Overview of CISM Validation and Metrics
Harlan E. Spence on behalf of the Validation and Metrics Thrust Team

1 Boston University, Boston MA

Metrics
Operational

• Metrics objectively measure a model’s overall improvement over time (Skill Score-SS)
• Operational (top) and science (bottom) CISM metrics are shown in the left-hand column.
• Baseline models, data sets, and physics models from which SS’s are computed shown in other columns.
• Metric studies currently underway to identify most useful SS and prediction efficiency measures.

Metrics Plan:
• CISM, with inputs from science and user communities, selected a limited number of key metrics that will be used to continuously measure progress over the lifetime of the Center.
• Several factors influenced (and continue to influence) CISM metrics:
  • must be comprehensive to measure wide range of models;
  • must be based on direct measurements or derived quantities that will be continuously and reliably available in the foreseeable future;
  • must be quantities related to key space weather effects that we are trying to predict;
  • must be recognized to be important by the space physics science community and/or the operational user community.

Validation Plan:
• “Validation” refers to the broad assessment of a model’s output through comparison with observations.
• What motivates these comparisons?
  • Do models possess all the most important physics and also the algorithms needed to solve the governing equations?
  • Do computational limitations lead to simplifying assumptions that are unjustifiable?
  • Do poorly specified boundary conditions and initial values compromise model predictability?
• To quantify and understand these shortcomings and to guide a developer’s path to eliminating or reducing them, model validation is an essential element of CISM.

Validation Philosophy and Procedure:
• Independent code validation essential for a robust coupled model
• Independent validation efforts will:
  • Fully exercise models to establish physical range of usefulness:
    • Comprehensive event studies — large number of detailed point comparisons
    • Statistical studies — do codes have correct climatology?
    • Note: this is something that has never been attempted systematically in space physics and is a CISM novelty
  • Provide feedback to code/coupling developers
  • Contribute to overall CISM documentation:
    • Internal feedback to code developers: Can the code be run easily with the provided User Manual?
    • Produce community report identifying range(s) of validity of CISM models

CISM Validation and Metrics Status:
• Validation efforts underway with all core models as well as with their available coupled versions
• Detailed studies to define and refine metrics are underway (green), pending (yellow), and future (red)
• Both Operational and Scientific Metrics are being worked
• Examples of activities represented in remaining V/M posters

Overarching Validation/Metric Thrust Goals:
The goals of the Validation/Metrics (V/M) thrust are twofold:
• to identify and document integrated model strengths and weaknesses comprehensively, systemically, and quantitatively through the detailed comparison of model output with observations (Validation); and,
• to measure objectively, using a small set of targeted benchmarks, the long-term trends in model improvement (Metrics).

Background and Motivation:
• Appeal to tropospheric meteorology – metrics such as 36-h, 500-mb forecast provide assessment of long-term improvement to models
• Routine calculation of important operational- and scenario-motivated metrics permits us to measure objectively the ability of CISM models to predict essential space weather quantities
• Rationale for CISM metrics selection has been developed and the list of 29 metrics, along with the baseline models, first-generation physics models, and the data sets needed to compute skill scores, are outlined — work ongoing to define/refine metrics
• While metrics provide a means for the objective assessment of long-term model improvement, model validation—the comprehensive, systematic quantitative comparison of model output with observations—is required for identifying and documenting model strengths and weaknesses
• One CISM goal/outcome is to introduce the process of model validation to the space physics community and the space physics user community.

Cross-functional Integration:
• V/M team works cooperatively with the model developers and the knowledge transfer thrust

Metrics Selection:
• Metrics objectively measure a model’s overall improvement over time (Skill Score-SS)
• By using these models that will be continuously and reliably available in the foreseeable future;
• By using the key space weather effects that we are trying to predict;
• By being recognized to be important by the space physics science community and/or the operational user community.

CISM Validation and Metrics Thrust Goals:
The goals of the CISM Validation/Metrics (V/M) thrust are twofold:
• to identify and document integrated model strengths and weaknesses comprehensively, systemically, and quantitatively through the detailed comparison of model output with observations (Validation); and,
• to measure objectively, using a small set of targeted benchmarks, the long-term trends in model improvement (Metrics).

Background and Motivation:
• Appeal to tropospheric meteorology – metrics such as 36-h, 500-mb forecast provide assessment of long-term improvement to models
• Routine calculation of important operational- and scenario-motivated metrics permits us to measure objectively the ability of CISM models to predict essential space weather quantities
• Rationale for CISM metrics selection has been developed and the list of 29 metrics, along with the baseline models, first-generation physics models, and the data sets needed to compute skill scores, are outlined — work ongoing to define/refine metrics
• While metrics provide a means for the objective assessment of long-term model improvement, model validation—the comprehensive, systematic quantitative comparison of model output with observations—is required for identifying and documenting model strengths and weaknesses
• One CISM goal/outcome is to introduce the process of model validation to the space physics community and the space physics user community.

Cross-functional Integration:
• V/M team works cooperatively with the model developers and the knowledge transfer thrust

Metrics Selection:
• Metrics objectively measure a model’s overall improvement over time (Skill Score-SS)
• By using these models that will be continuously and reliably available in the foreseeable future;
• By using the key space weather effects that we are trying to predict;
• By being recognized to be important by the space physics science community and/or the operational user community.