoidable

- Mapping magnetic topology to laboratory coordinates
- Equilibrium specification [fixed or variable; subject to modeling]
- Limitations on physical processes included [missing fields, missing kinetic effects, boundary representation, inhomogeneities not included (flow)]
- Limitations on grid resolution [in singular layers; scale ranges]
- Integration time [long time correlations, coupling of transport to turbulent time scale]
- Artificial constraints [fixed profile, flux tube, missing or imprecise experimental data for input parameters]
- Resolution [large scale, small scale, time step]
- Representation of physical dissipative processes; artificial numerical dissipation

Uncertainties associated with diagnostics can be quantified with synthetic diagnostics in simulation

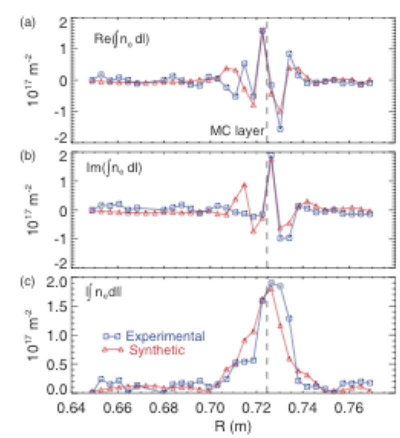
Synthetic diagnostics emulate experimental diagnostics in processing of raw input data

- Include spatial and temporal transfer functions
- Mimic Resolution and sensitivity limitations
- Replicate plasma modeling inherent in diagnostic signal interpretation



Useful for sensitivity studies of experimental data:

Can distinct inputs to diagnostic yield indistinguishable output signals?



Useful for quantifying modeling effects; physics of uncertainties in experimental diagnostics

Important to understand factors in experiment and models affecting fidelity and significance of validation comparisons

- Some measured quantities are more sensitive discriminators between different models
- Some measured quantities are poor discriminators

  Very different models seem to do about as well
- Some measured quantities can be susceptible to false positives
- Some measured quantities have model assumptions folded into them

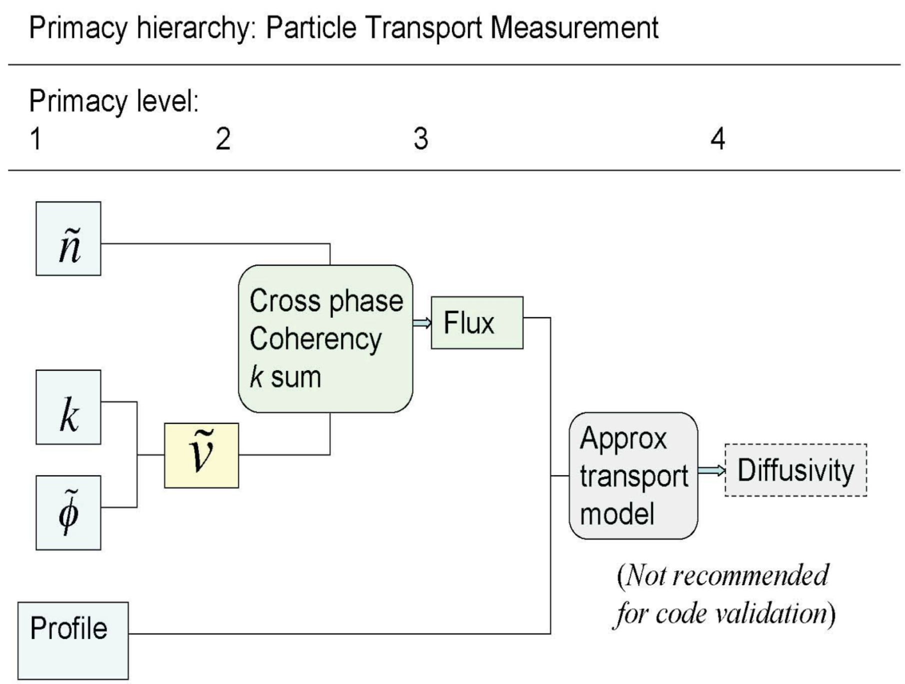
$\Rightarrow$ Not all measured quantities and comparisons are equally meaningful in validation

To assess and quantify these effects:

- Primacy hierarchy (mostly measured quantities)
- Sensitivity analysis (mostly models)

Primacy hierarchy: ranking of measured quantities in terms of extent to which other effects integrate to set value of quantity

Can be constructed in various ways for various types of comparisons



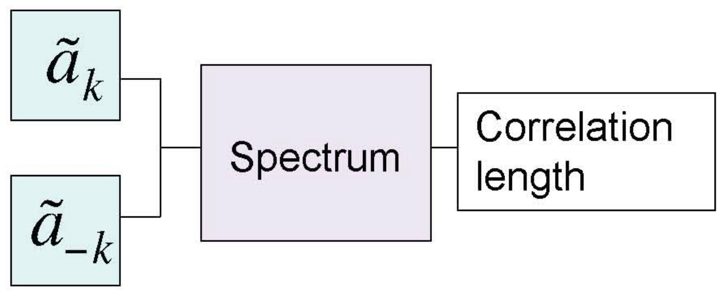
Lower primacy level: fewer effects integrated

Measurements at multiple levels recommended, with awareness of hierarchy

Another example of a primacy hierarchy

Primacy hierarchy: Wavenumber Spectrum Measurement

Primacy level:		
1	2	3

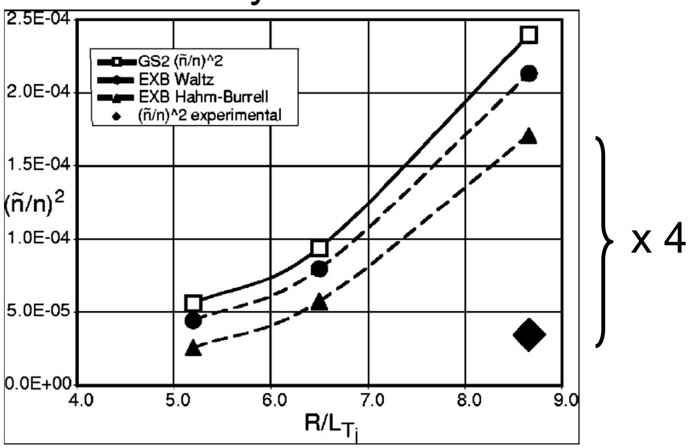


Primacy hierarchy in comparisons with gyrokinetic models: discrepancy with experiment a function of primacy level

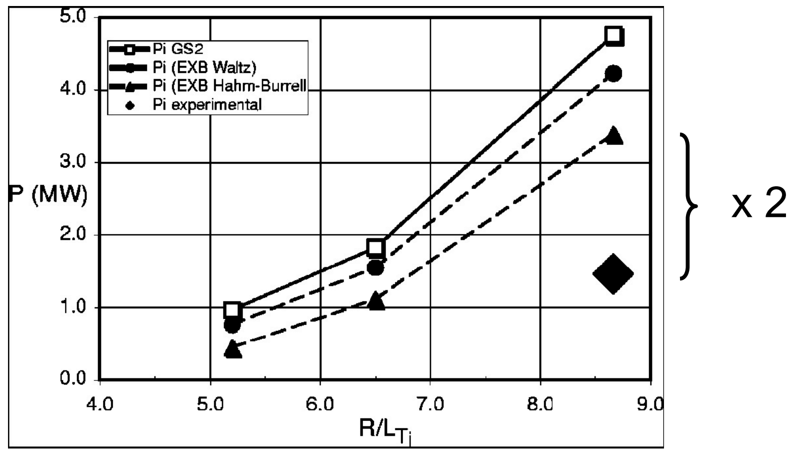
Fluxes (level 3) are in closer agreement than fluctuations (level 1)

Higher level - capability for discrimination between models may be reduced

density fluctuation



flux



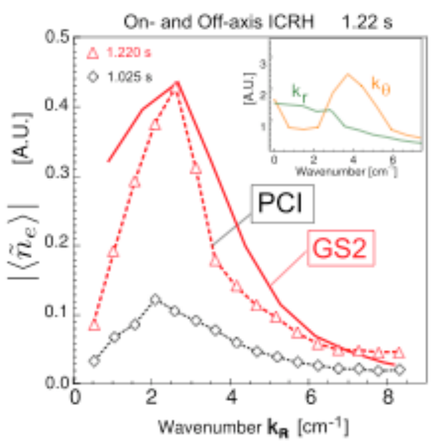
GS2 - Ross and Dorland, Phys. Plasmas **9**, 5031 (2002)

How discrepancy changes from level to level may be function of model, hierarchy

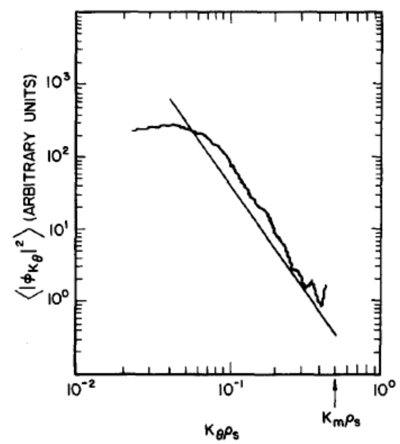
Understanding how effects integrate physically is also useful in assessing comparisons

Historically: *k* spectrum agreement easier to get than other quantities

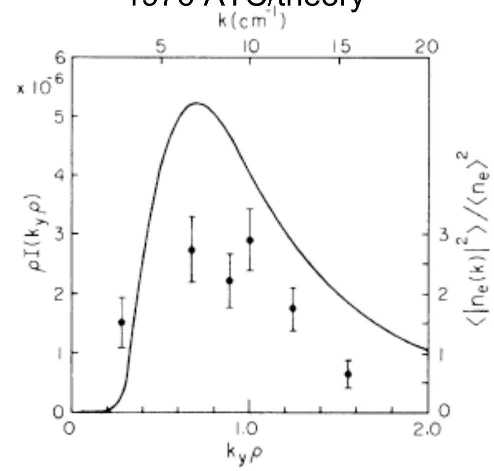
2006 CMod/GS2



1985 TEXT/theory



1976 ATC/theory



Decreasing model complexity, analysis sophistication ⟶

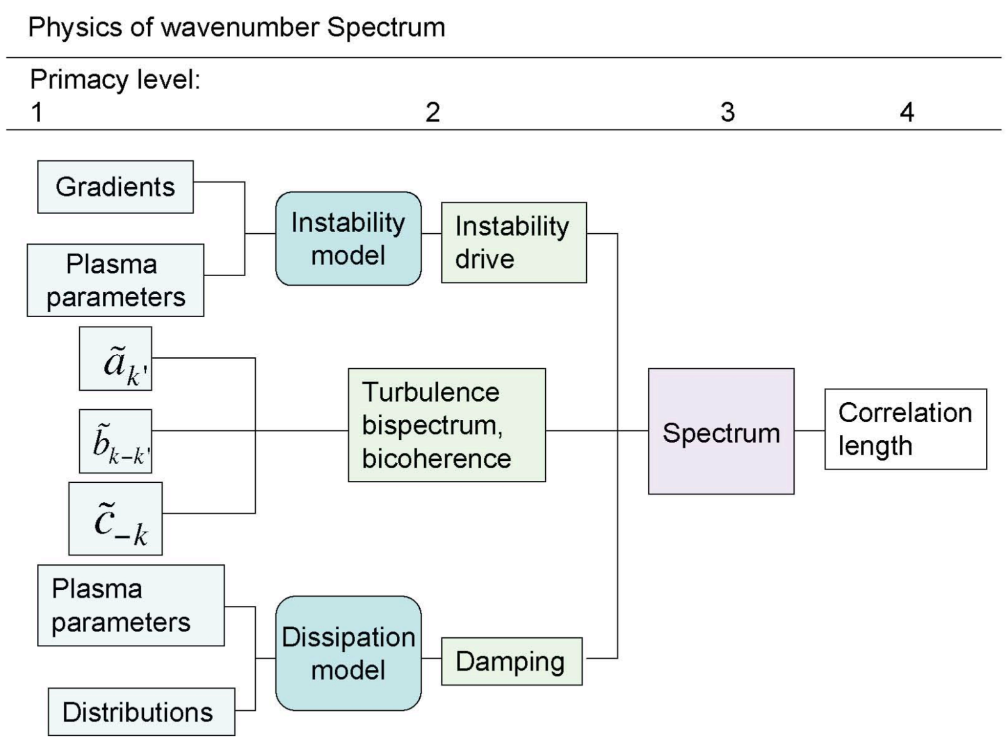
Text/Theory comparison: spectrum comparison great, others not; model discarded

Measurement at multiple levels crucial for credible validation

Poor discrimination capability of spectrum consistent with a primacy hierarchy

Spectrum is amalgam of lower-order processes

Significant physics at lower level goes into spectrum
Folding makes spectrum a poor discriminator between models



Primacy hierarchies are useful in assigning confidence level to validation activities and tracing effects of uncertainties

- Identify possibility that errors/uncertainties are canceling
- Sort out error/uncertainty propagation
  - Holistic view of error/uncertainty sources and folding paths
  - Tracing backwards through hierarchy helps identify most important uncertainties
- Assess ability of measurements to discriminate between different models
  - Synthetic diagnostics applied at higher levels might further degrade ability to discriminate between models → apply to lower levels
- Hierarchies not necessarily unique in form
  - Important to make comparisons at multiple levels
  - Grappling with way effects integrate in comparisons more important than detailed from of hierarchy

Complexity of plasma dynamics must be confronted in validation

Plasma dynamics is nonlinear and complex:

- Bifurcations
  e.g., transitions to enhanced confinement regimes
- Stiffness
  e.g., dependence of fluctuations, fluxes on profiles
- Many parameters
- Extreme sensitivity to certain parameters
  e.g., edge heat flux at L-H transition
- Different behavior in different parameter regimes
  e.g., collisionality switches nonlinear behavior on/off in electron dynamics

Any of above can pose serious problems for validation

How to deal with it:

- Basic theory understanding
- Sensitivity analysis

Theory understanding is crucial in validation

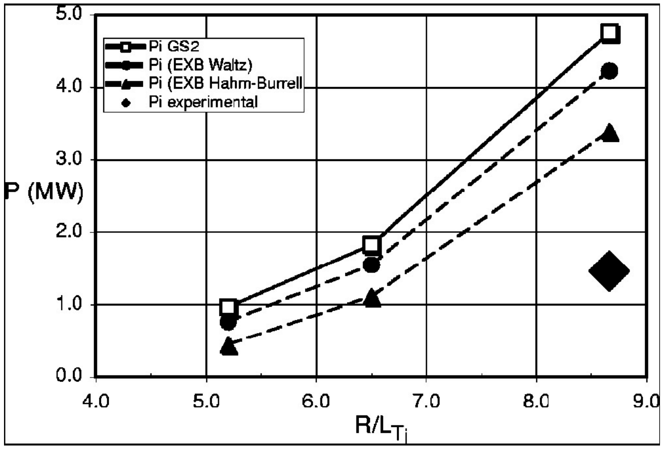
Creates conceptual framework

Identifies features of dynamical landscape

Lays out workings of processes creating landscape

Provides qualitative and quantitative description of dynamics

- Basic scalings
  - Which parameters crucial
  - Where most extreme sensitivities are
- Morphology of dynamical behavior



Identifies previously unknown effects

Example: E×B shear

- Effect that cannot be ignored
- Scalings for effect on fluctuations, transport
- Must be accounted for in validation, doesn't fully close gap in GS2 comparison

Validation will fail or lack credibility if done in theoretical vacuum

Commensurate development of theoretical understanding essential

Validation will not be credible without sensitivity analysis

Certain measurable quantities vary more strongly with certain parameters on which they depend than on other parameters

- Sensitivity of fluctuations, fluxes to profiles is problem in every comparison to date

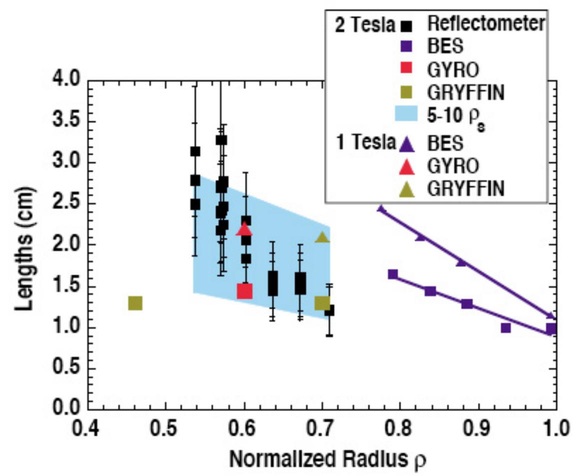
Difficulty:

- Agreement extremely difficult in some quantities
- Agreement too easy in others

Recommendations:

- Must map out sensitivity of all parameters
  - Use theory for guidance
- Looking at quantities that remove sensitivity may help agreement, but may limit ability to discriminate
  - Example: radial correlation length
- Sensitivity to computational effects also important
  - Particle noise
  - Simulation time
  - Resolution

  } assessed in verification



Uncertainties, primacy hierarchies, and sensitivities are rated in a validation metric

Assign confidence level to results of validation activity

Confront disagreement in quantitative detail, figure out its source

Uncertainty (and error) - how to grade it:

- Which uncertainties have been subjected to quantitative testing? Which have not?
- Are there bounds associated with reasonable variation?
  - Use synthetic diagnostics to bound uncertainties associated with resolution, sensitivity
- Are there nonlinear effects from combinations of uncertainties?
- What are their bounds?

Researcher develops grading scheme

- Low score - higher confidence level
- High score - lower confidence level

Validation metric - primacy hierarchy and sensitivity

Primacy hierarchies have ratings associated with primacy levels

Measurement and comparison at multiple levels better than single level

Sensitivity:

- Agreement in quantities with high degree of sensitivity is not rated as favorably as agreement in quantities with low sensitivity
- May be able to use robust quanatities to remove sensitivity
  - Examples: χ_i/χ_e , wavenumber spectrum peaks, have low sensitivity
  - But these may remove ability to discriminate between different models
  - Agreement in quantities with poor ability to discriminate is not rated as favorably as agreement in quantities with good ability to discriminate
- Are there robust predictions that also discriminate?
- High sensitivity: large output uncertainties even for validated models within validation domain
- May be possible to beat down sensitivity problem by reducing uncertainty in source parameters

Special experimental conditions can remove complicating factors or probe lower levels of primacy hierarchy

Special experiments

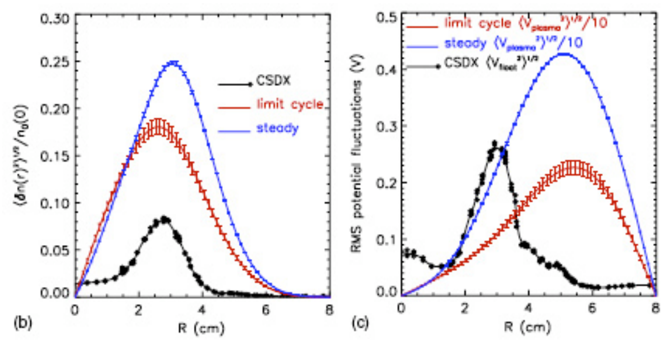
- Simplified geometry/magnetic topology
- Freeze quantities that vary in general
- Parameters in regime of simpler physics
- Fewer disparate effects integrated
- Enhanced diagnostic access

CSDX: linear geometry, controlled turbulence level

Collisional, passing particle drift wave regime

Hasegawa-Wakatani model not optimal for comparison

Comparison with appropriate gyrokinetic model?



Other examples: LAPD, Helimak, Columbia linear machine. . . .

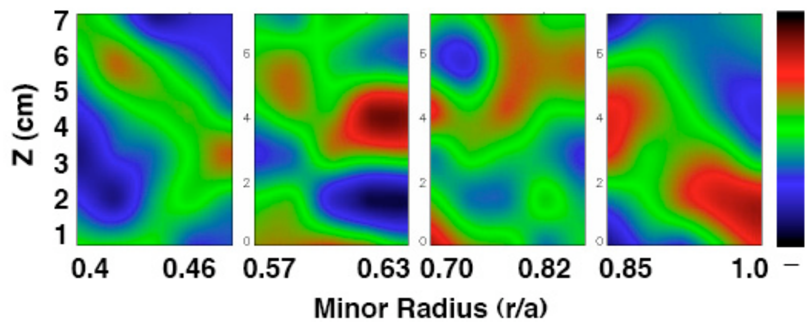
New experiments to propose?

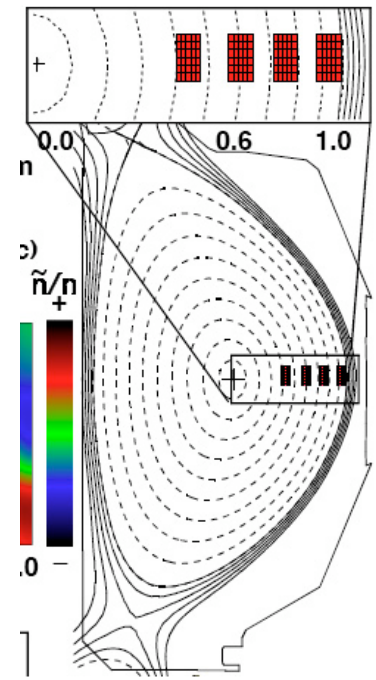
Apply alternate concept experiments to model validation?

Enhanced diagnostic capability, special discharges expand comparison possibilities

Examples of payoffs from enhanced capability

BES sensitivity improvements: fluctuations over wider range of *r/a*





High wavenumber diagnostics: probe electron scale fluctuations

Future development: welcome anything in direction of

More fluctuating fields

Bispectra, bicoherence

Direct sampling of wavenumber

Special discharges: boring for showcasing expt, crucial for verification

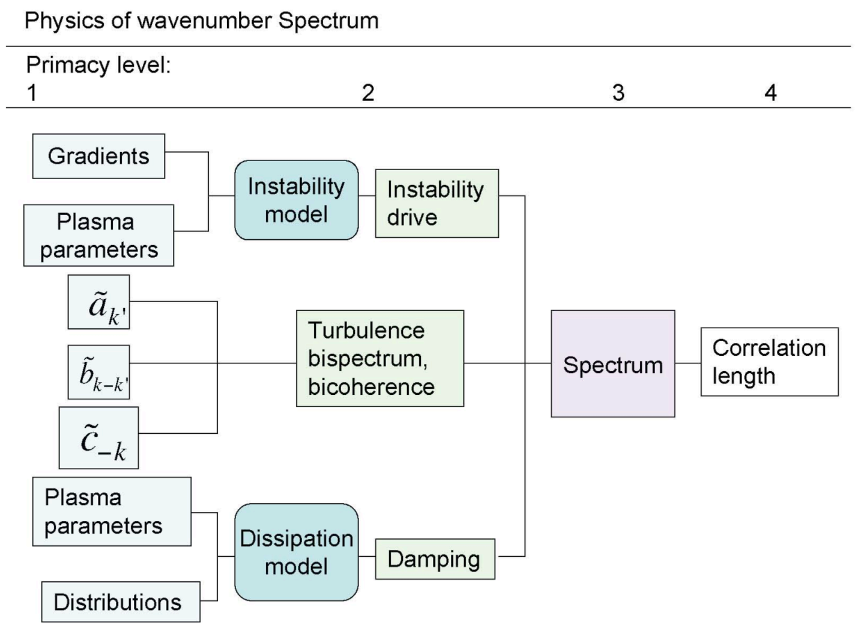
L mode

Long duration, steady state

Develop, use techniques to undo integration of effects

Wavenumber spectrum is poor discriminator between models

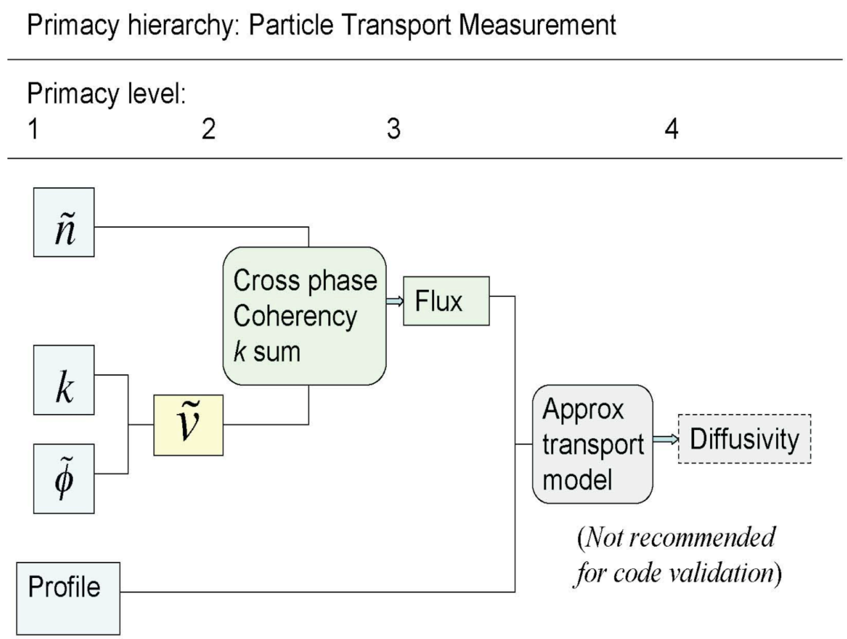
Many effects integrate



Measure bispectrum - infer underlying instability drive (bispectral deconvolution)

Develop, use techniques to undo integration of effects

Diffusivities impose extreme model assumptions



Model fluxes with fractional derivatives

Seek better analysis tools

Create culture of validation

- Joint activity between modelers, experimentalists, theorists
  - TTF has developed right forum for reporting validation efforts
- Run codes in predictive mode
  - Blind, double blind comparison
- Validation as important scientific activity
  - Pursue independently of code building
  - We are working with journals (editors, referees) to welcome V&V papers
- Open reporting of difficulties, shortcomings in comparisons
  - Remove stigma of reporting imperfect results
  - Skepticism about favorable results: hallmark of good science
  - Don’t stop tweaking when agreement obtained (is it really agreement?)
  - *Have seen good examples of openness ⇒ isn’t a career killer*

Where we go from here

Creating guidelines and good practices

- Initial proposals
- Feedback
- Refinement
- Iteration

Technical development

- Robust quantities, sensitivity and discriminating between models
- Ideas for validation experiments
- Diagnostic and analysis technique development
- Do validation with validation metric

Programmatic opportunities

- Fusion Simulation Project - impacting way it is set up
- 5 year planning for major facilities - including validation activities

Questions for discussion

- How do we deal with quantities important for comparisons that are sensitive?
  - Are there robust (not sensitive) quantities that discriminate between models?
- What experimental devices (new or existing) could simplify validation?
- What diagnostic advances (within reach) could facilitate validation?
- What new analysis techniques could enable comparisons?
- Validation metric - how do you assign weights to different uncertainties, errors, primacy level, etc.?
  - How do these weights translate into a confidence level for a reduced code?
- How do we get major experiments to devote runtime to validation?
- How do we reward researchers for doing serious validation work, thereby creating a culture of validation?