AMSR-E/Aqua Data Quality and Data Uncertainty

TABLE OF CONTENTS

1 DATA QUALITY

1.1 Quality Assessment Files

For all AMSR-E products, each HDF-EOS file is accompanied by a Quality Assessment (QA) file. The QA files can be accessed through HTTPS or Earthdata Search. The contents of each QA file is unique to the product. The quality assessment is performed during standard product generation at the Global Hydrology and Climate Center (GHCC) - Science Investigator-led Processing System (SIPS). Should a product fail the assessment, the file is quarantined and the Team Leader Science Computing Facility is notified. After corrective action is taken, the file is reprocessed and forwarded to NSIDC for archive and distribution.

1.2 Quality flags

The quality flags found in individual products are detailed in each product's guide document.

2 DATA UNCERTAINTY

The following discussions outline the uncertainty assessment of AMSR-E data products available from NSIDC, by providing:

- A synopsis on sources of uncertainty
- Best estimates of data uncertainty under optimal conditions for each measurement, including a confidence interval where possible
- A description of how to interpret quality flags to understand the conditions under which uncertainty may be greater

2.1 Brightness Temperatures

Applicable data sets
AMSR-E/Aqua L2A Global Swath Spatially-Resampled Brightness Temperatures (AE_L2A)
Applicable version(s): Version 3
Algorithm Name: Remote Sensing Systems (RSS) Version 7 Microwave Brightness Temperature Calibration
Algorithm Author: Frank J. Wentz

Synopsis on sources of uncertainty.

One of the main sources of brightness temperature error is radiometer noise. We have seen that overall brightness temperature uncertainty can be significantly reduced in the resampling process because individual random errors tend to cancel each other. Mis-specification of antenna ground patterns would contribute uncertainty to resampled channels. Other major sources of brightness temperature uncertainty include errors in the calculated effective hot load temperature, effective cold space temperature, emissivity of the main reflector, non-linear response functions, geolocation errors (especially earth incidence angle), center observation frequencies, Antenna Pattern Correction (APC) and spillover, residual errors in the radiative transfer model (RTM), and Radio Frequency Interference (RFI). Swath edge cold contamination, while corrected reasonably well over ocean, remains problematic over land and ice.

The best estimates of data uncertainty under optimal conditions, for each measurement, including a confidence interval where possible.

Over 95% of brightness temperature values over water are within .5K.

Description of how to interpret quality flags to understand the conditions under which uncertainty may be greater.

In addition to passing along JAXA calculated Data_Quality from the input L1A files, the L2A algorithm produces Scan_Quality_Flags, Channel_Quality_Flags, and Resampled_Channel_Quality_Flags. The usage of these flags is documented in the quality information section of the AE_L2A user guide.

2.2 Soil Moisture Products

Applicable data sets
AMSR-E/Aqua L2B Surface Soil Moisture, Ancillary Parms, & QC EASE-Grids (AE_Land)
AMSR-E/Aqua Daily L3 Surface Soil Moisture, Interpretive Parameters, & QC EASE-Grids (AE_Land3)
Applicable version(s): Version 2
Algorithm Name: AMSR-E Land Surface Product Algorithm
Algorithm Authors: Eni Njoku, Steven Chan

Synopsis on sources of uncertainty.

AMSR-E retrievals of soil moisture are directly sensitive only to the top 1 cm of soil averaged over approximately 60 km spatial extent. Significant uncertainty may therefore arise when these measurements are compared against point-derived in-situ data, due to differences in sampling depth and spatial extent between satellite and in-situ observations.

Measurements of soil moisture are most accurate in areas of low vegetation. Attenuation from vegetation increases the retrieval error in soil moisture. A surface type classification based on the AMSR-E data is assigned to indicate low and moderate vegetation, and retrievals are not performed in dense vegetation.

In low and moderate vegetation areas the retrieval algorithm performs a correction for vegetation and roughness based on an empirical formulation tuned to specific targets (consistent with externally available Normalized Difference Vegetation Index (NDVI)-derived vegetation data). After correction, a second soil moisture retrieval step is performed with coefficients adjusted to cover the soil moisture dynamic range anticipated for AMSR-E (at the 60-km scale). Global representativeness of these coefficients will have an impact on the bias and sensitivity of the soil moisture retrievals.

The retrieval algorithm does not explicitly model or correct for effects of urban areas, water bodies, topography, snow cover, clouds, and precipitation. Other potential error sources include anomalous inputs from bad radiometric data and low-level processing errors. The processing algorithm includes checks to identify these and other anomalies and assign appropriate flags.

Soil moisture retrievals represent averages over the horizontal footprint area and vertical sampling depth in the top ~1 cm of soil. The actual sampling depth varies with the amount of moisture in the soil. Soil moisture deeper than ~1 cm below the surface may not be sensed by AMSR-E.

The 6.9 GHz channel is shared with mobile communication services; therefore, retrievals using this frequency are subject to Radio Frequency Interference (RFI), particularly near large urban land areas. The soil moisture algorithm uses the 10.7 GHz channel to alleviate the RFI problem, however some RFI has also been observed at 10.7 GHz.

The best estimates of data uncertainty under optimal conditions, for each measurement, including a confidence interval where possible.

Several papers have performed comparative analyses of AMSR-E soil moisture data products. Among these are:

Crow, W. T., Diego G. Miralles, and M. H. Cosh. 2010. A Quasi-global Evaluation System for Satellite Surface Soil Moisture Retrievals. *IEEE Transac. Geosci. Remote. Sens.* 48(6): 2516-2527. doi:10.1109/TGRS.2010.2040481.

Draper, Clara S., Jeffrey P. Walker, Peter J. Steinle, R. A.M. de Jeu, and T. R.H. Holmes. 2009. An Evaluation of AMSR-E Derived Soil Moisture over Australia. *Remote Sensing of Environment* 113(4): 703-710. doi:10.1016/j.rse.2008.11.011.

Jackson, T. J., M. H. Cosh, R. Bindlish, P. J. Starks, D. Bosch, M. Seyfried, D. Goodrich, S. Moran, and D Du. 2010. Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products. *IEEE Transactions on Geoscience and Remote Sensing* 48(12): 4256-4272. doi: 10.1109/TGRS.2010.2051035.

Description of how to interpret quality flags to understand the conditions under which uncertainty may be greater.

2.3 Ocean Products

Applicable data sets
AMSR-E/Aqua L2B Global Swath Ocean Products derived from Wentz Algorithm (AE_Ocean)
AMSR-E/Aqua Daily L3 Global Ascending/Descending .25x.25 deg Ocean Grids (AE_DyOcn)
AMSR-E/Aqua Weekly L3 Global Ascending/Descending .25x.25 deg Ocean Grids (AE_WkOcn)
AMSR-E/Aqua Monthly L3 Global Ascending/Descending .25x.25 deg Ocean Grids (AE_MoOcn)
Applicable version(s): Version 2
Algorithm Name: Ocean Products derived from Wentz Algorithm
Algorithm Author: Frank J. Wentz

Synopsis on sources of uncertainty.

Some degree of uncertainty is inherent to the input data, namely resampled brightness temperatures, geolocation, and earth incidence angle. Under optimal conditions, additional uncertainty is mainly due to residual errors in the radiative transfer model (RTM) and its inverse, which is essentially the ocean retrieval algorithm. Certain environmental conditions can significantly increase errors and uncertainty, in many cases making some retrievals unfeasible. These environmental factors include land, sea ice, rain, high wind speeds, sun glint, and Radio Frequency Interference (RFI). Some of these factors are stable (land), or at least somewhat consistent (sea ice, rain, winds). Sun glint effects can vary with solar activity, especially solar flare events. RFI is a continuously evolving source of errors.

Very Low Resolution Sea Surface Temperature (SST) errors increase with proximity (~75km) to land, sea ice, rain, wind speeds above ~20 m/s, sun glint less than ~25°, 7 GHz RFI, 11 GHz RFI, and 18 GHz RFI.

Low Resolution SST errors increase with proximity (~50km) to land, sea ice, rain, wind speeds above ~20 m/s, sun glint less than ~25°, 11 GHz RFI, and 18 GHz RFI.

Low Resolution Wind Speed errors increase with proximity (~50km) to land, sea ice, rain, sun glint less than ~25°, 11 GHz RFI, and 18 GHz RFI.

Medium Resolution Wind Speed errors increase with proximity (~25km) to land, sea ice, rain, sun glint less than ~10°, and 18 GHz RFI.

Medium Resolution Water Vapor errors increase with proximity (~25km) to land, sea ice, heavy rain, and 18 GHz RFI.

High Resolution Cloud Liquid Water errors increase with proximity (~25km) to land, and sea ice.

While all of these factors are flagged and filtered by the L3 algorithm, if any of these factors are undetected, the errors will increase uncertainty in the L3 products.

Some retrievals tend to be more challenging and prone to larger errors at the extremes of their range. Although SST is less certain in colder waters due to lower signal to noise ratio, AMSR-E SST is performing better than many expected in cold water scenes. Wind retrievals at higher wind speeds become increasingly challenging for both radiometers and buoys.

The best estimates of data uncertainty under optimal conditions, for each measurement, including a confidence interval where possible.

Under optimal conditions, meaning land, sea ice, rain, high wind speeds, sun glint, and RFI have been detected with reasonable accuracy of over 95%, then

Variable	Values are within
SST	0.5 K
Wind Speed	1.0 m/s
Water Vapor	1.0 mm
Cloud Liquid	0.02 mm

Description of how to interpret quality flags to understand the conditions under which uncertainty may be greater.

The L3 algorithm provides an excellent guide to reading and interpreting the L2B flags to determine conditions in which uncertainty is large enough that data should be excluded.

The L2B Ocean_products_quality_flag gives detailed analysis of sea ice, rain, land, sun glint, and RFI glint. The meaning of these flags is documented in the quality assessment section of the AE_Ocean user guide.

2.4 Rain Products

Applicable data sets
AMSR-E/Aqua Monthly L3 5x5 deg Rainfall Accumulations (AE_Rain)
AMSR-E/Aqua Monthly L3 5x5 deg Rainfall Accumulations (AE_RnGd)
Applicable version(s): Version 2

Algorithm Name: Goddard Profiling Algorithm 2010V2 (GPROF2010V2)
Algorithm Authors: Christian Kummerow, David Randel, Ralph Ferraro

Synopsis on sources of uncertainty.

For the L2/orbital retrievals, the primary sources of uncertainty can be summarized as follows:

- **Database Construction and Representativeness** -- The databases constructed are considered the most robust over the 35 S to 35 N domains where Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) and precipitation radar (PR) data are constructed to get a realistic set of hydrometeor profiles. There is still some limitations in the PR itself due to its lack of sensitivity to lighter rain conditions. Outside of this domain, then the databases are built using an assortment of information coincident satellite and numerical model information, where the AMSR-E radiances are simulated through radiative transfer. There are large uncertainties in the radiative transfer calculations, especially in the presence of frozen hydrometeors and over the higher AMSR-E observation frequencies.

- **Bayesian Inversion** -- Within the Bayesian retrieval, there are far more degrees of freedom than the set of AMSR-E brightness temperatures that attempt to constrain the inversion. This uncertainty has been reduced over ocean in GPROF2010V2, where the databases are constructed as a function of sea surface temperature and total precipitable water. However, over land the degrees of freedom are larger and only convective and stratiform databases are use, thus, the uncertainty is much larger. (Notionally, we estimate that the uncertainty is twice as large over land as it is over ocean).

- **Non-linearity of rain rate to brightness temperature** -- The non-linear relationship between observed AMSR-E brightness temperatures and non-uniform precipitation within the AMSR-E field of view (i.e., commonly referred to as "beam filling") can lead to errors in the inversion. The lower AMSR-E channels used in precipitation have the most linear response to rainfall but have the largest footprints whereas the highest AMSR-E frequencies have the least linear response, but have the smallest footprints.

- **Surface Types** -- Over land, surface features such as snow cover and desert sand can be hard to separate from the rain signal, causing potential misclassifications. Melting snow can be extremely problematic, especially in the spring season. Retrievals along coastlines and inland water bodies are also difficult because of the large emissivity contrast between land and water.

Level 3, Global Monthly -- For the L3 global monthly rainfall, because of the twice a day (at best) AMSR-E overpass times (130 am/130 pm local time), large sampling errors are expected when translating the L2/orbital rain rates into a monthly rain accumulation.

The best estimates of data uncertainty under optimal conditions, for each measurement, including a confidence interval where possible.

Several papers on GPROF have been published over the years. For the most recent version of GPROF2010 (but not GPROF2010V2), here are the most appropriate references:

Gopolan, K., N.Y. Wang, C. Liu and R. Ferraro. 2010. Status of the TRMM 2A12 Land Precipitation Algorithm. *J. Appl. Clim. Meteor.* 27: 1343-1354.

Kummerow, C. D., S. Ringerud, J. Crook, D. Randel and W. Berg. 2011. An observationally generated a-priori database for microwave rainfall retrievals. *J. Atmos. and Oceanic Tech.* 28: 113-130. doi: 10.1175/2010JTECHA1468.1.

The table below summarizes the best estimates of uncertainty on a global scale. By using the TRMM PR as the reference source, the L2 retrievals are considered to be unbiased on a global scale however, regional biases do exist (e.g., see Gopolan et al. 2010).

	L2 Root Mean Square Error (RMSE)/Bias (%)	L3 RMSE/Bias (%)
Ocean	20% and 0%	10% and 3%
Land	50% and 0 %	25 % and 5%

Description of how to interpret quality flags to understand the conditions under which uncertainty may be greater.

In the L2 orbital products and the L3 monthly products, the quality flags are 0 (High quality - retrieval is good), 1 (Medium quality - use with caution) or 2 (Low quality - recommend qualitative use only). Specifically, this means:

<u>Ocean Algorithm</u>:
High: Grid box contains only good pixels
Medium: Grid box contains pixels where retrieval used extended database and/or expanded search radius for apriori database. Also where potential sun-glint occurs.
Low: Grid box contains pixels where retrieval used excessive search radius to find matches in apriori database.

<u>Land/Coast Algorithm</u>:
High: Grid box contains only good retrievals
Medium: Grid box contains ambiguous pixels
Low: Grid box contains missing or unable to retrieve pixels

2.5 Sea Ice Products

Applicable data sets
AMSR-E/Aqua Daily L3 6.25 km 89 GHz Brightness Temperature Polar Grids (AE_SI6)
AMSR-E/Aqua Daily L3 12.5 km Brightness Temperature, Sea Ice Concentration, & Snow Depth Polar Grids (AE_SI12)
AMSR-E/Aqua Daily L3 25 km Brightness Temperature & Sea Ice Concentration Polar Grids (AE_SI25)
Applicable version(s): Versions 2 and 3
Algorithm Name: Bootstrap Sea Ice Algorithm
Algorithm Author: Josefino C. Comiso

Synopsis on sources of uncertainty.

The main sources of uncertainties in the estimates of ice concentration are the spatial and temporal variabilities of the emissivity of 100% sea ice and the average temperature of the ice surface. The Boostrap Algorithm addresses these problems through the use of a multichannel algorithm that uses two sets of channels: 37 GHz channels (called VH37) at vertical and horizontal polarizations and 18 and 37 channels (called V1837) at vertical polarization. The VH37 set does very well in accounting for the changes in emissivity and surface temperature in a large fraction of the ice cover but it does not do well in surfaces where there are layering effects that would cause large changes in the brightness temperature at horizontal polarization but not as much at the vertical polarization. These cases are better handled by the V1837 set. The uncertainty is more predictable during the cold/dry seasons when changes in emissivity and surface temperature can be taken into account effectively. The uncertainty is larger during the wet periods of spring and summer when the snow cover goes through different stages that makes the effective emissivity unpredictable and when meltponding occurs making ice free areas virtually indistinguishable from meltponded areas.

There are other sources of uncertainties, and although they are not related to the basic ice concentration algorithm, they are very relevant, especially if the parameters of interest are ice extent, actual ice area and ice volume. One such source of uncertainties is imperfect land mask that may be due in part to temporal changes in land ice/ocean boundaries because of the occurrence of iceberg calving and the demise of ice shelves. A regular update of the land mask using high resolution data (not currently done) would solve the problem. Another source of error is the land-ocean boundary and antenna side-lobe effect which tend to smear the contrast between land covered areas and ocean areas. This results in the retrieval by the algorithm of significant fraction of sea ice in areas that are not known to be free of sea ice (like the coastal regions of Spain) all year long. An algorithm that takes this into account has been developed but is not perfect. A low-end threshold for ice concentration is also used because below this threshold, the microwave signature of sea ice cover areas and ice free areas are basically the same. An algorithm has been developed to establish the threshold which is currently between 5% and 10%. During storms and harsh weather conditions, this does not work very well and sometimes, significant fraction of ice is retrieved by the algorithm in open ocean areas far from the ice pack. Such effect has been minimized using sea surface temperature (SST) data and assuming that areas with SST greater than a threshold (usually 3 or 4°) is ice free.

The best estimates of data uncertainty under optimal conditions, for each measurement, including a confidence interval where possible.

Under optimal conditions, the best estimate for the uncertainty of ice concentration retrieval is 5% with a 95% confidence interval. This applies to a large fraction of the ice cover, especially in the Arctic basin where the ice is mainly consolidated and the data points are so well defined and clustered together that the impact of variations in emissivity and temperature is only within a standard deviations of about 2.5%. In other regions and especially in the marginal ice zone, the uncertainty is higher averaging between 10 and 15% because of the presence of so many types of ice and surfaces. The uncertainty is even higher (between 15% and 25%) during the summer period when the emissivity of snow and ice surface is unpredictable and meltponding occurs as indicated above.

Description of how to interpret quality flags to understand the conditions under which uncertainty may be greater.

The most important quality flags are the flags for land covered areas and ice free areas. Flags for ice and surface types would be useful but the available surface classification techniques are not yet mature enough for use in creating the flags. No other flags are considered necessary.

2.6 Snow Products

Applicable data sets
AMSR-E/Aqua Daily L3 Global Snow Water Equivalent EASE-Grids (AE_DySno)
AMSR-E/Aqua 5-Day L3 Global Snow Water Equivalent EASE-Grids (AE_5DSno)
AMSR-E/Aqua Monthly L3 Global Snow Water Equivalent EASE-Grids (AE_MoSno)
Applicable version(s): Version 2
Algorithm Name: AMSR-E/Aqua Global Snow Water Equivalent Algorithm
Algorithm Authors: M. Tedesco, R. Kelly, J. L. Foster, and A. T. C. Chang

Synopsis on sources of uncertainty.

There are several known algorithm-related error sources. Some are intrinsic to the problem of retrieving snow water equivalent (SWE) at large spatial scales from passive microwave observations (e.g., the small sensitivity of brightness temperatures to snow depth with respect to other parameters such as grain size, for example) where other factors are related to the heterogeneity of the scene observed by the sensor (e.g., mixed pixel problem) and by the simplicity of the algorithm (which is suggested by the operational nature of the approach). Some of these error sources are the temporal and spatial evolution of grain size and density, obscuration by forest, inability to map the water equivalent of partially wet snow covers, presence of water bodies, effects of the atmospheric attenuation.

A significant source of uncertainty is linked to the dynamic nature of retrieval coefficients. These depend on snow grain size and density, which can change dramatically during the snow season, even on short-term scales. After snow deposition on Earth's surface, snow crystals metamorphose in response to vapour gradients within the snowpack (either kinetic or equi-temperature forms) and as a result of melting and refreezing cycles. Also snowpack bulk density usually increases during the snow season. Although empirical and physically-based models have been developed to predict the growth of the snow crystal, it is not straightforward to select a general model that will account for regional to global scale conditions. The same can be said with regard to the snow density.

The presence of liquid water within the snowpack increases the absorption, reducing the penetration depth, making the SWE/snow depth retrieval impossible. To reduce the number of occurrences of erroneous retrievals in wet snow conditions, knowledge of surface or air temperature might not be enough: melting can occur also from the bottom of the snowpack from geothermal heat fluxes.

The presence of water bodies within the area under study can affect the retrieval of SWE because of the strong brightness temperature gradient between liquid and frozen water. Analysis of airborne passive microwave data acquired in the Northwest Territories, Canada in April 2005 showed the relationship between 37 GHz brightness temperature and lake cover fraction is reversed across the northern boreal forest compared to the open tundra. Over forested sites, lower 37 GHz brightness temperatures were measured over lakes relative to land, while the 19 GHz data showed little sensitivity to lakes. Conversely, at tundra sites the 37 GHz brightness temperatures were higher over lakes than over terrestrial surfaces. This difference in response to lake ice at 37 GHz will have an effect on SWE retrievals because the increase in brightness temperature at 37 GHz across lake rich tundra areas will decrease the 37-19 GHz (or 37-10) difference, and therefore decrease SWE estimates with conventional algorithms.

Forest cover represents an important source of error, representing a major challenge to the refinement of a robust passive microwave SWE retrieval algorithm. Indeed, the presence of forest attenuates the radiation emitted by the underlying snowpack, affecting the retrieval accuracy of the algorithm. In general, the problem shows a high degree of complexity; both fractional volume and stem closure within a footprint are important modulators of the passive microwave emission. Crown closure, basal area and foliage biomass are all inversely related to visible reflectance and are also directly related to microwave emission.

Upwelling microwave radiation emitted from the Earth's surface passes through the atmosphere before being detected by the space-borne sensor, and thus it is subject to the effects of atmospheric absorption and (re-)emission. These effects have been neglected so far in the retrieval algorithm but recent studies show that they should be accounted to reduce retrieval errors further.

The best estimates of data uncertainty under optimal conditions, for each measurement, including a confidence interval where possible.

A new Algorithm Techical Basis Document (ATBD) has been submitted to NASA for evaluation of a refined operational AMSR-E SWE product. The table below reports the monthly averaged Root Mean Square Error (RMSE) (in cm) of the current operational and the new proposed algorithms obtained considering ~ 2000 WMO ground observations.

	Current algorithm	**New proposed algorithm**
	RMSE [cm]	**RMSE [cm]**
January	20.661	17.931
February	22.775	19.173
March	27.336	24.703
April	32.123	27.897
October	11.4344	10.3453
November	13.6451	13.156
December	17.1517	16.1324

SWE values are also currently being assessed using other SWE products produced by European colleagues. As an example, the figure below shows the comparison between monthly January SWE values obtained with the new AMSR-E proposed algorithm and those generated by the GLOBSNOW project for the period 2003 - 2011. The SWE distributions obtained with the two products are fitted with Gaussian distributions and the corresponding fitted parameters are reported in the figure.

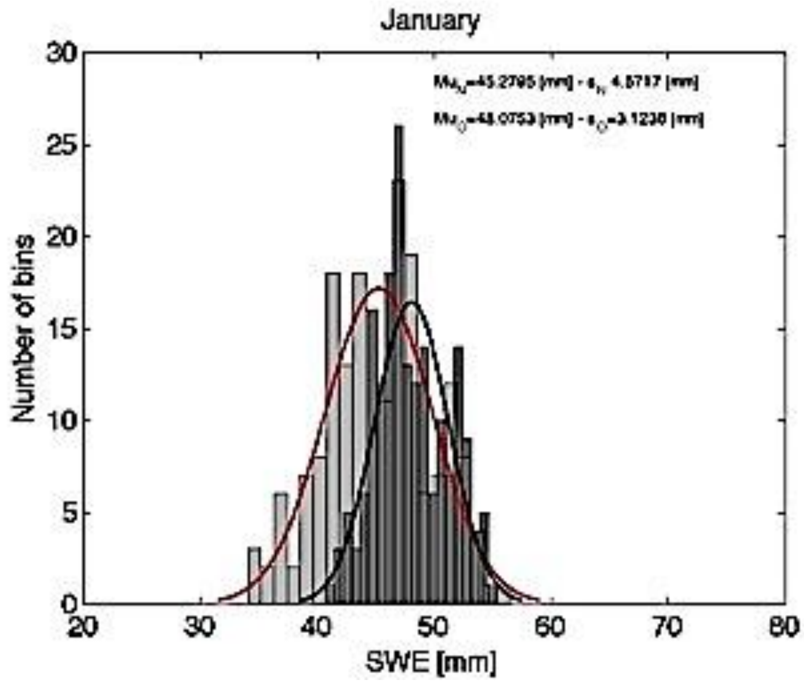

Figure 1. SWE Distribution

Description of how to interpret quality flags to understand the conditions under which uncertainty may be greater.

The current AMSR-E SWE product does not produce maps of quality flags. These will be generated in the new proposed product (if approved) and it is part of the ongoing and final work of the current sponsored project.