PEHRPP Error Metrics WG Summary

G.J. Huffman, F. Hossain, co-chairs

K. Aonashi, E. Ebert, R. Joyce, C. Kidd, C. Marcos, D. Michelson, S. Nesbitt, A. Rodriguez, J. Turk

The Issue

General feeling that the current understanding of HRPP quality/certainty/errors suffers from a lack of adequate error metrics that are pertinent to users and well-understood

Long-term Vision

Physically based error characterization of retrievals (note GPM)

Consistent set of “basic” metrics

Comprehensive quantitative error model that allows users to specify time and space scales, give the space-time… coefficients associated with a precip data set, and obtain estimated RMS error (diagnostic) or create synthetic precip fields (prognostic):

- basis for computing coefficients will be determined at representative sites in different climate regime for each season and extended to global coverage
- similarly for underlying time/space correlation structure
- both determined for each sensor
- minimum set of coefficient bases, correlation structures, and regimes
- expression for assessing correctness of the error estimate

Short-term Recommendations

Develop a standing working group on error metrics

Agree on a short list of error metrics – each needs confidence intervals

- "traditional" metrics that give insight at the scales of interest
- other metrics suggested by the long-term vision
- fuzzy validation framework
- WWRP/WGNE Joint Working Group on Verification list of metrics
- diagnostics (PDFs, conditional statistics, …)
- examine using transformed data in metrics

Test practicality of these metrics for producers and utility for users

- note Joyce subsetted gridded (30-min, 0.25°) precip data sets from ~15 satellites/sensors (with access software)
- use the metrics results for "more analysis" – engage algorithm developers to understand early results (i.e., multi-temporal RMSE cross-over at 5.5 days between CMORPH and 3B42v.6)
- scales of interest are: finest (by product); 0.25° 3-hourly, daily

Short-term Recommendations (cont.)

Work with input dataset producers to start providing information that supports calculating the error metrics:

- note OPERA end-to-end paradigm for quality information transfer
- probability of precipitation, spread of probability, estimate of detection …
- other input algorithm-specific quality indicators (ambiguous flags, …)
- algorithm-determined or ancillary-data regime information (surface type with terrain and sub-grid water bodies, convective/stratiform, ice content, cloud depth, freezing level, assim. meteorology, storm lifecycle, rain/snow…)
- observation geometry information (scan position, diurnal phase, …)

Get contacts into NWP ensemble error estimation community (D. Michelson to J. Turk and P. Arkin) and check for applicable approaches

Short-term Recommendations (cont.)

Systematize the various validation pages (Australia, U.S., Europe, …) across the various display products

Maintain HRPP quality page on IPWG site to

- link page for current product validations
- facilitate communication between developers
- summarize status for the larger community (guidance for use at different scales, regime; guidance for use of metrics)
- promote common understanding of error metrics

Investigate a withheld-data validation process and associated metrics

Short-term Recommendations (cont.)

Pursue the “full” error model

- recognize strong linkages to GPM work
- define “regime”
- characterize validation sites by regime
- what is the minimum set of coefficients, time/space correlations, and sensor types that is needed?
- which coefficients need to be posted for each grid box, and which can be provided as single values by regime, etc.?
- how to formulate the model such that calibration in various regimes can be “extended” outside the calibration locations to cover the globe?
- can combination data sets be treated as a sensor, or do we have to account for the inter-sensor time/space correlation structure?
- Work toward an assimilation-like approach for combinations (to overcome previous issue)
- perform early tests against other approaches
- sensitivity to changes in algorithms?