Model Evaluation and Validation

T. G. Onsager   NOAA/SEC

Scientific case studies, statistical validation, and continuous measures of performance are all important for model selection, transition, and operational use.

- **Scientific and User-Oriented Validation**

- **Model Selection Criteria**

- **Continuous Validation in Operations**

Model Evaluation

Model evaluation requires consideration of operational needs and capabilities, as well as consideration of scientific validity.

- Scientific Validation
  - Provided by the researcher
  - Gives a "proof-of-principle" for the model
  - However, it typically is limited to a few specific conditions,
  - Often does not allow quantitative inter-comparison with other models

- User-Oriented Evaluation
  - Evaluate the relevance for operational needs
  - Determine usability within real-time operations
  - Estimate improvement to current capabilities

Internal SEC Process for Selecting New Models

- Solicit questionaire from modeler – available on Web

- Internal evaluation of models based on three main categories:

  - Strategic Importance to SEC

  - Operational Significance

  - Implementation Readiness

- Estimate cost/benefit to transition to operations

- Make recommendation to management for selection

- Develop test product and begin long-term validation

- Evaluate for final transition to operations

# Operational Execution: What is Required?

Operational execution requires a continuing effort, involving scientists, programmers, forecasters/users, and researchers.

- Maintain a controlled software configuration and concept of operations

- Obtain a continuous measures of performance
    - Long-term assessment of model capabilities
    - Establishes metrics to evaluate new models and/or data

- Make long-term validation available to the relevant research and user communities

- Support operational systems, including data ingest, model execution, user-interface, validation, and archival

Example: Costello Neural-Network K_p Predictor

- Model was acquired by SEC with a minimal validation

- Long-term statistical assessment was performed by Tom Detman to determine the probability of occurrence of the true Kp given a predicted Kp

- Validation results are made available with the real-time predictions to give users an indication of the statistical model uncertainty
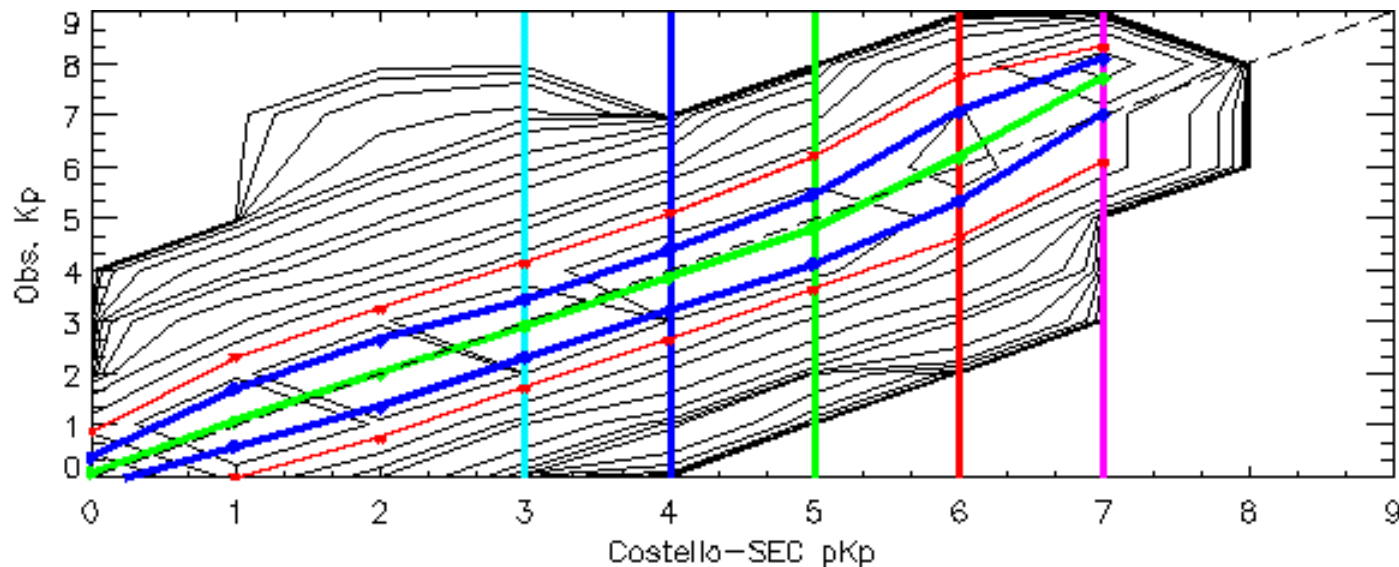
Example: Costello Neural-Network K$_p$ Predictor
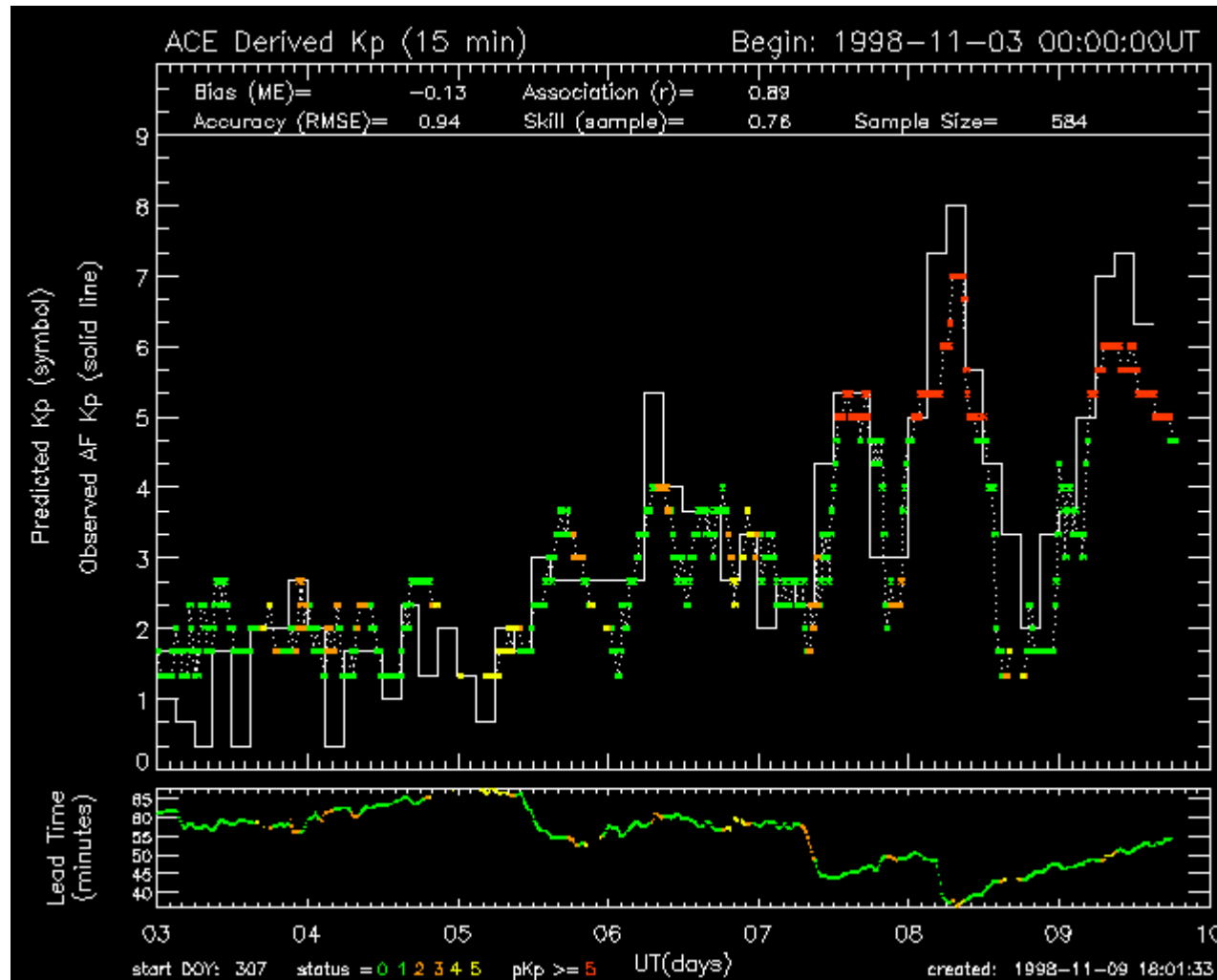
Electrojet Prediction and Nowcast Challenge

Time period chosen:

       March 19-20, 1999 - Steady Solar Wind Driving

Target output:

       Japan 210 chain (K. Shiokawa, Nagoya University and Y. Yumoto, Kyushu University)
       Greenland chain (J. Watermann, Danish Meteorological Institute)

Input data:

       ACE Magnetometer and Plasma data
       Canopus Magnetometer Chain
       Intermagnet Magnetometer Chain
       IMAGE Magnetometer Chain

Evaluation criteria: Root Mean Square Error: $\langle |B_{pred} - B_{obs}|^2 \rangle^{\frac{1}{2}}$