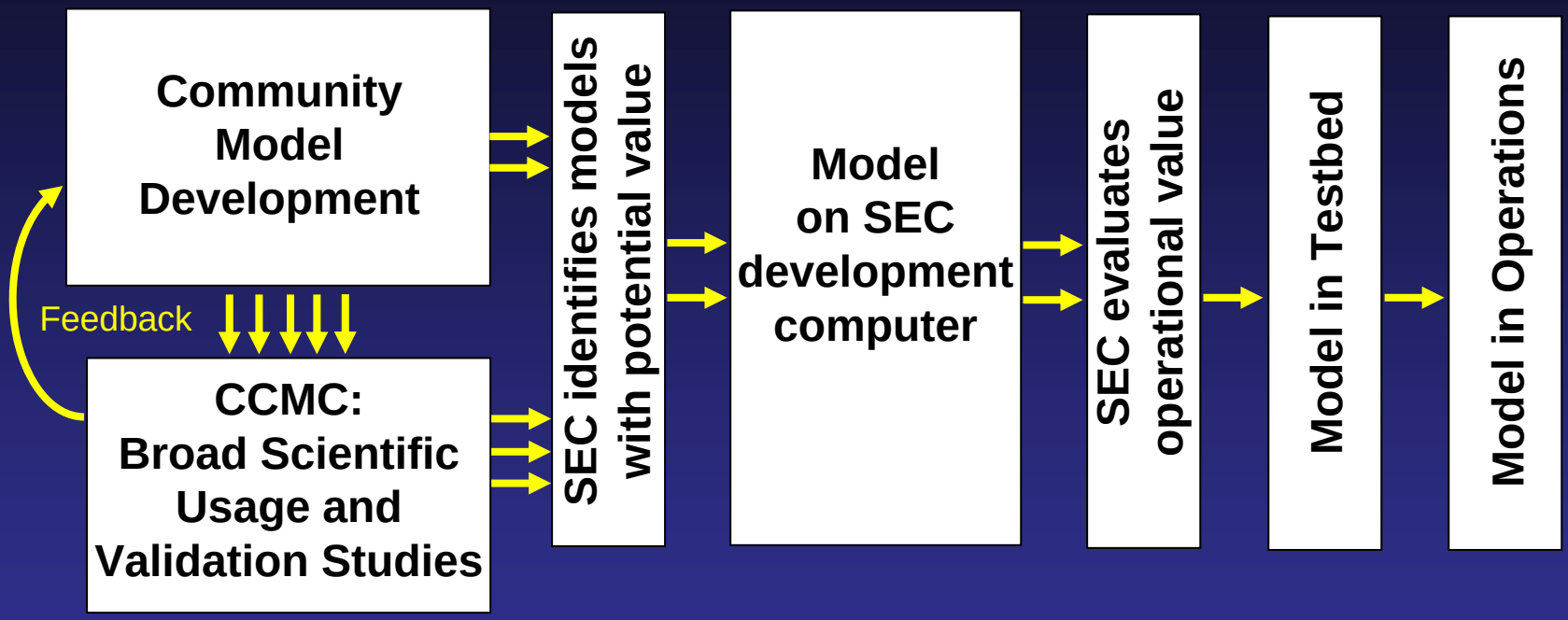
NOAA Metrics and V&V Needs

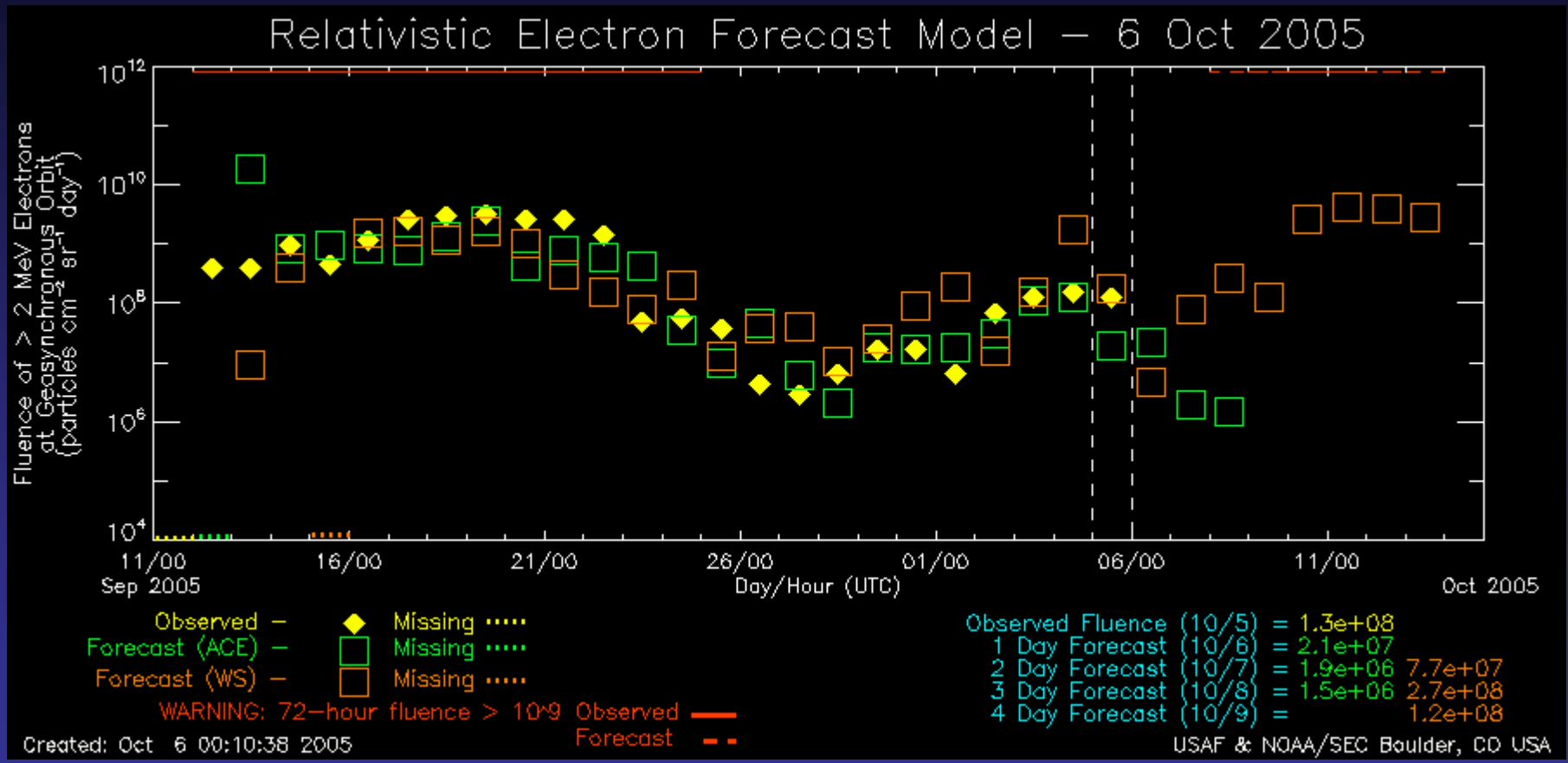
- SEC routinely calculates metrics for forecast products – alerts and warnings
- Some models have performance statistics calculated in real time, others only have historical performance statistics
- Our validation and verification needs fall into three catagories:
  - Scientific Validation – Papers, conference presentations, …
  - Operational Validation – Long-term performance using operationally available real-time data
  - On-going Measures of Performance – How has the model been doing recently under current conditions

Validation Needs Throughout Operational Model Flow



1-day to 8-day Radiation Belt Electron Predictions

Chris Smithtro, AFIT



Recent history of predictions gives a valuable on-going measure of performance.

REFM Skill Scores - 1/1/1996 through 12/31/1999

$$SS = 1 - MSE_{target} / MSE_{reference}$$

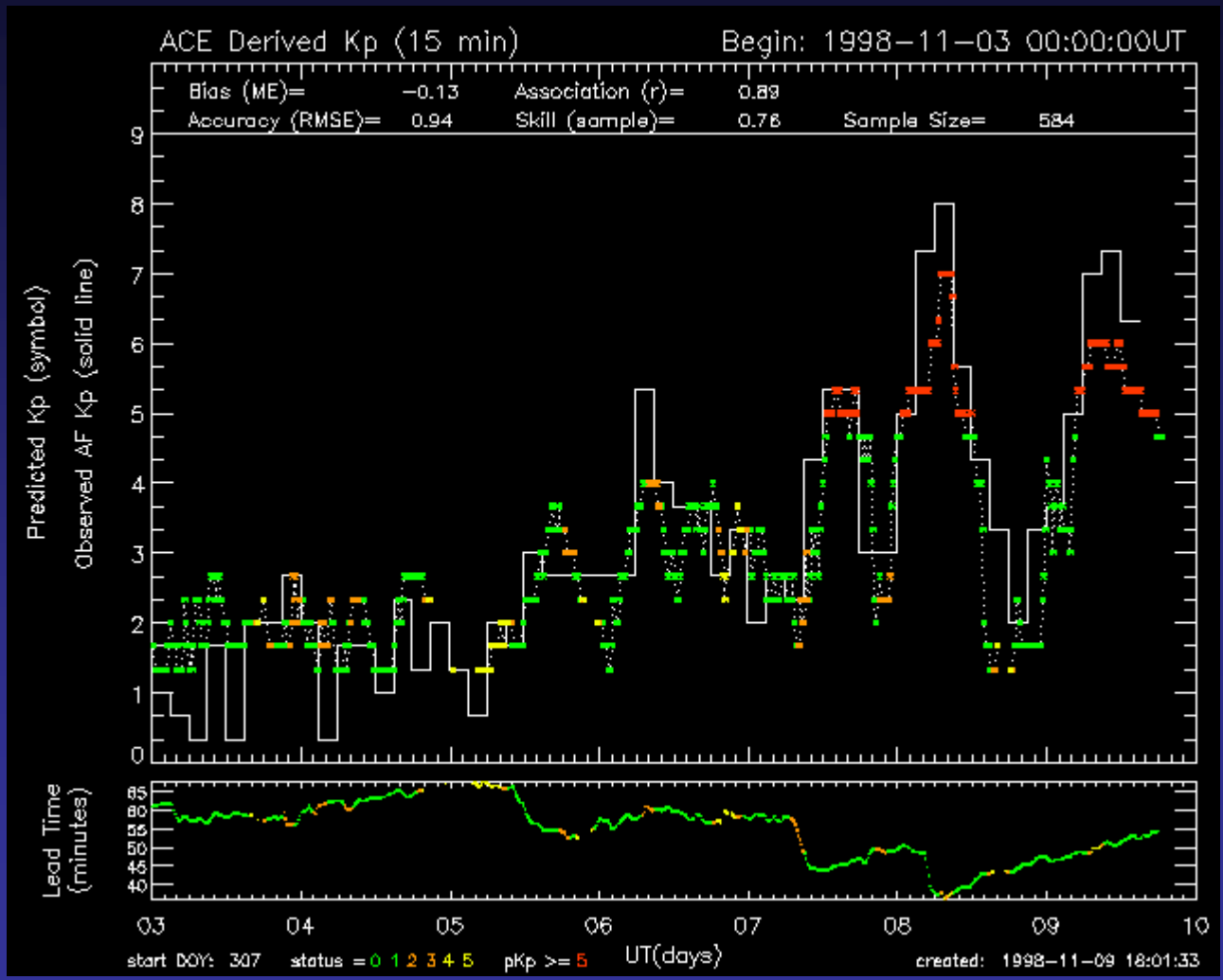
$$MSE = \frac{1}{Np} \sum_{i=1}^{Np} (observed_i - forecast_i)^2$$

Skill Score versus:

Forecast	Mean (PE)	Persistence	Recurrence
+1 Day (ACE)	0.71	0.13	0.79
+2 Day (ACE)	0.39	0.15	0.56
+3 Day (ACE)	0.24	0.26	0.45
+2 Day (WS)	0.32	0.00	0.48
+3 Day (WS)	0.19	0.18	0.39
+4 Day (WS)	0.09	0.26	0.31
+5 Day (WS)	0.06	0.33	0.29
+6 Day (WS)	0.03	0.35	0.26
+7 Day (WS)	-0.01	0.36	0.23
+8 Day (WS)	-0.02	0.37	0.22

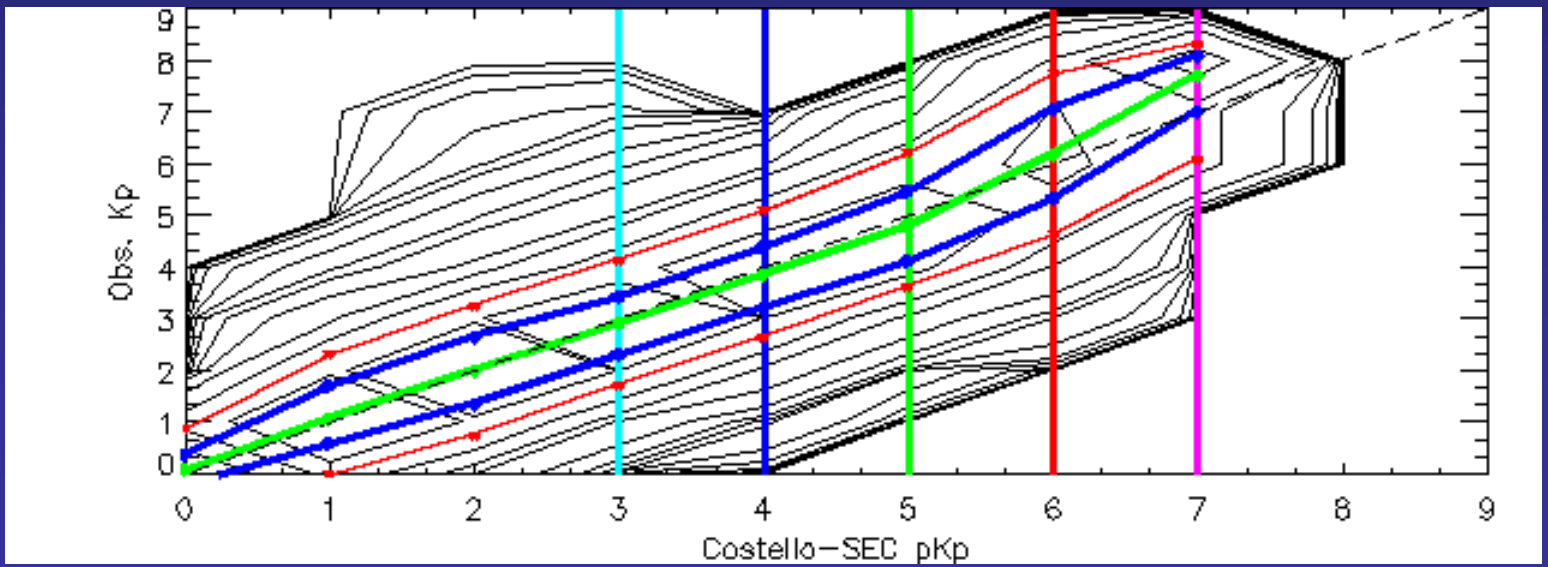
Continuously updating skill scores for the past 30 days are also available on the web.

Costello Neural-Network K_p Predictor



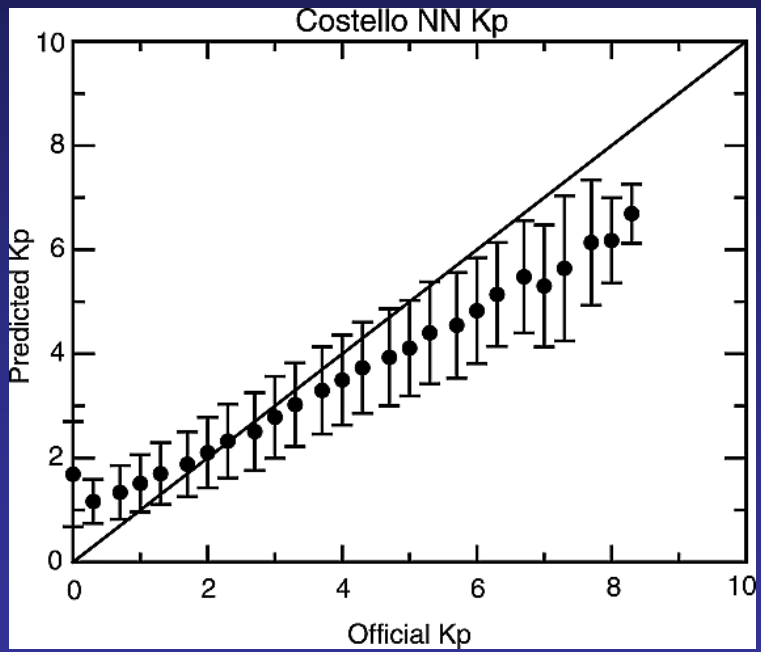
Costello Neural-Network K_p Predictor

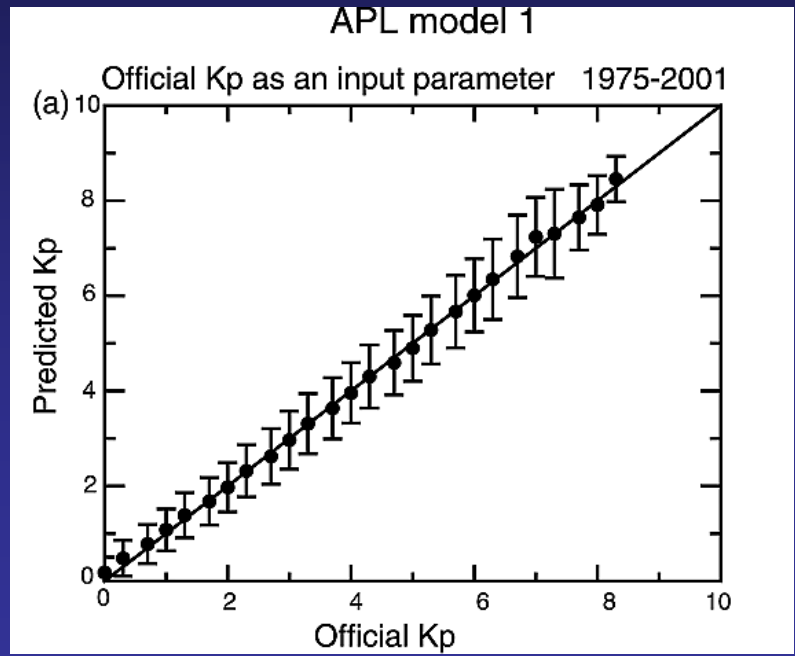
- Model was acquired by SEC with a minimal validation
- Long-term statistical assessment was performed to determine the probability of occurrence of the true Kp given a predicted Kp
- Validation results are made available with the real-time predictions to give users an indication of the statistical model uncertainty



UPOS K_p Versus Costello K_p Predictor

- Comparison of a new model with one in operations – Wing et al., JGR, 2005

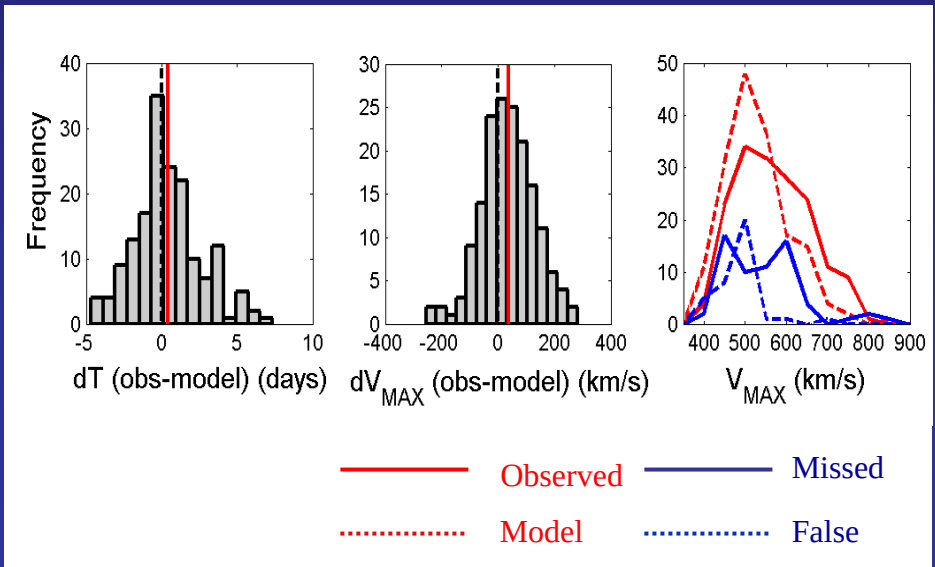




Wang-Sheeley-Arge Model Validation

Contingency tables

		Observed	
		HSE	No HSE
Model	HSE	166	36
	No HSE	64	-



Owens et al., 2005

- WSA is used now in operations to predict the background solar wind speed and IMF polarity
- Validation study demonstrates the model capabilities
- New models will need to show improved capabilities to merit the cost of implementing in operations

Summary

- Our validation and verification needs fall into three catagories:
  - Scientific Validation – Papers, conference presentations, …
  - Operational Validation – Long-term performance using operationally available real-time data
  - On-going Measures of Performance – How has the model been doing recently under current conditions
- Our model selection is based in part on:
  - Is the model scientifically accurate with the best of inputs?
  - Does the model perform sufficiently well over a broad range of conditions of interest?
  - Does the model provide sufficient new value to space weather customers over the value of current services?