STAR PRECIPITATION AND WATER VAPOR VALIDATION SYSTEM

Version 1.0 – September 2020

Ralph Ferraro, Pierre Kirstetter, Mark Kulie, Patrick Meyers, Limin Zhao, John Forsythe, Tony Reale, Veljko Petkovic, Yalei You, Jun Dong, Malar Arulraj, Chris Grassotti, Sheldon Kusselson, Stan Kidder, Huan Meng, Yongzhen Fan, Bob Kuligowski

DOI: 10.25923/0hxp-wx63

1) Introduction and Objectives

NESDIS generates a suite of precipitation and water vapor (layered and total precipitable water, henceforth denoted as Blended Water Vapor (BWV) and precipitable water (PW)) products to support NWS operations. The routine validation of these products is critical to evaluate their performance over various regions, seasons and varying synoptic conditions. Such information is vital for product leads and end users. Over a decade ago, a system was developed, in conjunction with the CGMS/IPWG, to monitor 24-hr rainfall totals over the CONUS, using Stage IV radar and a CPC rain gauge analysis. This system has served us well, but it is no longer maintained and is grossly outdated. Some recent improvements to the system included adding in some preliminary swath validation for the GCOM AMSR2 rain rate product. Similarly, the NPROVS system has been a cornerstone for satellite temperature and moisture sounding validation, but has not been exploited routinely to validate the BWV/PW products.

As NESDIS moves to more enterprise solutions for precipitation and water vapor products, a more robust validation system is needed in order to make assessments of algorithm performance, to aid end users in their continued use of the products, and to form the basis for a system that somehow combines the best attributes of the different algorithms and products. This will then lead to a product suite that will have improved performance and will be easy to maintain in an operational environment.

It should be noted that this system would not be the only source of information used by the science teams to determine if product requirements are met. This will still be the responsibility of each science team, which may require a more detailed use of the validation data or use data sets that are not routinely available to this validation concept (i.e., field campaign data, data from external sources, etc.)

A “tiger team” of subject matter experts across NESDIS (STAR, OSPO, CI partners) was assembled to map out such a validation system. This short white paper will describe the requirements for such a system. It should be noted that the concept paper will try to leverage many of the existing tools in place for each of the algorithms that in many cases, use the same reference data sets for validation.

2) Product suites for consideration

The intent of this system is to consider all operational and emerging NESDIS product systems. It may also include other mature algorithms that are generated by other NESDIS partner agencies, such as NASA, JMA, JAXA and EUMETSAT. At a minimum, products listed in Table 1 will be considered part of the NESDIS precipitation and water vapor “portfolio.”

System or Algorithm Name	Parameter(s)	Focal Points	Product Class	Comments and Product Value	User Focal Point(s)
Microwave Integrated Retrieval System (MiRS)	Rainfall Rate Layered and total precipitable water	C. Grassotti Q. Liu L. Zhao	Operational	Will be extended to J-2, etc. and upcoming EPS-SG/MWS sensor data. Potential science improvements for TPW and precipitation being developed and tested.	CPC: P. Xie, R. Joyce; CIRA: J. Forsythe; CIMSS: T. Wimmers
Snowfall Rate (SFR)	Snowfall Rate	H. Meng J. Dong Y. You	Operational; Moderate Assurance	Ongoing effort to 1) expand to over ocean, 2) improve coverage in cold conditions, 3) enhance RTM and 1DVAR, 4) extension to new missions, J2, Metop-SG MWS and MWI	CPC: P. Xie, R. Joyce; NWS: A. Jacobs, K. White; NASA/SPoRT: E. Berndt
GAASP-GCOM-W AMSR2 Algorithm Software Processor (GAASP)	Rainfall Rate	V. Petkovic M. Arulraj L. Zhao	Operational	Global coverage; Hurricane monitoring; Monitoring over complex terrain Running GPROF2010V3; evaluating GPROF2017	CPC: P. Xie, R. Joyce
Global Hydro-Estimator (GHE)	Rainfall Rate	B. Kuligowski L. Zhao	Operational	Legacy, low latency rainfall rates	WPC: A. Orrison; NHC: M. DeMaria
Enterprise Rainfall Rate	Rainfall Rate	B. Kuligowski L. Zhao	Operational	Will replace GHE for GOES-16, GOES-17, Himiwari and METEOSAT.	WPC: A. Orrison; NHC: M. DeMaria
Blended Rainrate and TPW (bRR/bTPW)	Rainfall Rate TPW	S. Kidder J. Forsythe L. Zhao	Operational	https://www.ospo.noaa.gov/Products/bRR/Product_Animation.html seeing rain beyond radar coverage. https://www.ospo.noaa.gov/Products/bTPW/Product_Animation.html seeing atmospheric rivers of moisture.	WPC: A. Orrison; NHC: M. DeMaria NWS: A. Jacobs; K. Kasper
Layered PW	PW in four layers	J. Forsythe L. Zhao	Planned to Transition to Operation beginning in 2021	http://cat.cira.colostate.edu/sport/layered/advected/LPW_alt.htm seeing atmospheric rivers of moisture at four distinct layers. Heavy precipitation.	WPC: J. Nelson, M. Klein, A. Orrison
Ensemble Tropical Rainfall Prediction (eTRaP)	6-hr, 24-r rainfall forecasts	B. Kuligowski L. Zhao	Operational	http://www.ssd.noaa.gov/PS/TROP/etrap.html	WPC: A. Orrison; NHC: M. DeMaria
Cloud optical properties	Rain/No-Rain; Rain rate	A. Heidinger	Operational; Developmental	Rain/no-rain detection for shallow/warm top systems	
Non-NESDIS products for consideration include CMORPH2, IMERG, GSMaP					

Table 1: Products/algorithms that are candidates for routine validation and their associated attributes.

3) Validation Attributes

To help the validation design and selection of attributes it is useful to know the goals of each precipitation product (e.g. water resource management, extreme events, nowcasting vs. forecasting, input for gridded products, etc.). The value of each product, from the developers' perspective, is shown in Table 1.

The primary questions that drive the design of the validation system should come from both the algorithm developers and end users. There are many topics that need to be considered, including the monitoring of product performance and any changes over time with algorithm updates; routine statistics to see performance on various scales of weather events; are the products meeting expected requirements?; are there circumstances where the algorithm may fail? (e.g., region, synoptic conditions, time of the year, etc.). As an example, Figure 1 shows conditional first-order (systematic bias) and second-order (random error) discrepancies, which can be used to monitor a product performance or determine if product requirements are met.

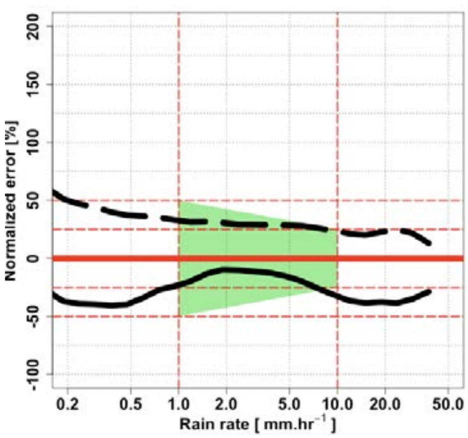


Figure 1: Evaluation of the GPM-Ku precipitation product. The green area indicates the Level 1 requirements in terms of errors. The solid line indicates the systematic error and the dashed line indicates the random error.

Some commonly used parameters used by the precipitation community include (adapted from https://www.cawcr.gov.au/projects/verification):

- **Discrimination:** ability of the estimate to discriminate among reference, that is, to have a higher prediction frequency for a reference value whenever this reference occurs. Examples: maps of hits/misses/false alarms contingency statistics by occurrence and volume; Probability Distribution Function (PDF) comparison by occurrence and volume;

- **Reliability or Bias:** the average agreement between the precipitation estimate and the reference (or correspondence between the mean estimate and mean reference). Example: Mean Relative Error (MRE);

- **Conditional bias:** if the reference or estimate are stratified into different ranges or categories;

- **Accuracy:** the level of agreement between the estimate and the reference, quantified with the individual differences between the estimate and the reference (i.e. error; the lower the errors, the greater the accuracy). Examples: scatterplots, Root Mean Square Error (RMSE), Mean Absolute Error (MAE);

- **Association:** the strength of the linear relationship between the products estimate and the reference. Example: correlation coefficient;

- **Skill:** the relative accuracy of the estimate over some benchmark. The benchmark is generally an unskilled estimate such as random chance, persistence (defined as the most recent set of observations, "persistence" implies no change in condition), or climatology. Skill refers to the increase in accuracy due purely to the "smarts" of the precipitation product. Examples: Heidke Skill Score (HSS), Nash;

- **Sharpness:** the tendency of the product to predict extreme values. Sharpness is a property of the product only, and it can have this attribute even if it's wrong (in this case it would have poor reliability).

The most useful validation metrics and graphics to (1) give a description of the water vapor and precipitation estimates performances and (2) meet the needs of different users and developers were identified through an online survey and inputs from the user and developer communities. The survey focused on different categories of attributes that diagnose (1) the occurrence of events (e.g. precipitation occurrence or extreme water vapor occurrence) which requires the use of categorical statistics, and (2) quantitative differences (how the values of the estimates differ from the values of the references). Questions were generic and applied to both water vapor and precipitation (two quantities with different physical properties, that are observed by different instruments, and retrieved with different techniques). Note that (1) precipitation and water vapor estimates are generally assumed to be deterministic estimates, and (2) the survey focused on matched-point statistics which are suitable to assess precipitation and water vapor gridded products. These a priori choices may be revisited down the road as needs arise and the system evolves. The survey was targeted to an expert audience from research laboratories and from the National Weather Service. Answers allowed to design the scope of the Validation System. An analysis of the answers indicates that a majority of the 21 responders work with precipitation products (76%), tend to be developers (62%), and work with deterministic products (90%). Specific outcomes of the survey are:

- Users' responses generally match developers' responses;

- While both categorical and quantitative statistics are identified to be useful for the assessment of precipitation product performances, it appears that quantitative statistics are more relevant than categorical statistics to the assessment of water vapor products (see table below). It certainly reflects the different properties of water vapor and precipitation fields, especially the high intermittency of the later;

- Matched-point statistics are relevant, and it is generally desired/recommended to complement them by statistics based on specific events (e.g. a landfalling hurricane, weather regimes, etc.) at a later time;

- In the situation where a decision-maker wants to receive only a summary score, the preferred option is to discuss for the use of a few key statistics instead of only one (67% of the responses), followed by combining several statistics into one diagram (29%). Diagrams may be perceived as complicated for non-expert users; one responder suggests to offer a set of statistics accompanied with a diagram;

- Preferred visuals for comparing precipitation or water vapor estimates w.r.t. a given reference are maps (95% of responders chose this option), followed by histograms (occurrence and volume; 90%), (density) scatterplots (76%), times series (67%), box plots (57%), and diagrams (52%);

- On conditional statistics: global metrics can be considered as a starting point. As the system develops they can be broken down according to seasonal, geographic, meteorological settings. According to a responder, this will be relevant to NWS users who are concerned about “how the products work upstream from them and do not care if the global / large scale statistics are good if it doesn't work for them”. The impact of data source to the final retrievals in the blended products should be considered. Specifically regarding water vapor, quantitative metrics should be calculated separately over land and water;

- Note: Interestingly some responders are aware of the limitations associated with bulk statistics; these alone do not necessarily describe the usefulness of a product to support NOAA operations.

Table 2 indicates the metrics evaluated in the survey and their rate of approval. This rate is computed as the cumulated rates (in percent) of the responses to “important” and “very important”.

4) Validation data sources

This section identifies viable data sources for the routine generation statistics within the system. It should be noted that specialized data sets (e.g., field campaigns or data sets with considerable latency) are not considered.

a) Precipitation (Rain and Snow)

- *Multi-Radar/Multi-Sensor (MRMS)*[1] – The GPM Ground Validation Multi-Radar/Multi-Sensor (GV-MRMS) Precipitation Reanalysis for Satellite Validation Product dataset contains precipitation rate and type estimates, quality control products, and precipitation corrective factors products. These data products were created using the NOAA MRMS System which ingests

[1] See https://wallops-prf.gsfc.nasa.gov/NMQ/index.html and https://ghrc.nsstc.nasa.gov/hydro/#/details?ds=gpmmrms

Assessment	Metric	Rate of approval Overall | W.V. | Prec.	Associated question
Event detection	Contingency (occurrence & volume)	85% | 40% | 100%	How do the estimates identify the occurrence of specific events?
Event detection	Accuracy	90% | 60% | 100%	What fraction of the estimates is correct in identifying the occurrence of an event?
Event detection	Bias	85% | 60% | 93%	How does the estimated freq. of event occurrence compare to the reference freq. of event occurrence?
Event detection	Probability of Detection (POD)	90% | 60% | 100%	What fraction of the event occurrences in the reference is correctly estimated?
Event detection	False Alarm Rate (FAR)	85% | 40% | 100%	What fraction of the estimated event occurrences actually did not occur?
Event detection	POFD (Probability of False Detection)	90% | 60% | 100%	What fraction of the reference non-occurrences are incorrectly estimated as occurring?
Event detection	Threat score / Critical Success Index (CSI)	90% | 60% | 100%	How well the estimated events correspond to reference events?
Event detection	Equitable Threat Score (ETS)	84% | 60% | 93%	How well the estimated event occurrences correspond to reference event occurrences, accounting for hits due to chance?
Event detection	Hanssen and Kuipers (HK) discriminant	95% | 80% |100%	How well the estimates separate the event occurrences from the non-occurrences?
Event detection	Hiedke Skill Score (HSS)	74% | 60% | 79%	What is the accuracy of the estimates relative to that of random chance?
Event detection	Odd ratio skill score	63% | 40% | 71%	What is the improvement of the estimates over random chance?
Quantification	Mean Relative Error (MRE)	90% | 100% | 87%	What is the average precipitation or water vapor estimate error w.r.t the reference?
Quantification	MRE	90% | 80% | 93%	What is the average sign of errors in estimates w.r.t the reference?
Quantification	Mean Absolute Error (MAE) or Root Mean Square Error (RMSE)	95% | 100% | 93%	What is the average magnitude of errors w.r.t the reference?
Quantification	Linear Error in Probability Space (LEPS)	90% | 100% | 87%	What is the error expressed in terms of probabilities?
Quantification	Correlation	100% | 100% | 100%	How well do the estimated values correspond to the reference values?
Quantification	Anomaly correlation	85% | 100% | 80%	How well do the estimated anomalies correspond to the reference anomalies?
Quantification	S1 score	70% | 100% | 60%	How well do the estimated gradients correspond to the reference gradients?

Table 2: Summary of survey of validation system requirements from both users and developers.

Weather Surveillance Radar 88 Doppler (WSR-88D) radar data, Rapid Update Cycle (RAP) model analysis fields, and gauge data. It should be noted that these data products are not standard MRMS products. Significant postprocessing is applied to the operational MRMS output to generate products specifically adapted to satellite purposes and needs over North America. These data are available from March 2, 2014 through October 30, 2018 in ASCII format in the NASA Global Hydrological Resource Center archive. The conservative post-processing is designed to maximize accuracy, minimize uncertainties and standardize the precipitation reference products across the CONUS. Thanks to their higher spatial resolution than any satellite precipitation product (0.01deg), these data are designed to be pixel matched in both time and space, and to build statistics for comparing reference precipitation intensities to satellite-based estimates.

Regarding snow, two potential drawbacks are associated with GV-MRMS (or any other dataset that utilizes scanning radars): (1) Quantitative Precipitation Estimation (QPE) currently relies on a static reflectivity (Z) to liquid equivalent snowfall rate (S) conversion (Z=75S^2.0) that might not appropriately represent different snowfall regimes and environmental conditions; (2) QPE accuracy and detection capabilities naturally degrade when using radar observations at longer distances from the radar site due to both beam filling effects and increasing radar beam height relative to the ground. This second challenge is an especially important consideration for shallow snowfall events like orographic and lake-effect snow. Radar beam blockage in mountainous regions also complicates snow QPE and leaves large data voids. The first deficiency can be partially mitigated through applied research (see Specialized ground-based testbeds sub-section below) and incorporating NEXRAD dual-pol fields in the near future should improve QPE. The second deficiency can be mitigated by adopting a range threshold from each respective radar site for satellite evaluation purposes. Two other complications arise when applying GV-MRMS for snowfall-related QPE: (1) standard MRMS QPE are not gauge-adjusted for likely surface snowfall events and (2) accurate surface precipitation phase identification. GV-MRMS utilizes operational NWP fields to discriminate between solid and liquid surface precipitation, but further applied research can be undertaken to verify the veracity of precipitation phase detection capabilities under complex conditions (e.g., surface temperature in the 0-4 C range that can be rain, snow, freezing rain, or mixed precipitation). However, simple screening methods could be applied to only consider extremely confident snowfall classifications when evaluating satellite datasets.

- *Stage IV*[2] - The Stage IV product is a regional radar mosaic product adjusted for gauge readings. It provides hourly precipitation analysis at 4 km resolution over the contiguous U.S. Its multi-step production is initiated at National Weather Service (NWS) River Forecast Centers (RFCs) shortly after the end of the accumulation hour. Combination of automated processes and manual quality control leads to a high-quality product on the national grid. Data typically become available within 30-40 min past the end of the hour of accumulation. These near real-time precipitation analyses

[2] Refer to Smalley et al. [2014] and Prat and Nelson [2015]

are complemented with any originally missing RFC reports to form the final mosaics 12–18 h after accumulation time on a Hydrologic Rainfall Analysis Project (HRAP) grid and are archived at the National Center for Atmospheric Research (NCAR). Despite its uncertainties, the final product's availability and latency have led to a wide range of Stage IV dataset applications, including validation and calibration of other observational products such as satellite merged data and microwave retrieval development (Meng et al. 2017).

- *Community Collaborative Rain, Hail & Snow Network* (CoCoRaHS)[3] – CoCoRaHS is a community-based network of volunteer weather observers who report 24-hr precipitation accumulations measured at as many locations as possible. These rather dense measurements are a potentially useful independent data stream that provide spatial variability statistics to augment automated gauge observations. While the measurement consistency and veracity pose potential issues, these observations can be found useful for regional applications. Daily files are readily accessible with some known daily latency challenges, which are being addressed through continuous improvements by the CoCoRaHS team. These data might be best used for daily post-hoc evaluations, especially for significant weather events.

- *Specialized ground-based testbeds for snowfall* - Point-source ground-based observations are often difficult to use for satellite evaluation activities due to spatiotemporal sampling discrepancies, but a few useful datasets can be leveraged for snowfall evaluation purposes. For instance, a snowfall observatory (profiling radar and in situ snow microphysics observations) at the Marquette, MI National Weather Service has operated since January 2014 (Pettersen et al. 2020; Kulie et al. 2020). High quality one-minute snowfall rate and snow-to-liquid ratio observations have been measured. These observations have both been used to improve empirical NEXRAD QPE by honing regime-dependent Z-S conversions and provide a robust multi-year snowfall rate dataset that can be statistically compared to satellite-based estimates (e.g., comparing Upper Great Lakes snowfall detection statistics and snowfall rate distributions). Other high quality datasets should be identified and leveraged, such as the NCAR snowfall observatory in the foothills of Colorado and the Environment and Climate Change Canada datasets collected at numerous sites (e.g., near Toronto and relatively new Arctic observatory).

  Other possible datasets to leverage include the Snow Telemetry (SNOTEL) dataset collected in mountainous Western CONUS and Alaska regions. SNOTEL is administered by the US Department of Agriculture's Natural Resources Conservation Service and National Water and Climate Center. SNOTEL uses a combination of automated and manual measurements to create snowfall accumulation datasets on various time scales and is probably more appropriate for monthly or seasonal accumulation comparisons.

- *Global data sets - Integrated Surface Database (ISD)[4]* - The Integrated Surface Database (ISD) consists of global hourly and synoptic observations compiled from numerous sources into a

[3] See https://www.cocorahs.org/Login.aspx
[4] https://www.ncdc.noaa.gov/isd

single common ASCII format and common data model. The database includes over 35,000 stations worldwide, with some having data as far back as 1901, though the data show a substantial increase in volume in the 1940s and again in the early 1970s. Currently, there are over 14,000 "active" stations updated daily in the database. This dataset includes the precipitation occurrence record over the vast majority of the world. The precipitation intensity is only available over the western Europe, the United States and Japan with varying quality control procedures.

- *Validation over open ocean* - Satellite product validation over open ocean is even more challenging largely due to the data availability issue. There are three primary data sources which may be useful for satellite product validation:

  - *International Comprehensive Ocean-Atmosphere Data Set (ICOADS)*[5]- This dataset only includes the precipitation phase variable (i.e., precipitation or non-precipitation), which has been used extensively to train the snow/rain separation scheme over ocean;

  - *The Ocean Rain and Ice-Phase Precipitation Measurement Network (OceanRain*[6]*)* - OceanRAIN is an ODM470 optical disdrometer based in-situ shipboard precipitation, evaporation and freshwater flux data set over the global oceans. For the period between June 2010 and April 2017 OceanRAIN-1.0 offers more than 6.83 million minutes of measurements from 8 research vessels and provides high-quality validation and calibration data.

  - *GPM Radar* - Collocated GPM Ku-Ka precipitation radar measurements are regularly used for development and validation of passive microwave precipitation retrievals, applied to polar orbiting satellites, including the GPM constellation (e.g., S-NPP, GCOM-W1, MetOpA, MetOpB, NOAA-18, -19 and -20.

b) TPW and LPW

There are several viable data sets and existing systems that can be exploited to validate the TPW and LPW products. These are described below and also summarized in Table 3.

- *NOAA Products Validation System (NPROVS)*[7] - The NPROVS is well-positioned to perform validation of TPW and LPW. Radiosondes provide a source of vertical profiles, although biased towards land where passive microwave retrievals are less accurate than over ocean. TPW and LPW must be derived from the radiosonde data, and there are small errors due to how this integration is performed. Surface-based GNSS retrievals are useful for TPW only. GPS is a component of GNSS which includes navigational satellite systems operated by

[5] https://icoads.noaa.gov/
[6] https://oceanrain.cen.uni-hamburg.de/
[7] https://www.star.nesdis.noaa.gov/smcd/opdb/nprovs/

other nations (e.g. European Galileo, Russian GLONASS). NOAA ingests Radio occultation profiles from GNSS (e.g. COSMIC-2), but due to low spatial resolution and the use of forecast model constraints it is not recommended as a validation source for water vapor at this time.

- *CIRA Monitoring System*[8] - CIRA currently operates near-real time validation of the TPW (hourly via GPS) and LPW (twice daily via radiosondes). Hourly comparisons to GPS of the operational blended TPW, advected microwave only TPW, GOES-16 TPW, and GFS TPW are computed. Portions of this processing may be used along with NPROVS if desired.

- *Other Sources* - Other nations have their own surface GNSS systems, often denser in earthquake-prone regions (Japan has an extremely dense network of about 800 stations). We encourage NOAA to continue to pursue ingest of these other networks, as even over CONUS the network distribution follows agency and state boundaries leaving large gaps with no data.

  NWP Models also offer independent, spatially continuous data fields with global coverage that can provide somewhat independent validation for the satellite products on a routine basis. Performance stratified by surface type can easily be determined. They are not truly independent since satellite radiances are used in the data assimilation, and in some instances, are the same data that are used to derive the TPW from satellites directly. However, these are done much differently, so the final TPW values can be considered different.

Data Source	**Pros**	**Cons**
GPS-MET (Figure 2)	·All weather, hourly ·Established in many papers as a validation standard for TPW ·Available in NRT via MADIS (CIRA gets this now, requires government access)	· Clumped stations · Biased to land, CONUS ·Topography effect for TPW must be considered in mountains (think California, Central Valley vs Sierras)
Radiosondes (Figures 3 & 4)	· The best tool for profile validation · Somewhat global coverage · Could focus on GRUAN, but need longer time periods	· Some errors via country due to different models, time of day effects (sensor heating) · Typically 0/12 UTC

Table 3: Pros and Cons of Validation Data for TPW and LPW

[8] http://cat.cira.colostate.edu/GPS_TPW_Stats

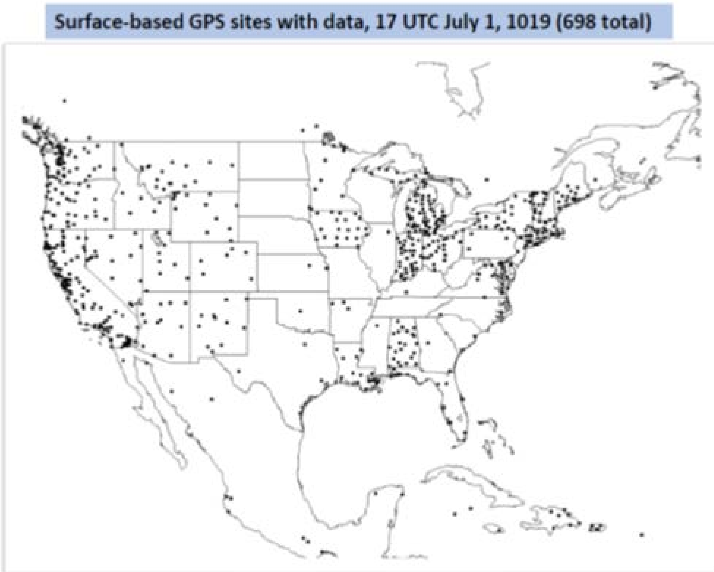


Figure 2: Surface GPS sites with TPW data, 17 UTC, 1 July 2019.

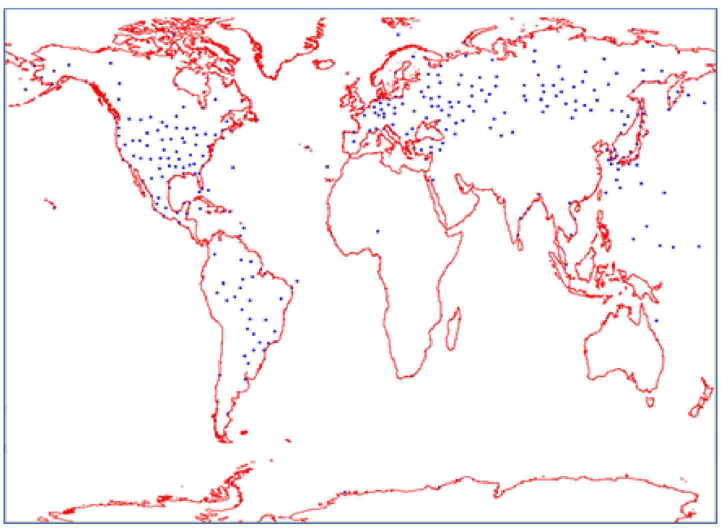


Figure 3: MADIS radiosonde sites, 12 UTC 19 May 2020.

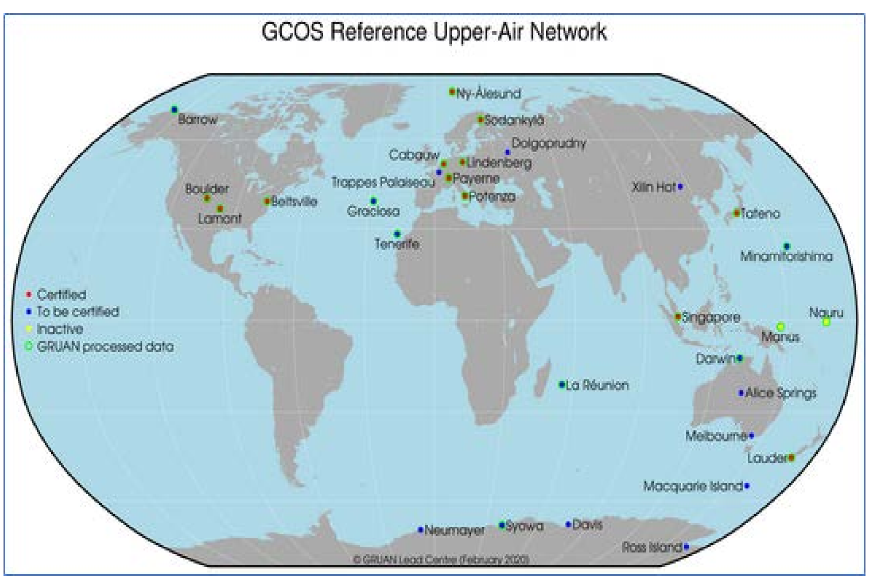


Figure 4: GRUAN radiosonde sites, as of 19 May 2020 from gruan.org (26 sites, 12 certified).

5) "Operational" vision

Based on the system requirements and also considering existing validation infrastructure, this section identifies the most "practical" approach to developing and operating a more robust validation system at NESDIS STAR. The term "operational" refers to routine, reliable validation delivery, most likely run as a moderate assurance system at STAR or at its CI partners. It is well understood that the resources required to make this a 24 x 7 fully operational system will be difficult to obtain and are not necessarily a requirement to bring this to fruition.

a) Precipitation validation System

For precipitation, the initial plan is to enhance an existing validation system. We plan to start with precipitation validation infrastructure at CISESS and that is partially supported by STAR JPSS. Currently in place to support the GAASP and bRR products (see Figure 5), the system can be extended to the other products listed in Table 1.

Initial development of the Precipitation Cal/Val system will rely on CISESS computing resources, which will leverage support for validation and monitoring the AMSR2 and bRR precipitation products. This effort will be led by Dr. Malar Arulraj, with the input of the Cal/Val team and product leads. The proposed validation system will be developed for AMSR2 (Level-2) and bRR (Level-3), and then vetted by the Cal/Val team to approve the methodology and display.

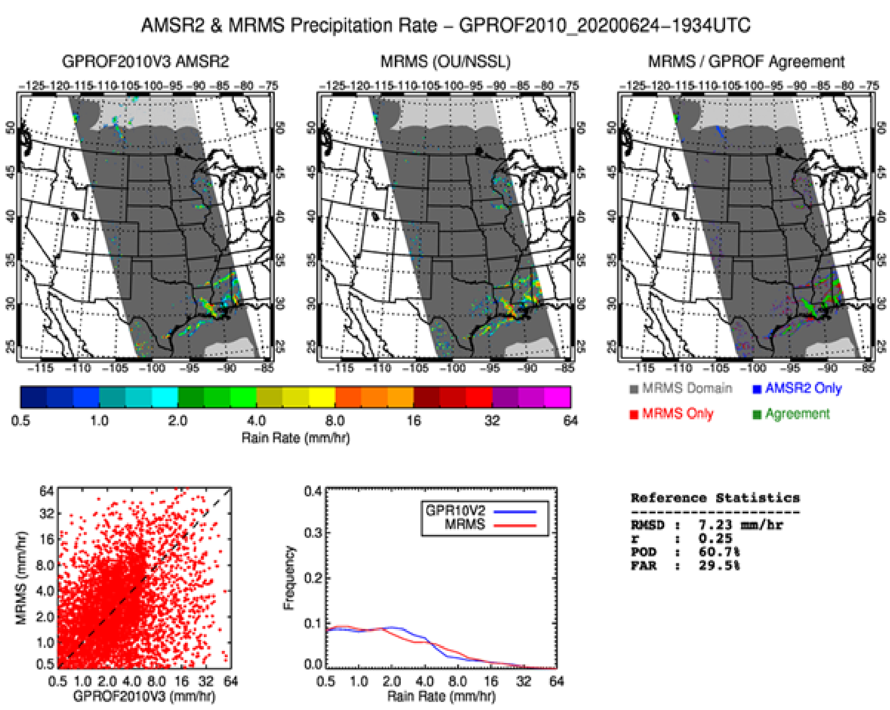


Figure 5: Example of AMSR2 validation with MRMS for 24 June 2020, 1934 UTC.

The precipitation validation will primarily rely on the radar-based MRMS dataset over CONUS described in Section 3 (GV-MRMS; Kirstetter et al. 2012; Carr et al. 2015). Gauge-adjusted instantaneous rain rates will be averaged to the nominal resolution of the field of view (FOV; Level-2) of each satellite or grid spacing (Level-3) for each product (Gebregiorgis et al. 2018). Despite differences in resolution between products, the validation should be presented at the original resolution so that forecasters and product developers can assess the accuracy of the product in its distributed form (Kirstetter et al. 2012).

For the Level-2 products, side-by-side comparisons of the satellite and radar observations, along with statistical metrics of point-by-point comparisons, will be generated each time a satellite estimate is available over the CONUS region. Plots will be generated and made available to the users within 1-hour of satellite products release. Similar comparisons will be generated for Level-3 products in 3hourly increments. The matchup data will be saved and available for download, including relevant environmental variables (i.e. surface temperature, surface type, environmental TPW) that could help in conditional evaluation of the product. Saving the data with its original geolocation information will facilitate comparisons by product developers in support to algorithm updates.

The plotting and statistical calculations will be consistent across Level-2 and Level-3 products, and available for online viewing. To facilitate bulk statistical comparisons across product lines, seasonal validation metrics (i.e. RMSE, bias, POD, FAR) and mapping will be performed on a universal grid. Cross-product comparisons will show the overall and conditional performance of the datasets.

The Cal/Val team will need assistance to get access to NOAA's cloud environment and to create a web interface. Approximately one year of collocated data will be saved and collectively require ~1 TB. Each product will require a unique reader and description of the sensor FOVs or product grid spacing. Python scripts for generating the comparisons will also be available for review and use by end-users.

This system is expected to be the primary validation system used to evaluate bRR and AMSR2, with the capability to be extended for other precipitation products from systems such as MiRS, GHE, and SCaMPR, Funding permitting, the system will be expended to support the snowfall rate products and offer annual status reporting.

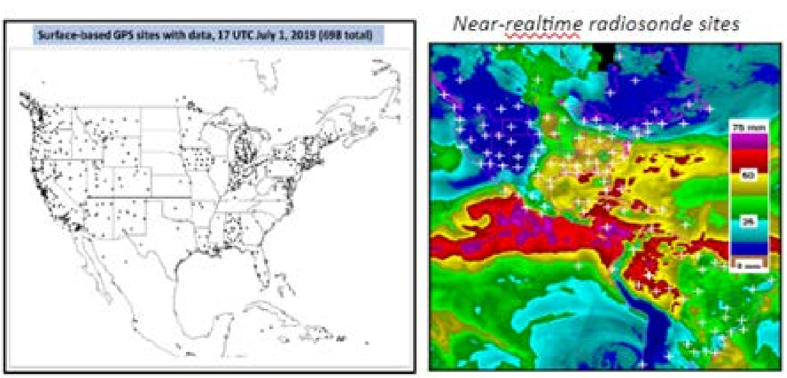
b) WV Validation System

The proposed WV Cal/Val system will be for the MiRS, blended TPW and Advected Layer Precipitable Water. It would leverage the Cal/Val capability at CIRA to evaluate blended Total Precipitable Water (TPW) and Advected Layer Precipitable Water (APLW), and Radiosonde data source and satellite collocation capability at the NOAA Products Validation System (NPROVS).

The WV products validation will primarily rely on the GPS over CONUS (see Figure 6), and radiosonde over global domains (see Figure 7). The ground-based observations will be collocated to the nominal resolution of the field of view (FOV; Level-2) or grid spacing (Level-3) for each product. Despite differences in resolution between products, the validation should be presented at the original resolution so that forecasters and product developers can assess the accuracy of the product in its distributed form.

For Level-2 products, each CONUS overpass will generate a figure with side-by-side comparisons of the satellite and GPS/Radiosonde observations, along with statistical metrics of point-by-point comparisons. Plots will be generated within 1-hour of receipt. Similar plots will be generated for Level-3 products, which will be generated every 3 hours. The matchup data will also be saved and available for download, including data (i.e. surface temperature, surface type, precipitation condition) that could help in conditional evaluation of the product. Saving the data with its original geolocation information will facilitate comparisons by product developers that are testing algorithm updates.

The plotting and statistical calculations will be consistent for Level-2 and Level-3 products, and be available for online viewing. To facilitate bulk statistical comparisons across product lines, monthly and annual validation metrics (i.e. RMSE, bias) and monitoring figures will be calculated on a universal grid. Cross-product comparisons will show the overall and conditional performance of the datasets.



	SNPP	NOAA-20
Number of Matches	2293	6690
r^2	0.84	0.86
RMS (mm)	3.9	4.2
Bias (mm)	1.5	2.3

Figure 6: Example of comparison plot that would be routinely generated to compare blended TPW retrievals to GPS.

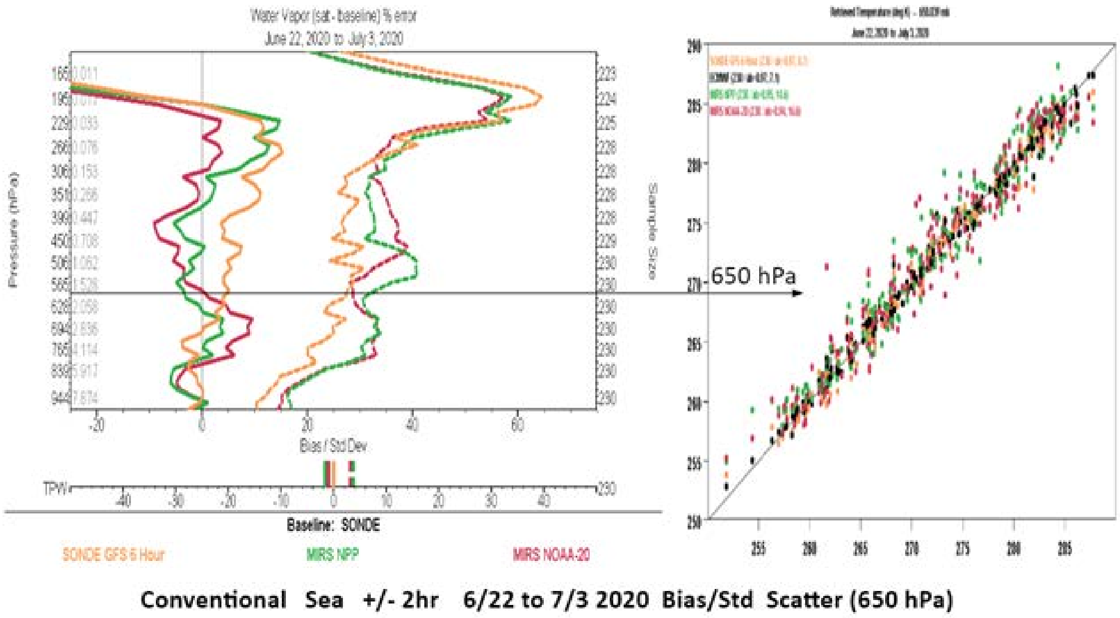


Figure 7 - Example of Vertical Statistics of SAT-Raob Difference that would be routinely generated to compare WV products with Radiosonde.

6) Next Steps (0-2 years)

The general plan is to present the findings of this white paper to the NESDIS/STAR Quarterly Program Reviews (QPR) as part of the Precipitation and Hydrology Team briefings. A companion briefing will also be given to the STAR science board. It is hoped then that the resources needed to move this activity into a more mature stage could be solicited by the NESDIS bi-annual planning meetings, where program needs are prioritized, and funding can be acquired. In the meantime, we can move a few topics forward, based on some resources that are already available through the JPSS program and part of our routine product quality assurance. As such, some next steps in the short term (0 to 3 years) can include:

a) Precipitation

- Optimize existing AMSR-2 and bRR validation code and transition from IDL to Python
- Perform sensitivity studies to determine optimal precipitation thresholds, spatial resolutions and land surface stratification
- Incorporate more baseline products (Table 1) such as MiRS, SCaMPR, SFR and CMORPH into the validation scheme
- Develop database of archived co-locations for further analysis
- Determine key statistics to generate in the first version of the system (based on Table 2)
- Develop web based interface to display statistics over a variety of geographic, temporal, and surface type domains. This should include time series plots of the statistics.

b) Water Vapor

Year 1:

- Validate MiRS swath TPW and LPW against radiosonde TPW / LPW within NPROVS, and provide daily and monthly updated statistics and plots on web.
- Validate blended TPW product against GPS, radiosondes. Use OCO-2 for clear sky ocean cases.

Year 2:

- Continue Year 1 validation and validate MiRS swath TPW / LPW against GRUAN sites, and also explore the feasibility to append GCOM/GAASP and operational bTPW.
- Continue Year 1 validation, and validate Blended TPW and ALPW against sondes / GPS, calculate statistics over a variety of geographic, temporal, and surface type domains.

Years 1 and 2:

- Develop and enhance web-based interface to display statistics over a variety of geographic, temporal, and surface type domains when resource available
- In-depth investigation of any anomalies reported by users or detected by team. These might involve rain or surface contamination, or sensor resolution impacts.

References

Carr, N., P.E. Kirstetter, Y. Hong, J.J. Gourley, M. Schwaller, W. Petersen, N.Y. Wang, R. Ferraro, X. Xue, 2015: The influence of surface and precipitation characteristics on TRMM TMI rainfall retrieval uncertainty. Journal of Hydrometeorology, 16, 1596-1614. 10.1175/JHM-D-14-0194.1

Gebregiorgis, A., P.E. Kirstetter, Y. Hong, J.J. Gourley, G. Huffman, W. Petersen, X. Xue, M. Schwaller, 2018: To what extent is the Day-1 IMERG satellite rainfall estimate improved as compared to TMPA-RT? Journal of Geophysical Research, 123, 1694-1707. doi: 10.1002/2017JD027606

Kirstetter, P.E., Y. Hong, J.J. Gourley, S. Chen, Z. Flamig, J. Zhang, M. Schwaller, W. Petersen and E. Amitai, 2012: Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar-based National Mosaic QPE. Journal of Hydrometeorology, 13(4), 1285-1300. doi:10.1175/JHM-D-11-0139.1

Kulie, M.S., C. Pettersen, A. Merrelli, T. Wagner, N. Wood, M. Dutter, D. Beachler, T. Kluber, R. Turner, M. Mateling, J. Lenters, P. Blanken, M. Maahn, C. Spence, S. Kneifel, P. Kucera, A. Tokay, L. Bliven, D. Wolff, W. Petersen, 2020: Upper Great Lakes snowfall: Lessons learned from a multi-sensor snowfall observatory. *Bull. Amer. Meteor. Soc*, In preparation.

Meng, H, J. Dong, R. Ferraro, B. Yan, L. Zhao, C. Kongoli, N-Y. Wang and B. Zavodsky, 2017: A 1DVAR-based snowfall rate retrieval algorithm for passive microwave radiometers. J. Geophys. Res. Atmos, 122, doi: http://dx.doi.org/10.1002/2016JD026325.

Pettersen, C., M.S. Kulie, L.F. Bliven, A.J. Merrelli, W.A. Petersen, T.J. Wagner, D.B. Wolff, and N.B. Wood, 2020: A Composite Analysis of Snowfall Modes from Four Winter Seasons in Marquette, Michigan. *J. Appl. Meteor. Climatol.,* **59**, 103–124, https://doi.org/10.1175/JAMC-D-19-0099.1

Acronyms

AMSR2 – Advanced Microwave Scanning Radiometer 2
BWV – Blended Water Vapor
CGMS – Coordination Group for Meteorological Satellites
CIRA – Cooperative Institute for Research of the Atmosphere
CISESS – Cooperative Institute for Satellite Earth System Studies
CMORPH2 – Climate Prediction Center Morphing Technique 2
CoCoRaHS – Community Collaborative Rain, Hail and Snow Network
CONUS – Continental United States
CPC – Climate Prediction Center
eTRaP – Ensemble Tropical Rainfall Potential
EUMETSAT – European Organization for the Exploitation of Meteorological Satellites
GAASP – GCOM-W AMSR2 Software Processor
GCOM-W – Global Change Observation Mission - Water
GHE – Global Hydro-Estimator
GNSS – Global Navigation Satellite System
GPM – Global Precipitation Measurement Mission
GPS – Global Positioning System
GSMaP – Global Satellite Mapping of Precipitation
GV – Ground Validation
GV- MRMS – Ground Validation Multi-Radar/Mult-Sensor rainfall system
HRAP – Hydrologic Rainfall Analysis Project
ICOADS – International Comprehensive Ocean-Atmosphere Data Set
IMERG – Integrated Multi-satellite Retrievals for GPM
IPWG – CGMS International Precipitation Working Group
ISD – Integrated Surface Database
JAXA – Japanese Aerospace Exploration Agency
JMA – Japanese Meteorological Agency
MiRS – Microwave Integrated Retrieval System
NASA - National Aeronautics and Space Administration
NPROVS – NOAA Products Validation System
NWS – National Weather Service
PW – Precipitable Water
QPE – Quantitative Precipitation Estimation
RAP – Rapid Update Cycle Model
RFC – River Forecast Center
SFR – Snowfall Rate
SNOWTEL – Snow Telemetry