

# Apparent Points Of Convergence

1. Model validation is one of CCMC's central responsibilities, and that activity should be continued.
2. The recent GEM metrics competition was a valuable first effort at evaluating space weather metrics. A lot of practical lessons were learned on how to make future contests more efficient and meaningful.
3. CCMC should continue to serve as unbiased judge of metrics competitions. Since CCMC does not have the resources to run all of the competing models, a way must be found for blind model runs to be carried out at modelers' institutions.
4. The present first-priority NSWP metrics (based on electric field for the magnetosphere and electron density for the ionosphere) represent reasonable first efforts at space weather metrics and should be minimally sufficient to satisfy the administrative need to quantify progress of the NSWP over the remainder of its lifetime. However, no single metric can be broad enough to really represent a major area of space-weather science.
5. The present NSWP magnetospheric and ionospheric metrics are not optimal and sometimes give results that conflict with scientific judgment.
6. Additional metrics for use in competitions should be defined with direct input from the users of space weather products.
7. Future metrics competitions must be "blind," with no opportunity for modelers to adjust their models to fit the data. Because the competition was not blind, the results of the GEM metrics competition cannot be used to judge the relative accuracy of the different models.
8. Metrics should be evaluated routinely and for longer time periods than were used for the initial GEM metrics competition.
9. More effort is needed to assure the cleanliness and accuracy and the observational data for metrics competitions. The observers should be asked to specify error bars, if technically possible.

# Points That Were Not Resolved And Will Require More Study

1. It is not clear whether the state of solar-heliospheric science has reached a state of maturity where a quantitative metric is meaningful.
2. Because it wasn't a blind test, it is not clear whether results of the GEM Metrics Challenge represent a scientifically valid first point on the long-time curve that will document the progress of NSWP.
3. The value of keeping scientific metrics distinct from application metrics is not clear. For example, the metric used for the Electrojet Challenge, which was aimed directly at a user needs, may be as scientifically meaningful as the presently adopted NSWP magnetospheric metric. Perhaps the electrojet-challenge metric should be implemented in future CCMC-judged metric challenges.
4. Though the present NSWP ionospheric and magnetospheric metrics could be improved for better conformity with scientific judgment, it is not clear whether that would be worth the effort.
5. It would be possible to develop more sophisticated multi-level sets of metrics that would be more diagnostic of the science. It is not clear whether or not that would be worth the effort.
6. There are several promising approaches to allowing metric challenges to operate efficiently over longer time periods, but it is not clear which is best. One possibility involves running a test one day per month (e.g. world day). Another involves running successive models through the same large "clean" dataset supplied by an operating agency, for which the "right answers" are kept confidential. There are possibilities for automating competitions.
7. There was no consensus on who should be responsible for cleaning input data sets (i.e., identifying and removing bad data points) for future metric competitions