Kidd Christopher (Orcid ID: 0000-0001-5906-8446)
Tan Jackson (Orcid ID: 0000-0001-7085-3074)

Validation of the Version 05 Level 2 precipitation products from the GPM Core Observatory and constellation satellite sensors.

Christopher Kidd[ab*], Jackson Tan[bc], Pierre-Emmanuel Kirstetter[d] and Walter A. Petersen[e]

[a]*Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA*

[b]*NASA Goddard Space Flight Center, Greenbelt, MD, USA*

[c]*Universities Space Research Association, Columbia, MD, USA*

[d]*National Weather Center, Norman, OK, USA*

[e]*NASA-MSFC Earth Science Office, National Space Science and Technology Center, Huntsville, AL, USA*

[*]Correspondence to: *C. Kidd, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, 20740, USA. E-mail: chris.kidd@nasa.gov*

Abstract

The accurate measurement of precipitation, an essential source of freshwater, is key for assessing water resources and precipitation-related hazards across the globe. The measurement of global precipitation is hampered by the inadequate spatial distribution of ground-based observations

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3175

such as gauges and weather radars. However, satellite instruments can provide global observations from which precipitation estimates may be generated. The satellites of the Global Precipitation Measurement (GPM) constellation carry passive microwave sensors that provide an essential set of observations from which the Goddard PROFiling (GPROF 2017v1) scheme is used to retrieve precipitation. The regional consistency and differences in the satellite retrievals need to be documented in order to refine and improve global precipitation estimates which form the basis of widely-used gridded precipitation products. Here the Level 2 instantaneous swath-based precipitation products generated by the GPROF scheme are evaluated using standard descriptive and statistical scores against national/international surface radar and dense gauge datasets, over the United States and Western Europe. Results show that over Europe the current GPROF retrieval technique tends to overestimate the occurrence of light precipitation, leading to an overestimation of the volumetric contribution by light precipitation intensities, while it underestimates moderate to heavy precipitation. Over the US the overestimation of light precipitation is reduced, with a more pronounced overestimation of moderate precipitation intensities, and an underestimation of heavier precipitation intensities.

Keywords: Satellite, precipitation, inter-comparisons, validation, GPROF, Global Precipitation Measurement.

1. Introduction

The measurement of the distribution and intensity of precipitation across the globe is critical to furthering our understanding of the complex systems and feedbacks that underpin the global water cycle, as well as for measuring water resources central to our environment and society. The accuracy of such measurements is crucial in achieving a complete description of not only the current state of global precipitation, but also to monitor and measure changes in the characteristics of precipitation over time. Although rain gauges may be seen as the *de facto* conventional surface measurement, they have a number of artefacts that require careful quality control (see Delahaye et al. 2015) while lacking representativeness on a global scale due to their paucity of numbers and non-uniform distribution (Kidd et al. 2017). Similarly, radar networks are inadequate on a global scale, largely restricted to land areas, and are accompanied by a number of artefacts that require additional processing to enable such data to be usefully employed. Therefore, satellite observations must form the basis of global-scale precipitation measurements.

Satellite observations useful for precipitation estimation may be categorised into a number of groups based upon the type of satellite orbit and frequency/wavelength band of the observations. Two broad orbital types are geostationary and low-Earth orbits, while observations may be grouped by visible, infrared and microwave portions of the spectrum. While geostationary satellites offer a wide view of the Earth, their capabilities are limited to visible and infrared

observations. Low Earth orbiting satellites are necessary for (passive and active) microwave sensors. (For a more complete review see Kidd and Levizzani, 2011).

Over the last decade there has been a concerted effort to provide an international constellation of precipitation-capable satellites/sensors to ensure good temporal and spatial coverage globally at scales commensurate with the variability precipitation systems. The launch of the Global Precipitation Measurement (GPM) Core Observatory (CO), on 27 February 2014, provides a focus for this constellation through advanced instrumentation, namely the GPM Microwave Imager (GMI) and the Dual-frequency Precipitation Radar (DPR) (see Hou *et al.* 2014). The constellation of international partner satellites currently provides 10 additional sensors, namely the Advanced Scanning Microwave Radiometer (AMSR-2), Special Sensor Microwave Imager/Sounder (SSMIS, 3 currently operational), Microwave Humidity Sounder (MHS, 4 currently operational), Sondeur Atmosphérique du Profil d'Humidité Intertropicale par Radiométrie (SAPHIR), and the Advanced Technology Microwave Sounder (ATMS). The availability of well-calibrated instrumentation on the GPM Core Observatory allows it to be used as a reference and transfer standard for the other constellation sensors to ensure consistent observations, and subsequent precipitation retrieval thereafter (Wentz and Draper, 2016).

Alongside the advances in satellite instrumentation, techniques have been developed and improved to identify and estimate the intensity of precipitation from the range of satellite

observations now available. Precipitation retrieval schemes can be broadly divided into those based upon visible/infrared observations, passive microwave (PMW), active microwave (AMW, i.e. radar) observations, or combined/merged techniques based upon a combination of these observations. The retrieval of precipitation from visible/infrared observations has a long association with satellite observations through the relationship of the presence of clouds and cloud-top characteristics with the underlying precipitation. However, despite the development of multi-spectral techniques, these still rely upon the cloud tops being representative of the precipitation falling from the cloud base. Techniques based upon PMW observations have been developed over the last forty years. Since these observations are affected by the precipitation sized particles within the atmospheric column, they are more directly associated with the surface precipitation. Satellite-based precipitation radars, although limited to the Tropical Rainfall Measurement Mission (TRMM) Precipitation Radar (PR) and the GPM DPR, provide the most direct measure of precipitation from space. Although techniques aimed to exploiting the better temporal and spatial sampling of the geostationary observations and the better retrieval ability of the microwave observations have been developed (e.g. CMORPH: Joyce et al., 2004, Joyce and Xie, 2011; PERSIANN-CCS: Yang et al. 2004), they have not necessarily achieved a full level of maturity.

The ground validation of the satellite-based precipitation estimates is an essential part of the development of retrieval techniques. Often short-period local studies are carried out to test the

viability of the algorithms. The GPM has conducted a series of dedicated ground validation (GV) campaigns primarily aimed at collection of detailed information from multiple sensors on multiple platforms, including satellites, aircraft and surface (see Hou et al., 2014). This multi-tier approach enables a full picture of the precipitation to be captured enabling a better understanding of the relationship between the *in situ* precipitation and the satellite observations (e.g., Jensen et al. 2016, Skofronick-Jackson et al., 2015; Petersen et al., 2016; Houze et al., 2017). However, large scale validation is necessary to assess their performance on a regional or global basis, over extended periods to capture the full range of precipitation systems and seasonality: Global studies based upon gauge analyses, such as that of the GPCC, suffer from the paucity of data over the oceans and the coarseness of the spatial and temporal resolutions. However, national weather radar precipitation measurements have improved significantly and provide quantitative estimates at temporal and spatial scales necessary for the validation of satellite estimates (Kirstetter *et al.* 2012, 2014). Some of these networks have been usefully employed in the validation or inter-comparison of precipitation products, although generally at daily, 0.25° × 0.25° scales (see Ebert et al. 2007; Kidd et al. 2012).

This paper provides an assessment of the latest precipitation products available from GPM constellation, namely the Level 2 (instantaneous, swath-based) V05 products of the NASA Precipitation Processing System (PPS) based upon the GPROF 2017v1 scheme for the radiometers and the DPR. Comparisons are made against surface reference data sets, both

surface radar and gauge, over Western Europe and the eastern US for a three year period from April 2014 through March 2017.

2. Methodology

The following section outlines the satellite and surface data sets and the methods of assessment used in this analysis: this analysis is performed at the instantaneous, 15 km x 15 km scale using co-located and co-temporal matchups for both radar-satellite and gauge-satellite observations.

2.1. Satellite retrievals

Two types of sensors provide Level 2 (L2) precipitation estimates for this study, the GMI, AMSR2, SSMIS, MHS and ATMS radiometers, and the DPR radar. The data products are the current (and latest) V05 data sets which are freely available online the NASA Precipitation Processing System (pmm.nasa.gov). Table 1 outlines the characteristics for each of these sensors included in this study.

The radiometer precipitation products analysed here are generated by the Goddard Profiling (GPROF) retrieval scheme, described as a fully parametric physically-based retrieval scheme (Kummerow et al. 2015). Although the retrieval of precipitation built upon basic radiometric properties have achieved a degree of success, providing an unambiguous retrieval of precipitation is often difficult due to the variability of the surface background, or the non-unique

spectral signature to hydrometeor profile/surface rainfall relationships. A greater insight into the radiometric signature may be obtained through inverse radiative transfer modelling, although such techniques are ultimately limited by the model itself and the computational requirements (see Olson et al. 2007). Nevertheless, the development of a physical technique, such as GPROF, offers the flexibility to model the characteristics the sensors, allowing multiple frequencies, resolutions, etc. to be tailored to individual sensors through a common set of atmospheric profiles to ensure consistency between the different sensors. In addition, consistency between the radiometers is facilitated by the inter-calibration between the sensors at the Level 1C (L1C) brightness temperature (Tb) level, the GMI providing the standard for radiance calibration in the GPM constellation (Berg et al., 2016).

The GPROF retrieval scheme comprises of three main processing steps as 1) the generation of an *a priori* database that contains radiances with associated surface rainfall, hydrometeor profiles and other pertinent data; 2) pre-processing of the satellite observations which attaches the surface types and model-derived surface (or 2 m) temperature (Ts) and total precipitable water (TPW) to the set of brightness temperatures for each satellite footprint, and; 3) the conditional probabilistic Bayesian retrieval of the precipitation by comparing the satellite observations to the radiances in the (appropriate) database with the same Ts and TPW. Key to the success of the GPROF scheme is the generation of the *a priori* database, or in practice, databases. The databases are populated by a range of suitable hydrometeor profiles that are categorised by Ts and TPW; this based upon

the findings of Berg et al. (2006) to account for regional differences in the relationship between surface precipitation and satellite radiances. This not only constrains the retrieval but also improves computational efficiency of the retrieval scheme. In addition, since the surface background influences the radiometric signatures, separate databases are generated for 14 different surface types. For the cross-track sensors (or sounders) these databases are generated for each of 3 different scan position groups (except V03) to account for differences in atmospheric path with scan position. Each of the sensor type has a separate group of databases to account for the different resolutions and sensor frequencies, but in V05 are all derived from the primary GMI database. At the retrieval stage the cross-calibrated L1C Tb data is read in along with model information and ancillary data sets (such as snow maps, land surface types) which allow the correct database to be accessed and the correct database bin to be interrogated: the Tbs of all eligible database profiles are then compared with the observed Tbs through the Bayesian scheme to allow a final result to be generated.

Since the launch of the GPM Core Observatory there have been three versions released: V03 (publically available from June 2014), V04 (2016) and the current V05 (May 2017). V03 represented the 'at-launch' retrieval and was based upon GPROF2014, a retrieval scheme was adapted from the TRMM scheme (see Kummerow et al., 2001) which necessarily included several approximations. This version used a pre-launch observational database based upon surface (Multi-Radar/Multi-Sensor; MRMS) and satellite (TRMM PR) radar data (Kummerow et

al. 2015), except for the MHS, ATMS and SAPHIR which was based upon a model-generated database (Kidd et al. 2015). A post-launch version (V04) was later released utilising the GPM DPR and simulations of the hydrometeor profiles to populate the databases and provide a unified scheme across all sensors.

The current (V05) uses databases that contain profiles where observed satellite radiances have been matched against the DPR-GMI combined product, together with the addition of profiles derived from the CloudSat Cloud Profiling Radar (CPR; see Stephens et al. 2002) to augment the profiles in low-intensity regimes to provide low-end sensitivity to light precipitation, including snowfall that might be missed in the DPR/GMI combined retrievals.

The DPR on board the GPM Core Observatory builds upon the success of the PR flown on TRMM between 1997 and 2015. The DPR consists of Ku- and Ka-band radars that scan cross-track with swaths of approximately 245 km and 125 km respectively. The Ku-band scan or 'normal swath' (NS) is used in this study and consists of observations of 49 scan positions each with a ground footprint diameter of 5.2 km at the surface. The minimum detectable radar reflectivity factors at 400 km range are approximately 12 dBZ for the DPR-Ku which relates to about 0.5 mm h^{-1} rainfall intensity.

Processing of the data is first done separately for the Ku-band and Ka-band channels, followed by the processing of Ku-Ka combined channels. For all three 'channels' (Ku, Ka and Ku+Ka), the basic algorithm flow is essentially the same: conversion of radar return powers into radar reflectivity factors, or normalized surface radar cross section, depending on whether the radar returns arise from the atmosphere or surface; a rain/no-rain determination; a classification algorithm to determine rain type (stratiform/convective) and ancillary information (presence of bright-band, snowfall at surface, hail detection, etc.); an algorithm using the surface return to estimate path integrated attenuation; and finally a retrieval algorithm to estimate, as a function of height, the water content, rainfall rate, and parameters of the particle size distribution (see Iguchi et al., 2016). The DPR-Ku near surface precipitation parameter is used in this study with the 5.2 km x 5.2 km footprints being resampled to approximately 15 km x 15 km (using a 3 x 3 averaging) to match the resolution of the radiometer retrievals.

2.2 Surface data

Surface reference data sets form an essential element of validating satellite-based precipitation. These reference data sets may be obtained from a variety of sources including surface-based weather radar and gauge data sets, and range from instantaneous scales through to long-term monthly temporal scales (Roca et al., 2010, Kirstetter et al. 2013c). Fine temporal/spatial comparisons are usually conducted to provide physical validation of the satellite observations

and retrievals, such as those carried out by the GPM-GV programme (see Petersen et al., 2016). These studies, usually carried out over a limited region for a limited time, often incorporate airborne data to provide information on the microphysical nature of the precipitation system to better understand the precipitation-observation relationships. Comparisons made at coarser resolutions (0.25° x 0.25° daily or monthly) have, however, been the mainstay of many inter-comparison projects (e.g. Adler et al. 2001, Ebert et al. 2007, Kidd et al. 2012) and rely upon surface gauge analyses or gridded radar data.

At the instantaneous scale surface radar and dense gauge networks provide an essential means for verifying satellite retrievals due to the available high spatial and temporal resolutions (Kirstetter et al. 2015b). This study exploits the availability of mature high quality radar networks over both the US and Western Europe, together with two small-scale dense gauge networks over the US (Pocomoke, MD) and Europe (WegenerNet, Austria). These networks are outlined below.

2.2.1 The Nimrod Western European radar product

The United Kingdom Meteorological Office (UKMO) generates a 5x5 km radar product over Western Europe extracted from the UKMO very short-range NIMROD forecasting system (CEDA; Met Office 2003). The radar composite is collected from a network of C-band weather radars at 15-minute intervals, covering the United Kingdom, France, Germany and the

Netherlands. Assessments for the radar quality over this region may be found in Harrison et al. (2009, 2010) and Fairman et al. (2015). The data undergoes basic quality control to account for fundamental radar artefacts although no quality index is available with the radar data. Consequently an initial comparison was undertaken to identify and ascertain any remaining artefacts in the radar data. Based upon a Heidke Skill Score analysis of GPM DPR and surface radar, a mask was generated to remove extraneous precipitation relating to surface radar range effects, beam blockages and anomalous propagation; this mask is then used to exclude these regions from the comparison over Western Europe (see Figure 1). In addition, since this radar product does not provide an indication of the phase of precipitation, for this study the satellite (model-based) rain/snow flag is used to exclude frozen precipitation. The 15-minute radar product generated during the satellite overpass is used and a nominal 15 km x 15 km resolution is generated to compare with the satellite products; while the actual retrieval resolution varies (see Table 1) the standardised resolution permits easier cross-sensor assessments to be made.

2.2.2. US MRMS radar

Over the US the MRMS radar network products generated by NOAA's National Severe Storms Laboratory/University of Oklahoma (Zhang et al., 2016) are utilised. The MRMS system incorporates observations from polarimetric WSR-88D radars (NEXRAD), automated rain gauge networks and model analyses over the continental US (CONUS). The system creates gridded mosaics of precipitation products from quality-controlled reflectivity data at 0.01° x 0.01° and 2-

minute update cycle. Uncertainties in the MRMS radar data (e.g. beam blocking, variations in vertical reflectivity, conversion from reflectivity to precipitation rate, etc.) are corrected in the system. Precipitation rate, phase and type (convective, stratiform) are generated along with a radar quality index (RQI) representing the radar QPE uncertainty associated with beam blocking and radar beam height relative to the melting layer. At the hourly time scale the radar QPE is blended with gauge data.

Kirstetter et al. (2012) describes the basis for comparing the MRMS to space borne retrievals through the use of high-quality reference precipitation data derived by applying additional quality and quantity controls to complement the correction procedures already in place within the MRMS system. The 0.01° x 0.01° 2-minute resolution of the reference precipitation provides flexibility in matching to the satellite sensor (Kirstetter et al. 2013b, 2014, 2015b; Carr et al. 2015). Only the best quality radar-based liquid precipitation estimates are kept by using the RQI to minimize radar sampling uncertainties. Pixel-by-pixel ratios between the hourly gauge-corrected and the hourly radar-only products are used to adjust the radar-only 2-min precipitation rates. As a quantitative data control procedure, pixels showing extreme disagreement between radar and gauge caused by residual artefacts (e.g. ground clutter) are discarded and no reference is generated. These procedures yield consistent and homogeneous levels of quality to the reference for directly evaluating precipitation from the sensors of the GPM constellation. For this study the 0.01° x 0.01° radar product is resampled (to improve computational efficiency) to a

5x5 km equal area grid over the eastern half of CONUS where radar and gauge networks sampling conditions are good relative to terrain and density of networks (see Figure 1). This 5 km x 5 km 2-minute data is then matched against the satellite products only where the RQI is at the highest (=1.0); a consequence of using the highest RQI is that only liquid precipitation (i.e. rainfall) in included in this study.

2.2.3. Gauges

Two dense well maintained gauge networks are used in this analysis to provide complementary independent validation sources to the radar networks described above.

The first network, the Pocomoke network (PCMK; e.g., Tan et al., 2016), is maintained and calibrated on a regular basis by the NASA GPM-GV program. PCMK is located in the mid-Atlantic region of the United States near NASA Wallops Flight Facility, a coastal environment with no significant orography. The main precipitation systems affecting this area are tropical cyclones (or remnants of tropical cyclones), airmass convection, nor'easters, and frontal rainfall (Tokay et al. 2014). The PCMK network contains up to 25 dual tipping bucket rain gauge platforms (50 gauges in total), each recording the time of every tip, with each tip corresponding to 0.254 mm of rain. The actual number of gauge platforms at any one time varies due to gauge maintenance and calibration, as well as addition and removal of platforms. The gauges are distributed over an area approximately the size of an individual 5 km DPR footprint, and hence

provide a robust measure of area-mean rainfall at the DPR footprint scale. As these gauges are not able to accurately measure the snow water equivalent rate, individual days, together with the following day, that are associated with snowfall reports in the daily reports from the NOAA Global Historical Climatology Network–Daily are excluded. The PCMK data is available at https://gpm-gv.gsfc.nasa.gov/Gauge/index.html.

The second gauge network is the WegenerNet (WEGN) located in the Feldbach region of southeast Austria (Kirchengast et al., 2014). The region lies in the southeastern Alpine foreland with moderate orography amidst a river valley. WEGN contains 153 stations, each providing calibrated rainfall data every 5 minutes. A quality control flag is available with the WEGN data to identify less reliable estimates as well as frozen precipitation, which is excluded in this analysis. WEGN data is available at http://wegenernet.org/.

To evaluate the GPROF estimates at the individual footprint level, gauges are matched to each footprint by identifying gauges that are within a nominal distance of 7.5 km from the footprint centre, i.e. a circular satellite footprint of 15 km in diameter is assumed. For the DPR-Ku footprints, a nominal distance of 2.5 km is used instead. To ensure representativeness within the footprint, only pixels with at least 12 gauges (or 6 pairs of gauges for PCMK) are considered for comparisons with GPROF products and at least 6 gauges (3 pairs for PCMK) for DPR-Ku

product. Gauge values averaged between ~~n30 min~~ ±30 min of the overpass time are used in the comparisons.

2.3. Statistical analysis

A statistical analysis is performed on the instantaneous co-located and co-temporal matchups for both radar-satellite and gauge-satellite at a nominal resolution of 15 km x 15 km. Standard statistical scores of bias (mean error), normalised root mean square error (NRMSE) and correlation are calculated for all valid coincident satellite-surface data. Additional metrics, namely the Probability of Detection (POD), False Alarm Ratio (FAR) and Heidke Skill Score (HSS) are used to assess the ability of the products to correctly identify precipitation (Wilks, 1995).

For the statistical scores only coincident data where both the satellite and surface data have precipitation rates greater than or equal to 0.2 mm h^{-1} are used (the lowest threshold specified for the GPM-CO and the minimum detectable threshold for the DPR-Ku; see Skofronick-Jackson et al. 2017): this avoids the statistics being overly influenced by the dominance of the non-raining and very light raining points that are prevalent at the instantaneous scale. For the POD/FAR/HSS scores a rain/no-rain threshold of 0.2 mm h^{-1} is used (i.e. <0.2 mm h^{-1} is considered to be no-rain, while e 0.2 mm h^{-1} is considered to be rain).

The surface and satellite radar data sets have been reduced in resolution to the nominal 15 km x 15 km resolution of the passive microwave radiometers. The Nimrod data is originally reported at a quantised resolution of 1/32 mm h^{-1} (or 0.03125 mm h^{-1}), although when the original 5 km x 5 km resolution is averaged over a 3x3 area to form the 15 km x 15 km comparable resolution, this becomes 0.00347 mm h^{-1}. Likewise, the DPR-Ku products have an original resolution of 5.2 km x 5.2 km with a lower detection threshold of about 0.22 mm h^{-1}. When this is averaged to approximate the 15 km x 15 km resolution for the comparisons a lower threshold of 0.031 mm h^{-1} is realised.

3. Results

The results of the analysis are presented as follows: first, descriptive analyses, using the distribution of observed precipitation rates by occurrence and accumulation are used to assess the ability of the products to capture the distribution of precipitation intensity characteristics. These are backed-up by 2D normalised density scatterplots. Comparisons are then made between the satellite products and the surface gauge networks: the gauge network data being independent of the surface radar data sets. Seasonal plots of the performance of the satellite products (with respect to the surface radar) are used to assess seasonal dependencies. Lastly, statistical analysis is used to quantify the performance of the precipitation products with respect to the surface data sets. All of the comparisons cover the full data period (April 2014 through March 2017) and are

based upon instantaneous matchups between the satellite products and surface radar/gauge data with a nominal resolution of 15 km x 15 km.

3.1. Representation of precipitation intensity

3.1.1. European region

The distribution of the occurrence of precipitation by intensity for each of the satellite and surface radar precipitation products is shown in Figure 2. The peak of the maximum occurrence (excluding no-rain) differs between the different sensors; in particular, all the SSMIS retrievals show a significantly higher maximum peak at about 0.13 mm h^{-1}, but commensurate with the location of a lower peak of the MHS, ATMS and GMI sensors. The AMSR2 retrievals peak at a high precipitation intensity around 0.25 mm h^{-1}. In comparison, the surface radar shows a much broader peak, reaching a maximum between about 0.30 and 0.70 mm h^{-1}. All the satellite products underestimate the occurrence of precipitation above about 0.40 mm h^{-1}. The DPR-Ku has a similar distribution as the surface radar but with a lower peak.

In terms of the accumulation of precipitation by intensity (Figure 2, *right*), although there is a preponderance of light precipitation occurrence, this contributes relatively little to the total accumulated precipitation. The peaks of the accumulated precipitation (by intensity) are very similar for the GMI and surface radar, although the GMI is slightly lower and has a broader distribution compared to that of the surface radar. The main peak in the SSMIS products is

slightly below that of the surface radar and has a not insignificant contribution from light precipitation with this secondary peak around 0.20 mm h^{-1}. The MHS products have similar maximum peaks as the SSMIS, but no low-end secondary peak. The AMSR2 distribution is somewhat unusual, with a low-end secondary peak at about 0.40 mm h^{-1}, and a sharply defined main peak at around 3.55 mm h^{-1}.

The distribution of occurrences by intensity for both the satellite product and surface radar are reflected in the 2D density scatterplots shown in Figure 3. A number of features can be noted: first, the detection of light precipitation by the surface radar (x-axis) is clearly evident, albeit with quantization issues at the very lowest precipitation intensities, while the GPROF retrievals for all satellites show little detection of precipitation below about 0.05 mm h^{-1}. Second, while most of the observations above about 1.0 mm h^{-1} fall close to the 1:1 line, below 1.0 mm h^{-1} the bulk of the GPROF retrievals lie generally below the 1:1 line before becoming almost insensitive to the precipitation rate as derived by the surface radar; this is evidenced by the horizontal maxima in the GPROF precipitation intensities around 0.2 mm h^{-1} (most noticeable in the SSMIS plot). The DPR-Ku has the least scatter amongst the data products with most points lying close to the 1:1 line. However, at high intensities the DPR-Ku overestimates the precipitation, although in relatively small numbers.

3.1.2 Eastern US region

The distribution of precipitation by intensity for the US is shown in Figure 4. While the overall form of the distribution is similar to that of the European region, there are some subtle but notable differences. In terms of the occurrences of precipitation (Figure 4, left), the distribution of the surface radar occurrences peaks at about 0.56 mm h^{-1}, similar to that of the European region, but has a greater number of high-intensity events. The distribution of the DPR-Ku product is very similar to the surface radar but lower, as was the case over Europe. All GPROF products generally overestimate the occurrence of precipitation (with respect to the radar) across most intensities, and particularly so at the low intensities. The SSMIS has a notable low-end peak at about 0.13 mm h^{-1}, while the MHS/ATMS have broader peaks at slightly higher intensities at about 0.18 mm h^{-1}. Both the AMSR2 and GMI precipitation products exhibit double peaks at about 0.08 mm h^{-1} and 0.79 mm h^{-1}, with AMSR2 showing a much sharper peak.

In terms of the accumulation of precipitation by intensity, the surface radar peaks at about 4.0 mm h^{-1}; the peaks of the MHS/ATMS satellite products are slightly higher at about 5.0 mm h^{-1}, but the SSMIS, AMSR2 and GMI products peak much higher at intensities around 7.1 and 6.3 mm h^{-1} respectively. The SSMIS and MHS/ATMS products overestimate the contribution of the accumulated precipitation at intensities up to about 10.0 mm h^{-1}, then underestimate the contribution. While the AMSR2/GMI is in broad agreement with the surface radar up to about 2.0 mm h^{-1}, they overestimate the volume of precipitation up to 11.0 mm h^{-1} before underestimating the contribution from higher intensities. It is notable that both the MRMS

surface radar and the DPR-Ku generates significantly more precipitation than the radiometer products above about 10 mm h^{-1}, highlighting the limitations of the current GPROF algorithm in producing sufficient intense rain rates.

Over the US the 2D scatterplots (Figure 5) show that the radiometer products tend to overestimate the precipitation, although the DPR-Ku is very close to the 1:1 line above 1 mm h^{-1}; below this it tends to underestimate precipitation. Regarding the distribution of values around the 1:1 line, the GMI/AMSR2/SSMIS are close although somewhat higher, although both the GMI and the AMSR2 products exhibit an anomaly at around 0.7- 0.9 mm h^{-1}, as evidenced by the horizontal maxima of values in the plots. The distributions of the MHS and ATMS products are very similar, both being broader than that of the imaging radiometers and above the 1:1 line.

3.2 Gauges

Due to the smaller sample size associated with the satellite-gauge matchups only scatterplots of the data will be considered here.

Figure 6 shows the DPR-Ku scatterplots for the US PCMK and the Austrian WEGN gauge networks. The smaller number of available matchups is evident in the PCMK scatterplot where the scatter of points seems to be greater particularly for the convective case (as identified in the satellite product and shown here as blue dots), although the stratiform cases have a similar

distribution to those from WEGN. Over the WEGN site the (satellite-identified) convective cases are in good agreement with the points lying close to the 1:1 line. The heavier stratiform precipitation is underestimated by the DPR-Ku product, while the lighter stratiform precipitation is over estimated by the DPR-Ku. This distribution is more evident in WEGN than in PCMK due to sample sizes. The overestimation of light precipitation and slight underestimation of moderate-to-heavy precipitation is consistent with the respective results obtained from ground radars.

Scatterplots for all of the GPROF products vs gauge data are shown in Figure 7; due to limitations in sample size all sensors are plotted together. The scatter of the GPROF retrievals is similar to that of the DPR-Ku retrievals, but with a larger spread in values about the 1:1 line. Over the PCMK site the GPROF products appear to slightly overestimate the mid-range intensities between 1.0-5.0 mm h^{-1}; conversely over WEGN the GPROF products slightly underestimate the mid-range of intensities, although the overall agreement is good. Overall, the comparisons with the gauge data are in line with the results from comparisons with surface radar data.

3.3. Seasonal performance

It is well known that the performance of satellite precipitation estimates has some seasonal dependence, particularly over the middle and higher latitudes where the seasonal cycles in

temperature and precipitation regimes are greatest. To assess the performance of the satellite estimates plots of the bias and correlation have been generated for each month at the instantaneous, 15 km x 15 km scale for both the European and US regions (see Figure 8). It is worth noting that the European region is located at between about 45°-60°N, compared with the US region between 20°-55°N. To aid the comparison a 3 month moving average was used on the resulting statistical scores.

It is clear that over the European region that there is a strong and regular seasonal cycle in both the V05 bias and correlation statistical scores. Across all seasons and all retrievals the bias is generally negative, although the AMSR2 (and GMI in 2014) just manages a positive bias during the summer months. Except for June 2014, the DPR-Ku has a negative bias, particularly so during the winter, similar to the radiometer products. Overall, the MHS, ATMS and SSMIS have the most negative bias. The correlation statistics reflect the seasonal cycle shown in the bias; the lowest values are during the winter period, and the highest during the summer. The DPR-Ku and GMI performance is quite similar, each attaining correlations close to 0.6 during the summer period. This seasonal cycle in the performance of the techniques is likely to be related to the types of precipitation systems observed over the year, being more convective during the summer versus cyclonic during the winter.

Over the US region variations in the bias are much less clearly organised; both negative and positive biases are present and the variation within each product is quite marked. The GMI and AMSR2 products have generally positive biases (except for Autumn 2014 and Winter 2016/2017), the DPR-Ku both negative and positive biases, while the MHS, ATMS and SSMIS have largely negative biases. In contrast to the results for Europe, the phases of the variations are not necessarily aligned between the different sensors, such as the MHS during Winter 2014/2015. In addition, most sensors have their greatest negative bias during the Summer, and greatest positive bias in the Spring. In terms of correlation, all products have a more steady performance across the whole period, again with slightly better performance during the Winter for the radiometer products. The DPR-Ku performs best across virtually the whole period, with correlations typically between 0.6 and 0.7. A hierarchy of performance is seen here: the DPR-Ku is best, followed by the GMI, AMSR2, SSMIS/MHS and finally the ATMS.

Over both the European and the US regions the correlation statistics of the retrievals are very much related to the directness of the products: the DPR performs best, followed by the GMI which is calibrated by the DPR. The other techniques are based upon the GMI retrievals and the decline in their performance is largely related to the (reduced) channel availability and (often poorer) resolution.

3.4. Statistical performance

Summaries of the statistical analysis of products are shown in Figure 9 and 10. Over Western Europe (Figure 9) the comparisons between the satellite products and surface radar data show a general underestimation (negative bias) with respect to the surface radar data set, which is also confirmed by comparison to the WegenerNet data. The NRMSE is notable for being consistent across the products and between the surface reference data sets. Correlations of the satellite vs surface data show reasonable correlations of over 0.5 for the GMI and AMSR2 conically scanning sensors, just below 0.5 for the DPR-Ku, while the cross-track sensors have correlations around 0.4. Interestingly, the GMI retrieval exhibits the highest correlation versus the surface radar; the GMI retrieval database forms the basis of the other radiometer databases, with the AMSR2 and SSMIS being most similar. Correlations of the satellite versus gauge data are somewhat more varied, with the MHS being the only retrieval with a correlation greater than 0.6, and the SSMIS performing quite poorly (correlation <0.4).

In terms of POD, FAR and HSS, the DPR-Ku performs very well against the surface radar. Most notable is the very low FAR which may be attributed to the poor sensitivity to very light precipitation (<0.25 mm h^{-1}) by the DPR-Ku, removing the detection ambiguity at the low precipitation intensities. The GMI has the second best HSS, with a gradual reduction of HSS for the other retrievals. Against the gauges the DPR has the highest HSS (0.72), while most the radiometers have HSS of around 0.61: the ATMS is the exception with an HSS of 0.54.

Over the United States the satellite products show low bias except for the ATMS product (negative bias) against the radar network and the GMI (positive bias) against the gauge network. The NRMSE is generally higher over the US when compared to the surface radar than it was over Europe (this despite being normalised by the mean rain rate). However, the higher NRMSE exhibited by the products against the surface radar is also replicated against the surface gauge data set suggesting a regional influence. Correlations against the surface radar over the US are led by the DPR attaining a correlation of 0.61. Similar to Western Europe, GMI has the best correlation of the radiometer products, declining with AMSR2, SSMIS, MHS and finally, ATMS. Correlations with the surface gauges are varied, and much lower than those obtained over the WEGN network: this is likely due to the smaller number of available matchups in the comparison leading to greater variability in the statistical quantities.

The POD, FAR and HSS scores over the US are better than those over Western Europe. Against the surface radar the DPR retrievals once again have low FAR leading to a high HSS (0.77). While the radiometers have equal or higher POD, they have high FAR scores, thus reducing their overall HSS scores. All of the HSS scores, for all sensors, against the surface radar are higher over the US the HSS scores over Western Europe; this is in contrast to the surface gauge comparisons where the HSS scores are lower.

3.5. Summary

Overall, over Western Europe the satellite products tend to underestimate the precipitation with respect to the surface radar and gauge data sets. Over the US the satellite products have little, or slight positive bias when compared against the surface radar although comparison with the gauge data is less conclusive. Correlations are generally good, with the DPR being good (0.49 over Europe and 0.61 over the US). The DPR-Ku has the best HSS scores resulting from low FAR scores, whereas the radiometers are hampered by their higher FAR scores. The GMI appears to perform best amongst the radiometers, not least since it utilises information from the DPR in the generation of the *a priori* database. Statistical scores for other radiometers tend to be poorer.

4. Discussion

This paper has assessed a collection of instantaneous Level 2 V05 GPROF and DPR products from the GPM constellation and surface reference data over both Europe and the US. The results have highlighted differences in the distribution of occurrence and accumulation of precipitation by intensity of the GPROF retrievals with respect to the surface radar, but similarities between the DPR-Ku and surface radar (Figures 2 and 4). Over the European region a consistent, regular seasonal cycle in performance is clearly observed while over the US, although this cycle is less clear, an evident hierarchy is seen in the performance of the retrievals by sensor type (Figure 8). These results are supported by the statistical analysis (Figures 9 and 10).

The comparison of satellite measurements with surface reference data sets is not necessarily straightforward and entails the consideration of a multitude of factors. Many of these factors become increasingly important when dealing fine scale measurements at full resolution and instantaneous scales (Kirstetter et al. 2015a,b). Improvements in the spatial resolution of passive microwave satellite observations together with the pressing need for finer-scale temporal observations has highlighted the need for the improved representation of both the surface and satellite estimates. Historically the low-resolution of the satellite observations and subsequent products has 'helped' the comparison with ground validation products since the scale of the spatial variability of the precipitation is within the resolution of the satellite observation; satellite products available in the order of 50 km x 50 km capture the general regional-scale precipitation features that are also captured in radar products of similar resolutions. However, increasing the resolution to 25 km x 25 km, to 10 km x 10 km (IMERG-like resolution), or to a more ideal resolution of 5 km x 5 km can produce very different results due to the fine-scale structure of precipitation systems. Comparison of fine-scale precipitation products also highlights differences in the relative position of 'surface precipitation' between that of the surface radar estimate and that of the satellite estimate.

Taking the surface radar measurement (or estimate) of precipitation first, one of the most significant factor in what the measurement represents (assuming the physics of the measurement to be 'true') is the height of the beam above the surface. To avoid backscatter from the radar

locality it is necessary to elevate the beam above the surface so that as the beam propagates away from the radar location the height above the ground level increases. While this reduces/avoids anomalous propagation and beam blockage errors, it means that the radar beam no longer measures precipitation at or close to the surface. This becomes increasingly problematic not only with increasing range but also with shallow precipitation where the radar beam might miss the underlying precipitation altogether. Furthermore, where the beam does encounter precipitation at altitude, it is not certain that this precipitation is the same as that at the surface since evaporation might occur (visible as virga), or even low level-enhancement (e.g. seeder-feeder mechanisms) (Kirstetter et al. 2013a).. There is also a time-impact since if precipitation is seen by the radar at 4 km altitude, assuming moderate intensity liquid precipitation, this would take about 10 minutes to reach the ground. Thus surface measurements will not necessarily be co-temporally matched with the radar measurement.

Satellite measurements are different but no less complex. Although passive microwave observations are essentially path-integral measurements, the spectral frequency of the sensor used for of the observation determines, to a large extent, the type of hydrometeor being observed and therefore the height at which the precipitation is being measured. For example, the higher frequency channels are more sensitive to ice particles, and therefore essentially measure the ice content of the precipitating system, while the lower frequencies measure liquid particles and therefore precipitation below the freezing level. Thus, different techniques are likely to provide

measures of precipitation that relate to different precipitation particle species. This is particularly true for over land versus ocean precipitation estimates; over land the higher frequency channels tend to be used (or have a greater weighting), while over the ocean the lower frequency channels are used. The situation is further complicated since the observations made by passive microwave sensors are usually along a slant path; for conically scanning radiometers this tends to be constant at about 52-53° (Earth Incidence Angle; EIA), while for cross-track sensors this varies from 0° at nadir to about 58° at scan edge. Thus, due to parallax, the high-frequency channel retrievals will locate the precipitation closer to the sub-satellite location (due to sensitivity to higher-altitude ice particles) than the lower frequency channel retrievals (with lower-altitude liquid precipitation). As noted above, with coarse resolutions this parallax is contained within the footprint size of the retrieval, but with improvements in resolution to those in the order of 5 km x 5 km parallax can account for at least 1 pixel shift in position, thus impacting the final retrieval and subsequent validation with surface datasets. Finally when the high frequency channels are used as the primary source of the estimator (as is the case over land) the satellite estimate may not be representative of the surface precipitation due to modification of the precipitation column as it falls to the ground. The decoupling of the rainfall process from the ice process leads to an elapsed time between the observed measurement of the satellite and the precipitation reaching the ground.

One compromise that has been made here is that all the comparisons have assumed a common spatial resolution of 15 km. It should be noted that there are differences in the resolutions of the observations and the retrieval resolutions of each sensor. While the GMI and AMSR2 have retrieval resolutions close to 15 km x 15 km, the SSMIS has a much poorer resolution (see Table 1). However, the SSMIS sampling is every 12.5 km both along track and along scan; thus there is a good degree of oversampling which users should be aware of. The MHS and ATMS, being cross-track sensors, have varying resolutions with scan position: at nadir the spatial resolutions are close to 15 km x 15 km, but degrade towards the edge of swath. Although Table 1 notes the nominal retrieval resolutions, these are further complicated by the weightings given to each frequency in the retrieval: over ocean the lower frequency channels are weighted more, while over land the higher frequency channels have greater weight. Thus, although the nominal retrieval resolution for the GMI, AMSR2 and SSMIS is that of the 18.7/19.35 GHz, over land the higher weighting of the 89/91 GHz channels actually provides a finer resolution. In addition, if ice is present the contribution of the higher (ice-sensitive) channels will lead to finer scale precipitation features in the retrievals than for a purely liquid precipitation event.

This study has exploited high-quality, large scale radar surface precipitation measurements to validate the satellite estimates, together with spatially-dense gauge networks. It is encouraging that the both the scatterplots of satellite-radar (Figures 3 and 5) and satellite-gauge are similar (Figures 6 and 7), as are the statistics and metrics (shown in Figures 9 and 10). Although

differences exist in the surface reference data sets both the surface radar and gauge networks have been carefully developed to ensure high quality data products. The US MRMS data has undergone extensive quality assurance and benefits from a largely homogeneous network, and while the Nimrod European radar product is a composite of radars from different operations systems, care has been taken to integrate these into one product. The two gauge networks used here are independent from gauges used to adjust and calibrate radar values; the density of the gauges and comparatively high time resolutions (time of tip for PCMK, 5 min accumulation for WEGN) allows representative match-ups the satellite pixels. This consistency in the results between the satellite products and these independent data sets suggests that there are regional biases within the satellite products; the seasonal dependency over the European region also suggests a dependence on the meteorological/precipitation regime.

The detail of this study also highlights the hierarchy of the satellite retrievals from the GPM sensors, and in particular of the derivation of the databases used within the GPROF retrieval scheme. The GPM DPR (through the DPR/GMI combined product) forms the basis of the GMI databases. The profiles within these databases are then used to simulate the profiles to populate the other sensor databases through the modelling of Tbs at the appropriate sensor frequencies and resolutions. Clearly, the databases of sensors with similar characteristics to that of the GMI would also be similar compared with sensors with greater differences (i.e. frequency or resolution); there is a loss of fidelity between the original database and subsequent databases.

This is evident from the results shown here: first the general form of the distribution of the occurrence and accumulation by intensities for the GMI is also reflected in that of the other sensors, and second; the statistics and metrics generally show that the GMI product performs best, with the overall decline in performance of the other sensors commensurate with the dis-similarities to the GMI sensor. Underlying physical reasons for differences in performance should also be considered. For example, Kidd et al (2017) noted that the METOPA and METOPB MHS retrievals outperform the retrievals made by identical and inter-calibrated instruments on the NOAA18 and NOAA19 satellites, which was attributed to the later overpass times of the MetOpA and MetOpB satellites and thus observing more organized precipitation systems.

5. Conclusion

High quality surface reference data sets over Europe and the US based upon surface radar and gauge networks have been used to validate the latest GPM V05 precipitation products from the GMI, AMSR2, SSMIS, MHS and ATMS radiometers together with the DPR-Ku precipitation product. The results show generally good agreement between the satellite products and the surface data sets. Systematic differences between the seasonal performance of the products between the US and the European regions were noted and will require further investigation. However, the results were consistent between satellite-radar and satellite-gauge comparisons suggesting more fundamental underlying issues relating to the identification and retrieval of

precipitation under differing meteorological regimes. Further detailed analysis is necessary to fine tune such comparisons and ascertain the sources of discrepancies such as effective retrieval resolution, parallax and variations in hydrometeor characteristics.

Past inter-comparison studies have shown that despite the multitude of precipitation-related observations and accompanying retrievals and products, it is often difficult to ascribe any single precipitation estimate to be better than any other estimate (Adler et al. 2001). This study has shown that high quality surface data can be used to provide insights into both the retrieval scheme and regional variations. However, it should be noted that this study has concentrated upon the European and US regions, and used different reference datasets from those regions; the performance of these products over other regions and seasons with unique precipitation regimes is likely to different.

Critically, the requirement of the actual precipitation product is dependent upon the final user; for the hydrologist and water resource manager the precipitation product should represent the best available estimate at the surface for a particular location, conversely, for atmospheric scientists a path-integrated value might be more appropriate whose ground location might not necessarily be the same as that prescribed for the hydrologist.

6. Acknowledgments

Chris Kidd is supported through the NASA Precipitation Measurement Missions (PMM) award NNX16AE22G and the NASA Collaborative Earth System Science Research between NASA/GSFC and University of Maryland, award NNX12AD03A. Jackson Tan is supported by an appointment to the NASA Postdoctoral Program at Goddard Space Flight Center, administered by USRA contract NNH15CO48B with NASA. Walt Petersen acknowledges support provided through the NASA GPM and PMM Programs. Pierre Kirstetter is supported through the NASA Precipitation Measurement Missions (PMM) award NNX16AE39G and the NASA Ground Validation Program award NNX16AL23G

We acknowledge the NASA/GSFC Precipitation Processing System (PPS) for the provision of the GPM and constellation precipitation products (*pmm.nasa.gov*), and the wider NASA PMM/GPM team. Finally, we acknowledge G. Kirchengast and his team at the Wegener Center (Univ. of Graz, Austria) for operating the WegenerNet high-resolution station network, sharing its data also into the NASA GPM/PMM ground validation team and valuable advice; detailed information on the network is found at *www.wegcenter.at/wegenernet* and the data are available through the WegenerNet data portal *www.wegenernet.org* .

7. References

Adler RF, Kidd C, Petty G, Morissey M, Goodman HM. 2001. Intercomparison of Global Precipitation Products: The Third Precipitation Intercomparison Project (PIP–3). *Bull. Amer. Meteor. Soc.* 82: 1377–1396. DOI: 10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2

Berg W, L'Ecuyer T, Kummerow C. 2006. Rainfall Climate Regimes: The relationship of regional TRMM rainfall biases in the environment. *J. Appl. Meteor. Climatol.* 45: 434-454. DOI: 10.1175/JAM2331.1

Berg W, Bilanow S, Chen RY, Datta S, Draper D, Ebrahimi H, Farrar S, Jones WL, Kroodsma R, McKague D, Payne V, Wang J, Wilheit T, Yang JX. Intercalibration of the GPM Microwave Radiometer Constellation. Journal of Atmospheric and Oceanic Technology 33, 2639-2654 DOI: 10.1175/JTECH-D-16-0100.1

Carr N, Kirstetter PE, Hong Y, Gourley JJ, Schwaller M, Petersen W, Wang N-Y, Ferraro RR, Xue X. 2015. The Influence of Surface and Precipitation Characteristics on TRMM Microwave Imager Rainfall Retrieval Uncertainty. *J. Hydrometeor.* 16: 1596–1614. DOI:10.1175/JHM-D-14-0194.1.

Delahaye F, Kirstetter PE, Dubreuil V, Machado LAT, Vila DA. 2015. A consistent gauge database for daily rainfall analysis over the Legal Brazilian Amazon. *J. Hydrology ,*527: 292-304. DOI: 10.1002/qj.2416

Ebert EE, Janowiak JE, Kidd C. 2007. Comparison of Near Real Time Precipitation Estimates from Satellite Observations and Numerical Models. *Bull. Amer. Meteor. Soc.* 88: 47-64. DOI: 10.1175/BAMS-88-1-47

Fairman JG, Schultz DM, Kirschbuam DJ, Gray SL, Barrett AI, (Submitted). Climatology of size, shape and intensity of precipitation features over Great Britain and Ireland. Submitted to *J. Hydrometeor.* 8 September 2016

Fairman JG, Schultz DM, Kirshbaum DJ, Gray SL, Barrett AI, 2015. A radar-based rainfall climatology of Great Britain and Ireland. *Weather,* **70,** 153–158, DOI: 10.1002/wea.2486.

Harrison DL, Scovell RW, Kitchen M. 2009. High-resolution precipitation estimates for hydrological uses. *Water Management*162: 125-135. DOI: 10.1680/wama.2009.162.2.125

Harrison DL, Norman K, Pierce C, Gaussiat N. 2012. Radar products for hydrological applications in the UK. *Water Management*165: 89-103. DOI: 10.1680/wama.2012.165.2.89

Hou AY, Kakar RK, Neeck S, Azarbarzin AA, Kummerow CD, Kojima M, Oki R, Nakamura K, Iguchi T. 2014. The Global Precipitation Measurement (GPM) Mission. *Bull. Amer. Meteor. Soc.* **95**: 701-722. DOI:10.1175/BAMS-D-13-00164.1.

Houze R, McMurdie L, Petersen W, Schwaller M, Baccus W, Lundquist J, Mass C, Nijssen B, Rutledge S, Hudak D, Tanelli S, Mace J, Poellot M, Lettenmaier D, Zagrodnik J, Rowe A, DeHart J, Madaus L, Barnes H. 2017. The Olympic Mountains Experiment (OLYMPEX). *Bull. Amer. Meteor. Soc.* doi:10.1175/BAMS-D-16-0182.1, in press.

Iguchi T, Seto S, Awaka J, Meneghini R, Kubota T, Chandra V, Yoshida N, Kawamoto N and Oki R. 2016. Precipitation rates estimated with GPM's dual-frequency radar. IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IGARSS 2016. 3917-3918. DOI: 10.1109/IGARSS.2016.7730017

Jensen MP, Petersen WA, Bansemer A, Bharadwaj N, Carey LD, Cecil DJ, Collis SM, Del Genio AD, Dolan B, Gerlach J, Giangrande SE, Heymsfield A, Heymsfield G, Kollias P, Lang TJ, Nesbitt SW, Neumann A, Poellot M, Rutledge SA, Schwaller M, Tokay A, Williams CR, Wolff DB, Xie S, Zipser EJ. 2016. The Midlatitude Continental Convective Clouds Experiment (MC3E). *Bull. Amer. Meteorol*. Soc. 97: 1667-1686. DOI:10.1175/BAMS-D-14-00228.1

Joyce RJ, Janowiak JE, Arkin PA, Xie PP. 2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeor.* 5: 487-503. DOI: 10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Joyce RJ, Xie PP. 2011. Kalman Filter-Based CMORPH. *J. Hydrometeor.* 12: 1547-1563. DOI: 10.1175/JHM-D-11-022.1

Kidd C, Levizzani V. 2011. Satellite precipitation retrievals. *Hydrology Earth System Science Discussions* 7: 8157-8177. DOI: 10.5194/hessd-8-8157-2010

Kidd C, Bauer P, Turk J, Huffman GJ, Joyce R, Hsu K-L, Braithwaite D. 2012. Inter-comparison of high-resolution precipitation products over northwest Europe. *J. Hydrometeor.* 13: 67-83. DOI: 10.1175/JHM-D-11-042

Kidd C, Matsui T, Chern J, Mohr K, Kummerow C, Randel D. 2015. Precipitation Estimates from cross-track passive microwave observations using a physically based retrieval scheme. *J. Hydrometeor.* 17: 383-400. DOI: 10.1175/JHM-D-15-0051.1

Kidd C, Petersen WA, Skofronick-Jackson G, Huffman GJ. 2016. Validation of Precipitation Products from the Global Precipitation Measurement mission. EUMETSAT Meteorological Satellite conference, Darmstadt, Germany, 26-30 September 2016.

Kidd C, Huffman G, Becker A, Skofronick-Jackson G, Kirschbaum D, Joe P, Muller C. 2017. So, how much of the Earth's surface *is* covered by rain gauges? *Bull. Amer. Meteor. Soc.* 98: 69-78. DOI: http://dx.doi.org/10.1175/BAMS-D-14-00283.1

Kirchengast G, Kabas T, Leuprecht A, Bichler C, Truhetz H. 2014. WegenerNet: A pioneering high-resolution network for monitoring weather and climate, *Bull. Amer. Meteor. Soc.* 95: 227–242. DOI:10.1175/BAMS-D-11-00161.1.

Kirstetter PE, Hong Y, Gourley JJ, Chen S, Flamig Z, Zhang J, Schwaller M, Petersen W, Amitai E. 2012. Toward a Framework for Systematic Error Modeling of Spaceborne Precipitation Radar with NOAA/NSSL Ground Radar-based National Mosaic QPE. *J. Hydrometeor.* 13: 1285-1300. DOI: 10.1175/JHM-D-11-0139.1

Kirstetter PE, Andrieu H, Boudevillain B**,** Delrieu G. 2013a. A physically-based identification of vertical profiles of reflectivity from volume scan radar data. *J. Appl. Meteor. Climatol.* 52: 1645-1663. DOI: 10.1175/JAMC-D-12-0228.1

Kirstetter PE, Hong Y, Gourley JJ, Schwaller M, Petersen W, Zhang J. 2013b. Comparison of TRMM 2A25 Products Version 6 and Version 7 with NOAA/NSSL Ground Radar-based National Mosaic QPE. *J. Hydrometeor.* 14: 661-669. DOI:10.1175/JHM-D-12-030.1

Kirstetter PE, Hong Y, Gourley JJ, Cao Q, Schwaller M, Petersen W. 2014. A research framework to bridge from the Global Precipitation Measurement mission core satellite to the constellation sensors using ground radar-based National Mosaic QPE. In L. Venkataraman, Remote Sensing of the Terrestrial Water Cycle. AGU books Geophysical Monograph Series, Chapman monograph on remote sensing. John Wiley & Sons Inc. ISBN: 1118872037.

Kirstetter PE, Gourley JJ, Hong Y, Zhang J, Moazamigoodarzi S, Langston C, Arthur A. 2015a. Probabilistic Precipitation Rate Estimates with Ground-based Radar Networks. *Water Resources Research* 51: 1422–1442. DOI:10.1002/2014WR015672

Kirstetter PE, Hong Y, Gourley JJ, Schwaller M, Petersen W, Cao Q. 2015b. Impact of sub-pixel rainfall variability on spaceborne precipitation estimation: evaluating the TRMM 2A25 product. *Quarterly Journal of the Royal Meteorological Society* 141: 953–966. DOI: 10.1002/qj.2416

Kummerow C, Hong Y, Olson WS, Yang S, Alder RF, McCollum J, Ferraro R, Petty G, Shin D-B, Wilheit TT. 2001. The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor,* 40: 1801-1820.

Kummerow CD, Randel DL, Kulie M, Wang N-Y, Ferraro R, Munchak SJ, Petkovic V. 2015. The Evolution of the Goddard PROFiling Algorithm to a Fully Parametric Scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280. DOI: 10.1175/JTECH-D-15-0039.1

Met Office (2003): 5 km Resolution Europe Composite Resolution Rainfall Data from the Met Office Nimrod System. NCAS British Atmospheric Data Centre, *last accessed 20 June 2017*. http://catalogue.ceda.ac.uk/uuid/d5ae8b92d8c884690592ce619f2eca07

Olson W, Yang SS, Stout JE, Grecu M. 2007. The Goddard Profiling Algorithm (GPROF): description and current applications, in Measuring Precipitation from Space: EURAINSAT and the future. Eds. Levizzani, V; Bauer, P; Turk, FJ, *Adv. Global Change Res.* 28: 179

Petersen WA, Houze RA, McMurdie L, Zagrodnik J, Tanelli S, Lundquist J, Wurmann J. 2016. The Olympic Mountains Experiment (OLMPEX): From Ocean to Summit. Meteorol. Tech. Intnl., Vol. Sept. 2016, 22-26.

Roca R, Chambon P, Jobard I, Kirstetter PE, Gosset M, Bergès JC. 2009. Comparing satellite and surface rainfall products over West Africa at meteorologically relevant scales during the AMMA campaign using error estimates. *J. Appl. Meteor. Climatol.* 49: 715–731. DOI:10.1175/2009JAMC2318.1

Skofronick-Jackson G, Hudak D, Petersen W, Nesbitt SW, Chandrasekar V, Durden S, Gleicher KJ, Huang G-J, Joe P, Kollias P, Reed KA, Schwaller M, Stewart R, Tanelli S, Tokay A, Wang JR, Wolde M. 2015: Global Precipitation Measurement Cold Season Precipitation Experiment (GCPEx): For measurement sake let it snow. *Bull. Amer. Meteorol. Soc*., 96**,** 1719-1741.

Skofronick-Jackson G, Huffman G, Petersen WA, Kirschbaum D, Kidd C. 2017. The Global Precipitation Measurement mission: Results after 2 years in orbit. *Accepted by the Bulletin of the American Meteorological Society.* DOI: 10.1175/BAMS-D-15-00306.1

Stephens GL, Vane DG, Boain, RJ, Mace, GG, Sassen, K, Wang ZE , Illingworth AJ, O'Connor EJ, Rossow WB, Durden SL, Miller SD, Austin RT, Benedetti A, Mitrescu C. 2002. The

cloudsat mission and the a-train - A new dimension of space-based observations of clouds and precipitation. *Bull. Amer. Meteor. Soc.*. 83: 1771-1790. DOI: 10.1175/BAMS-83-12-1771

Tan B-Z, Petersen WA, Tokay A. 2016. A Novel Approach to Identify Sources of Errors in IMERG for GPM Ground Validation. *J. Hydrometeor.* **17**: 2477-2491.

Tokay A, Roche RJ, Bashor PG. 2014. An experimental study of spatial variability of rainfall. *J. Hydrometeor.* 15: 801–812. DOI:10.1175/JHM-D-13-031.1.

Wentz FJ, Draper D. 2016. On-Orbit Absolute Calibration of the Global Precipitation Measurement Microwave Imager. *J. Atmos. Oceanic Technol.* **33**: 1393-1412, DOI: 10.1175/JTECH-D-15-0212.1.

Wilks DS. 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, pp 467

Yang H, Hsu K-L, Sorooshian S, Gao X. 2004. Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteor.* 43: 1834-1852. DOI: 10.1175/JAM2173.1

Zhang J, Howard K, Langston C, Kaney B, Qi Y, Tang L, Grams H, Wang Y, Cocks S, Martinaitis S, Arthur A, Cooper K, Brogden J, Kitzmiller D. 2016. Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. *Bull. Amer. Meteor. Soc.* 97: 621-638. DOI: 10.1175/BAMS-D-14-00174.1

Captions

Table 1. Characteristics and resolutions of the sensor used in this study together with their sampling and retrieval resolutions (km). Note the retrieval resolution relates to the along-scan vs across-scan dimensions.

Figure 1. Extent of radar coverage used in this study for (left) Western Europe and (right) eastern United States. The light grey-shaded region represents the areas excluded from the analysis to ensure good conditions for comparison. The locations of the well-calibrated dense gauge networks are also shown for Pocomoke (PCMK) and WegenerNet (WEGN).

Figure 2. Normalised distribution of the (left) occurrence and (right) accumulation of precipitation by intensity for GPROF and DPR Ku products over Western Europe.

Figure 3. Normalised density scatterplot of the V05 GPROF and DPR-Ku precipitation products versus surface radar data over the Western European region; all products are compared at a nominal resolution of 15 km x 15 km. (Note that zero values are plotted along the x and y axes).

Figure 4. Normalised distribution of the (left) occurrence and (right) accumulation of precipitation by intensity for GPROF and DPR Ku products over the Eastern United States.

Figure 5. Normalised density scatterplots of the V05 GPROF and DPR-Ku precipitation products versus surface radar data over the eastern United States region; all products are compared at a nominal resolution of 15 km x 15 km. (Note that zero values are plotted along the x and y axes).

Figure 6. Scatter plot of DPR-Ku precipitation retrievals against gauge measurements for the (a) Pocomoke (PCMK) and (b) WegenerNet (WEGN) gauge networks. Red crosses indicate stratiform precipitation and blue dots indicate convective precipitation, as identified by the satellite algorithm.

Figure 7. Scatter plot of all GPROF estimates from GMI, AMSR2, SSMIS, MHS and ATMS sensors versus gauge measurements for the (a) Pocomoke (PCMK) and (b) WegenerNet (WEGN) gauge networks. The colours indicate the density of points.

Figure 8. Monthly performance of instantaneous retrievals for the European region (upper) and the United States region (lower). Statistics have had a 3-month moving average applied.

Figure 9. Statistical performance of the retrievals over Western Europe for (top) the surface radar data comparisons and (bottom) the WegenerNet gauge network. Note that the DPR-Ku is excluded from the statistical analysis against the gauge data due to paucity of observations.

Figure 10. Statistical performance of the retrievals over the United States for (top) the surface radar data comparisons and (bottom) the Pocomoke gauge network. Note that the DPR-Ku is excluded from the statistical analysis against the gauge data due to paucity of observations.

Table 1. Characteristics and resolutions of the sensor used in this study together with their sampling and retrieval resolutions (km). Note the retrieval resolution relates to the along-scan vs across-scan dimensions.

Sensor	SSMIS	AMSR2	GMI	MHS	ATMS	DPR
Satellite	DMSP-F16, F17,F18,F19	GCOMW1	GPM	NOAA18,19, MetOp-A, B	NPP	GPM
Type	Conical	Conical	Conical	Cross-track	Cross-track	Cross-track
Frequencies	-	6.925/7.3VH	-	-	-	-
	-	10.65VH	10.65VH	-	-	13.60
	19.35VH	18.70VH	18.70VH	-	-	-
	22.235V	23.80VH	23.80V	-	23.8	-
	37.0VH	36.5VH	36.5VH	-	31.4	35.55
	50.3-63.3VH	-	-	-	50-3-57.3	-
	91.65VH	89.0VH	89.0VH	89V	87-91	-
	150H	-	165.6VH	157V	164-167	-
	183.31H	-	183.31V(2)	183.31H (2)	183.31(5)	-
	-	-		190.31V	-	-
Sampling resolution (km)	12.79 XT 12.59 AT	4.65 XT 4.28 AT	5.13 XT 13.19 AT	16.87 XT 17.62 AT	16.06 XT 17.74 AT	5.2 XT 5.2 AT
Retrieval resolution (km)	45x74 (19.35 GHz)	14x22 (18.7 GHz)	10.9x18.1 (18.7 GHz)	17.1x21.6 *at nadir*	16.5x16.2 *at nadir*	5.2x5.2 *at nadir*
FOVs per Scan line	180x	486x	221x	90x	96x	49x Ku 25x Ka

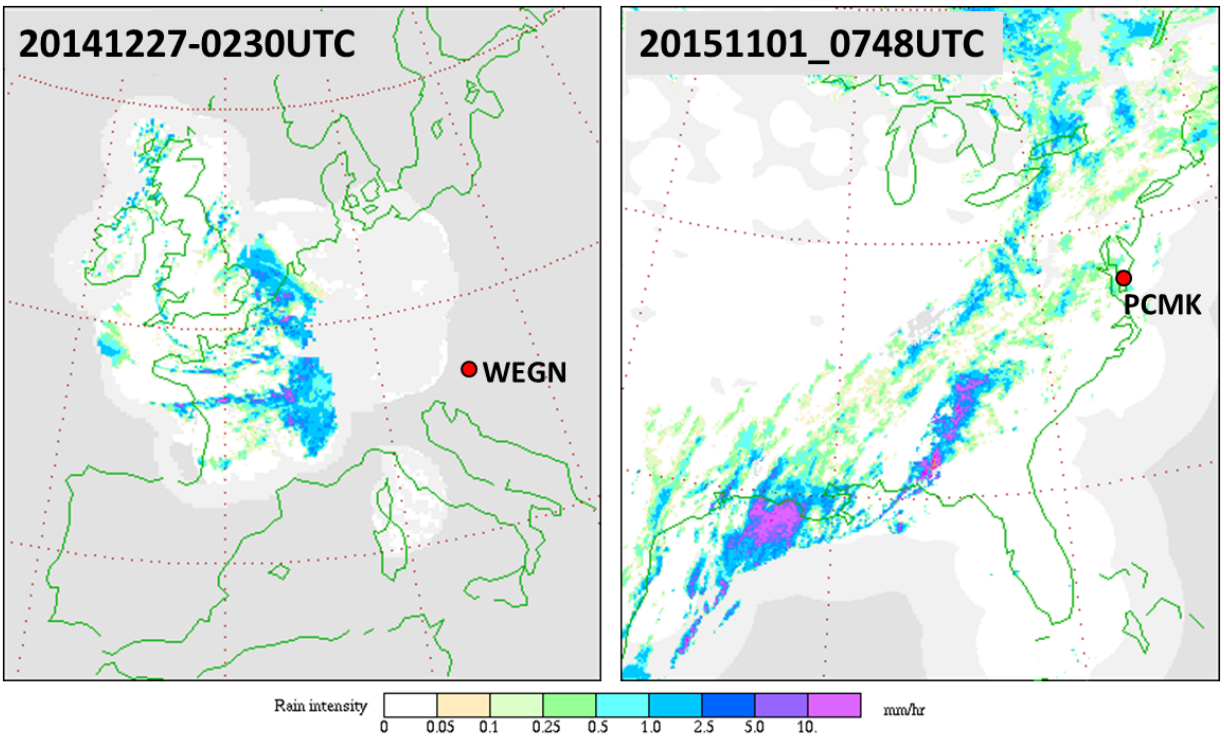


Figure 1. Extent of radar coverage used in this study for (left) Western Europe and (right) eastern United States. The light grey-shaded region represents the areas excluded from the analysis to ensure good conditions for comparison. The locations of the well-calibrated dense gauge networks are also shown for Pocomoke (PCMK) and WegenerNet (WEGN).

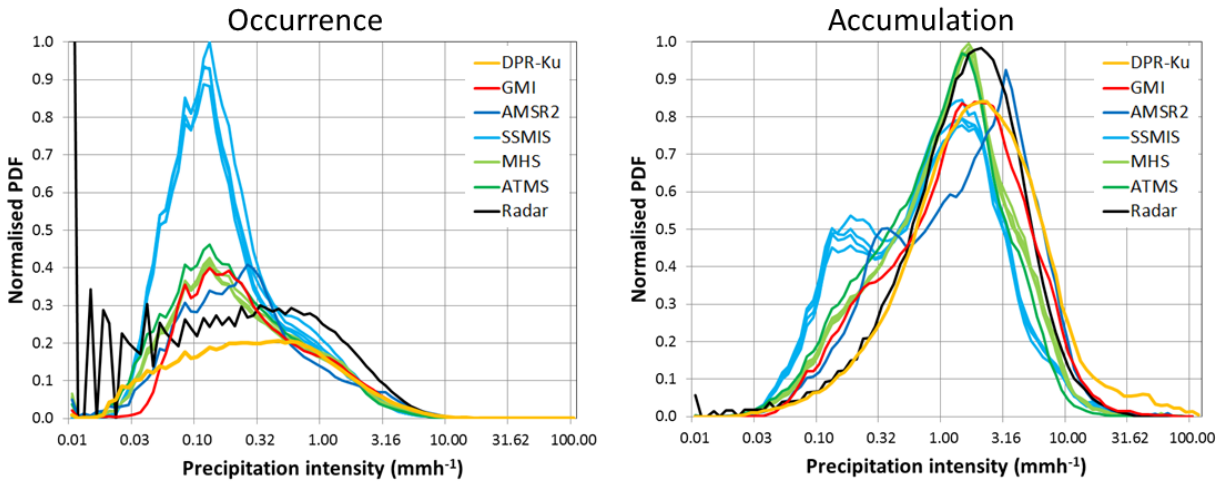


Figure 2. Normalised distribution of the (left) occurrence and (right) accumulation of precipitation by intensity for GPROF and DPR Ku products over Western Europe.

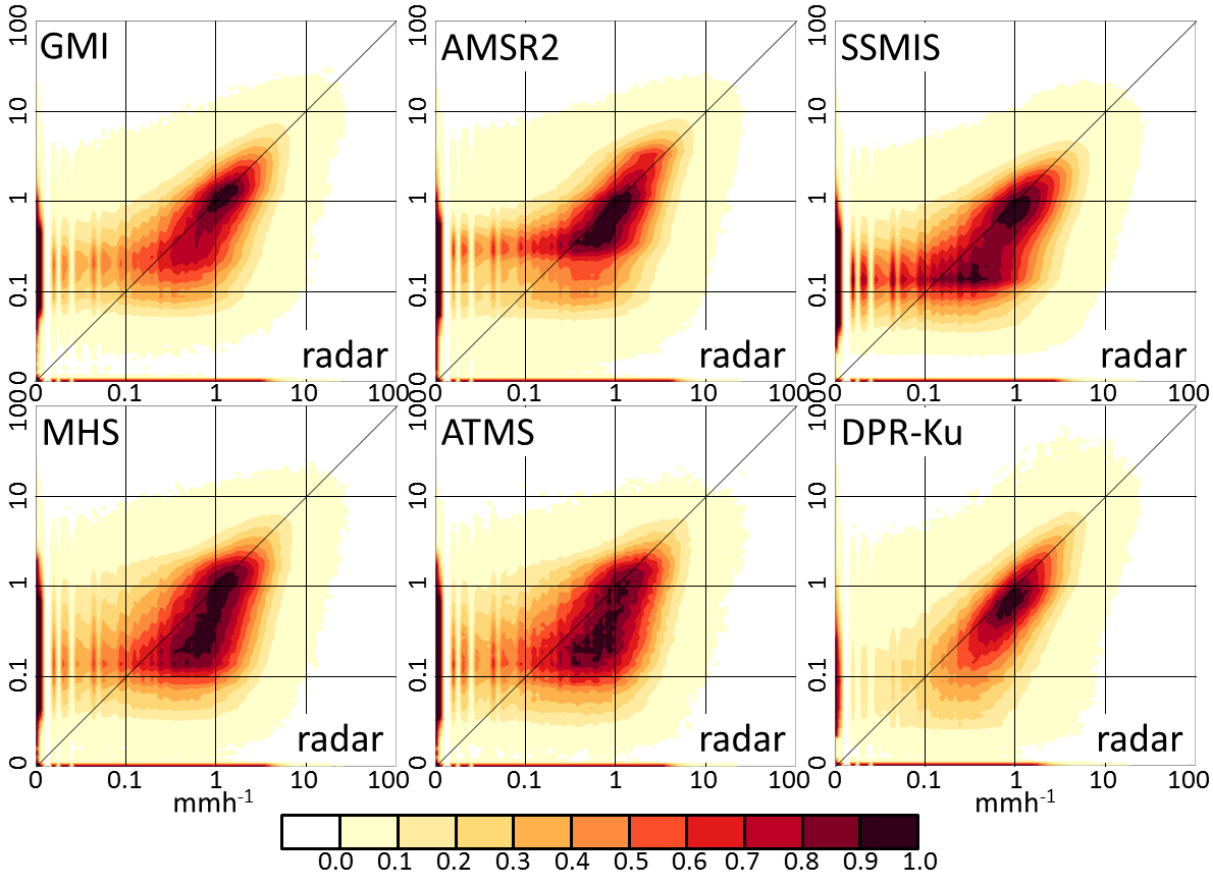


Figure 3. Normalised density scatterplot of the V05 GPROF and DPR-Ku precipitation products versus surface radar data over the Western European region; all products are compared at a nominal resolution of 15 km x 15 km. (Note that zero values are plotted along the x and y axes).

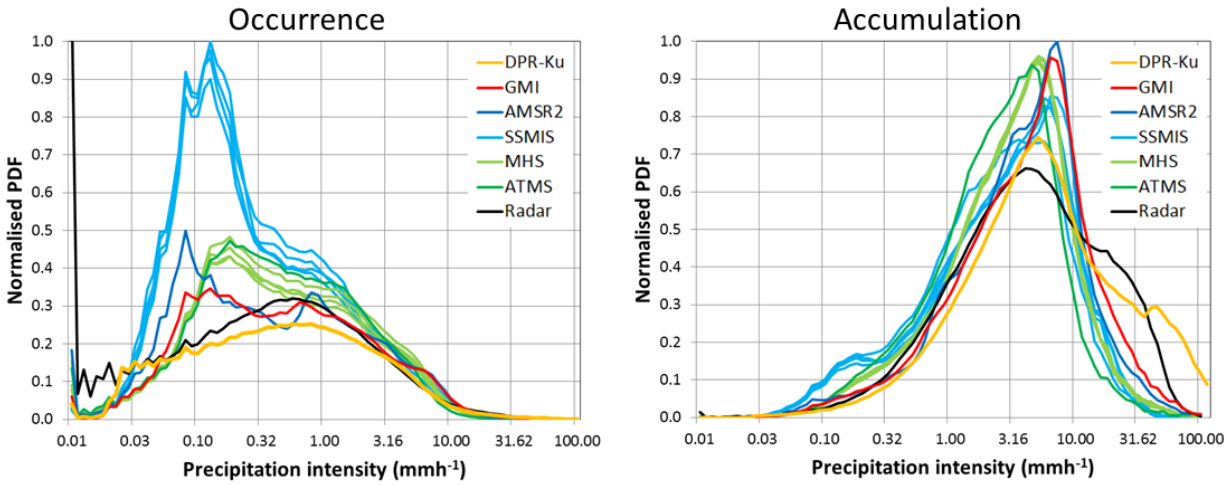


Figure 4. Normalised distribution of the (left) occurrence and (right) accumulation of precipitation by intensity for GPROF and DPR Ku products over the Eastern United States.

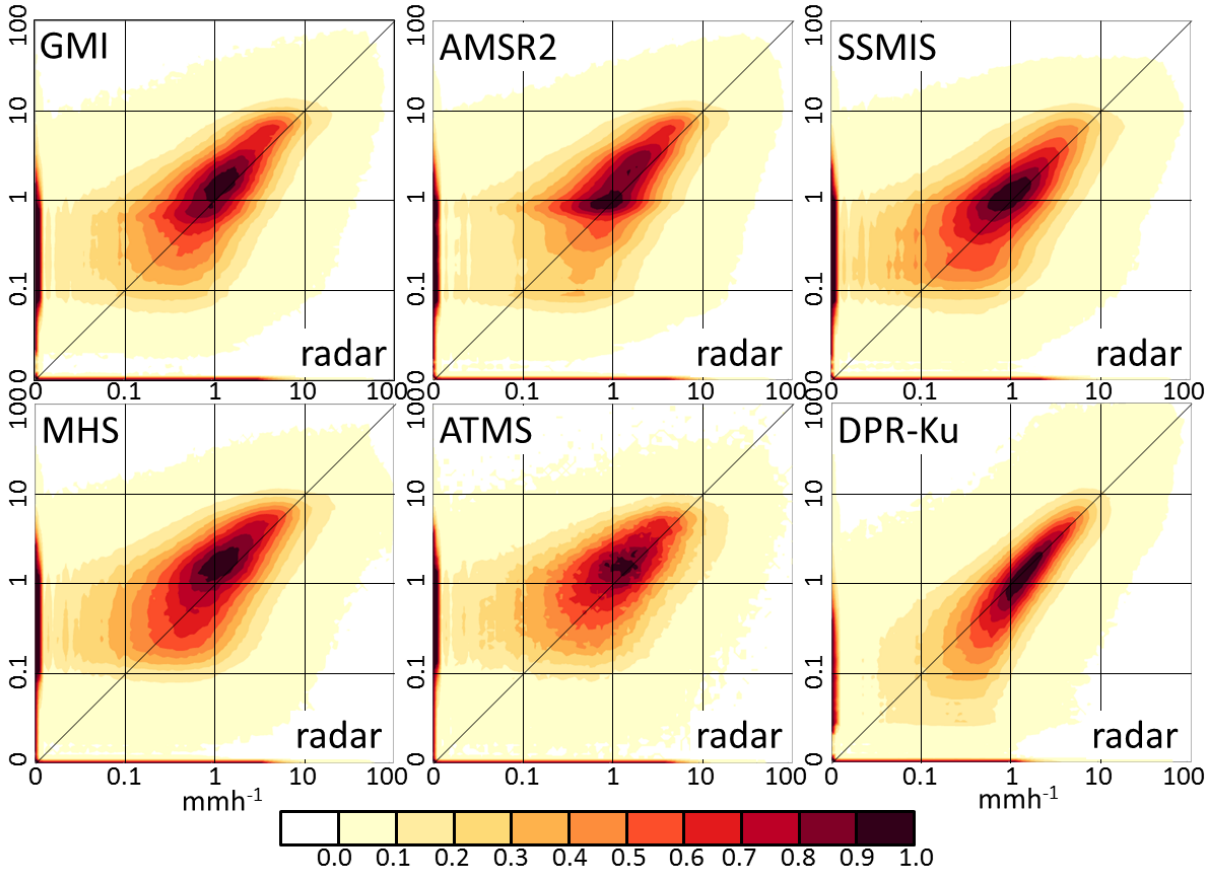


Figure 5. Normalised density scatterplots of the V05 GPROF and DPR-Ku precipitation products versus surface radar data over the eastern United States region; all products are compared at a nominal resolution of 15 km x 15 km. (Note that zero values are plotted along the x and y axes).

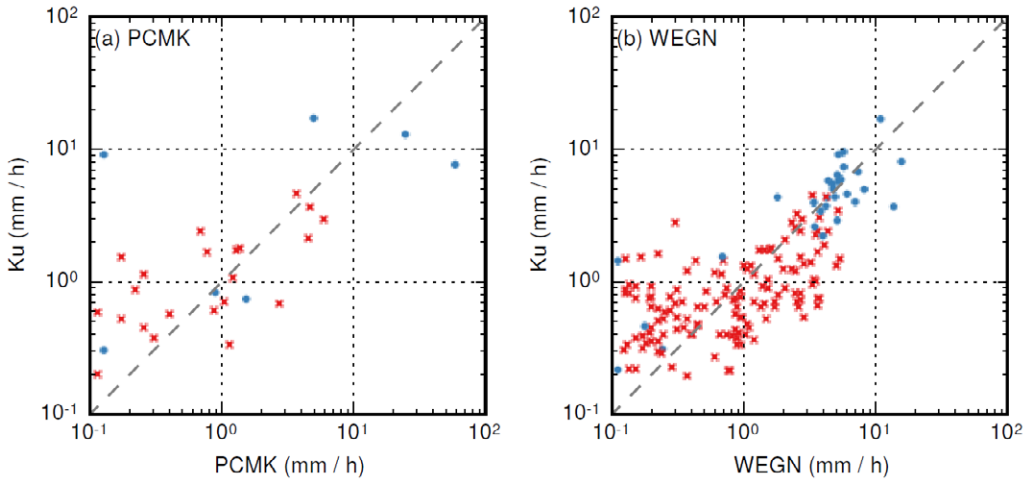


Figure 6. Scatter plot of DPR-Ku precipitation retrievals against gauge measurements for the (a) Pocomoke (PCMK) and (b) WegenerNet (WEGN) gauge networks. Red crosses indicate stratiform precipitation and blue dots indicate convective precipitation, as identified by the satellite algorithm.

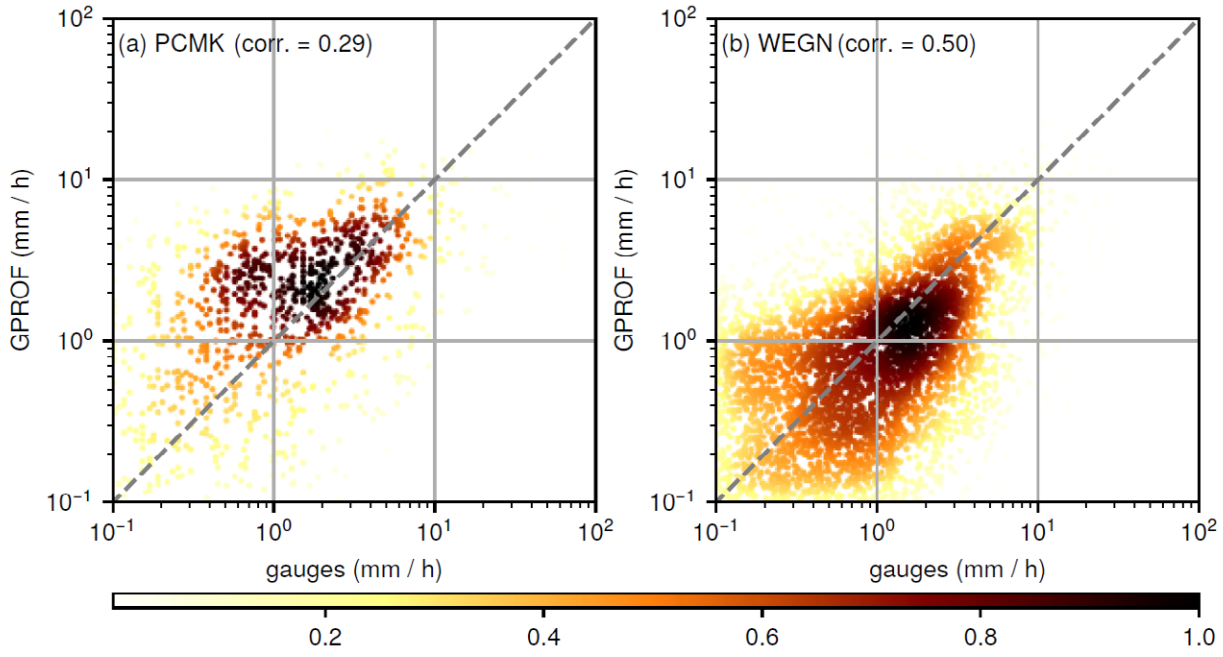


Figure 7. Scatter plot of all GPROF estimates from GMI, AMSR2, SSMIS, MHS and ATMS sensors versus gauge measurements for the (a) Pocomoke (PCMK) and (b) WegenerNet (WEGN) gauge networks. The colours indicate the density of points.

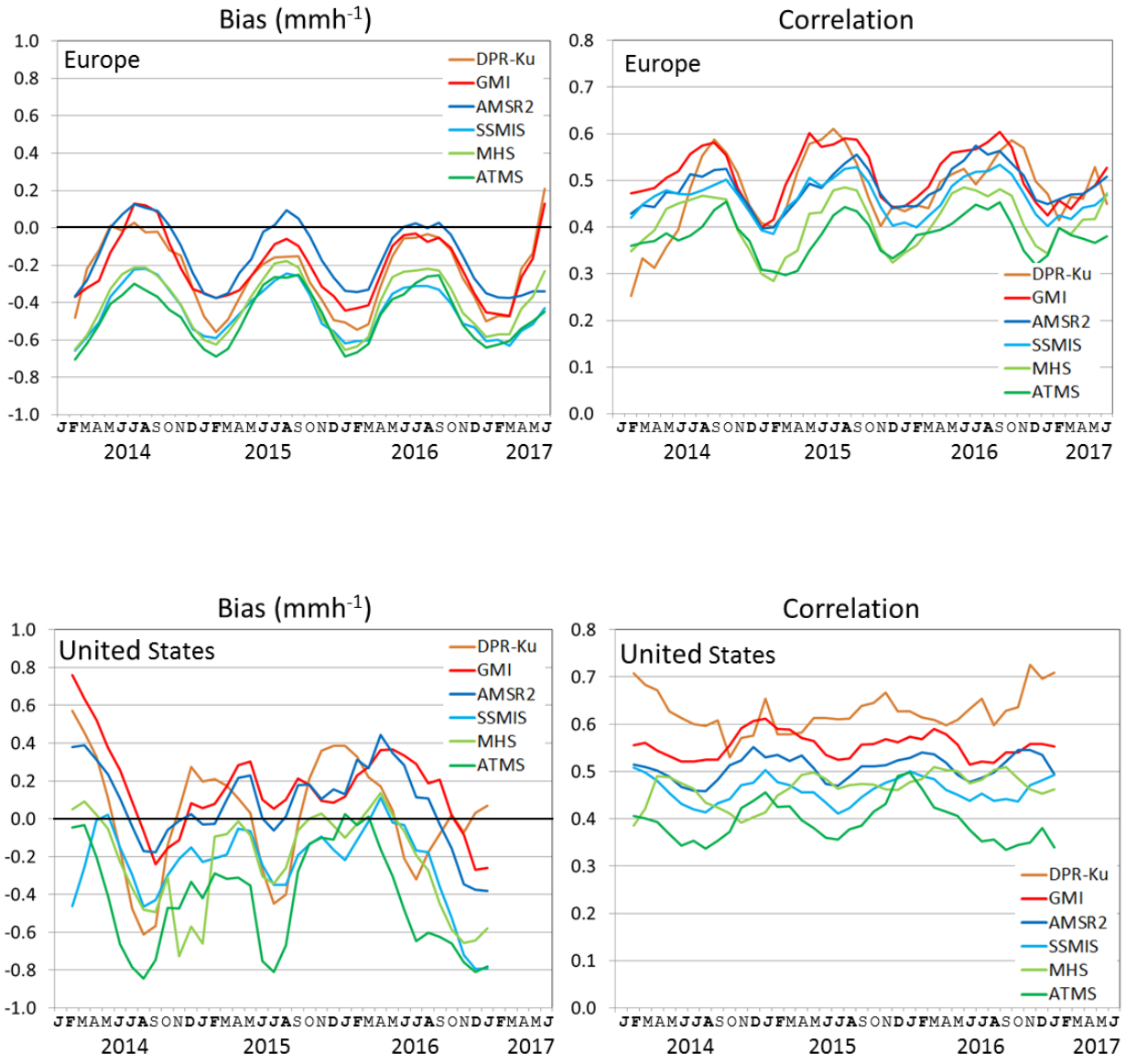
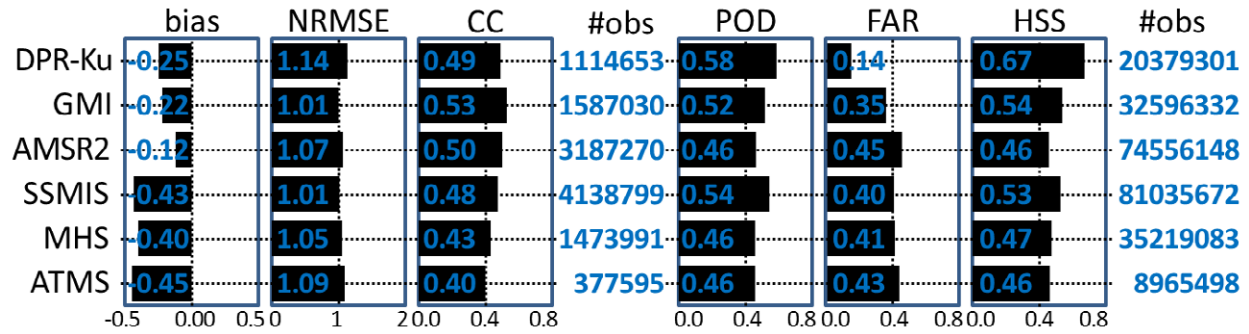


Figure 8. Monthly performance of instantaneous retrievals for the European region (upper) and the United States region (lower). Statistics have had a 3-month moving average applied.

Nimrod Radar Network



WegenerNet Gauge Network

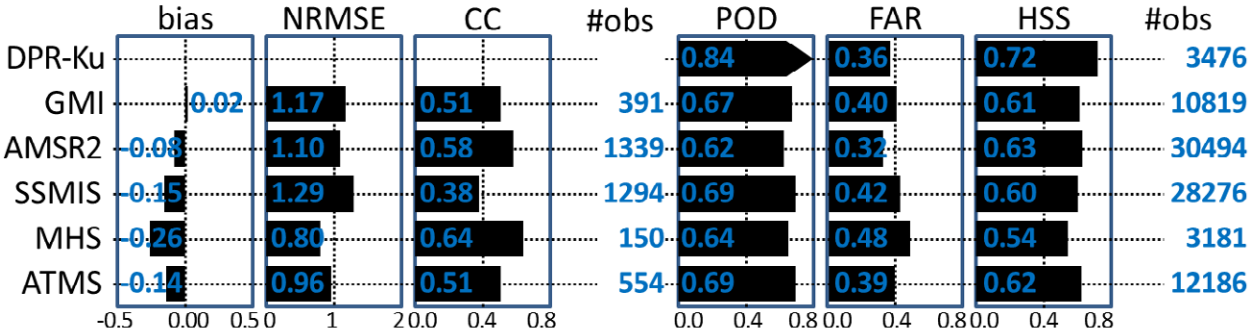
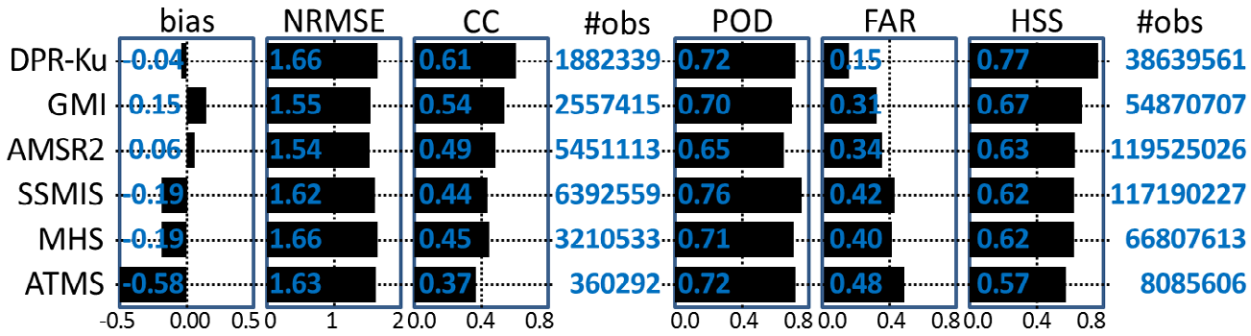


Figure 9. Statistical performance of the retrievals over Western Europe for (top) the surface radar data comparisons and (bottom) the WegenerNet gauge network. Note that the DPR-Ku is excluded from the statistical analysis against the gauge data due to paucity of observations.

MRMS Radar Network



Pocomoke Gauge Network

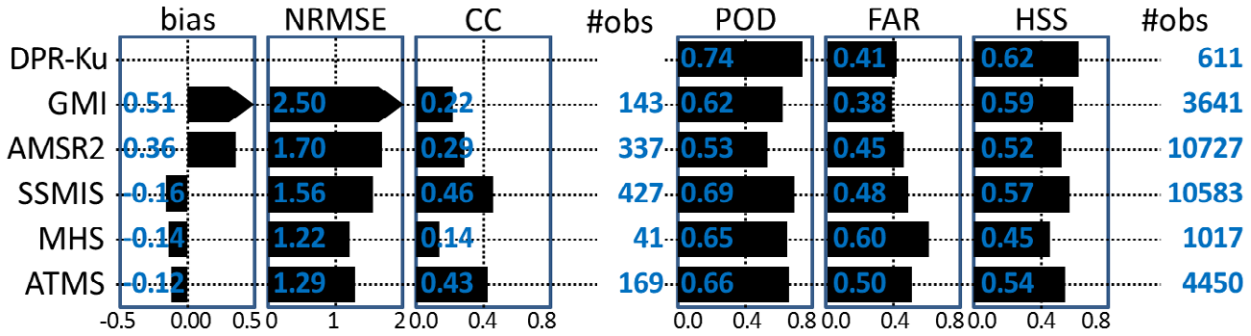
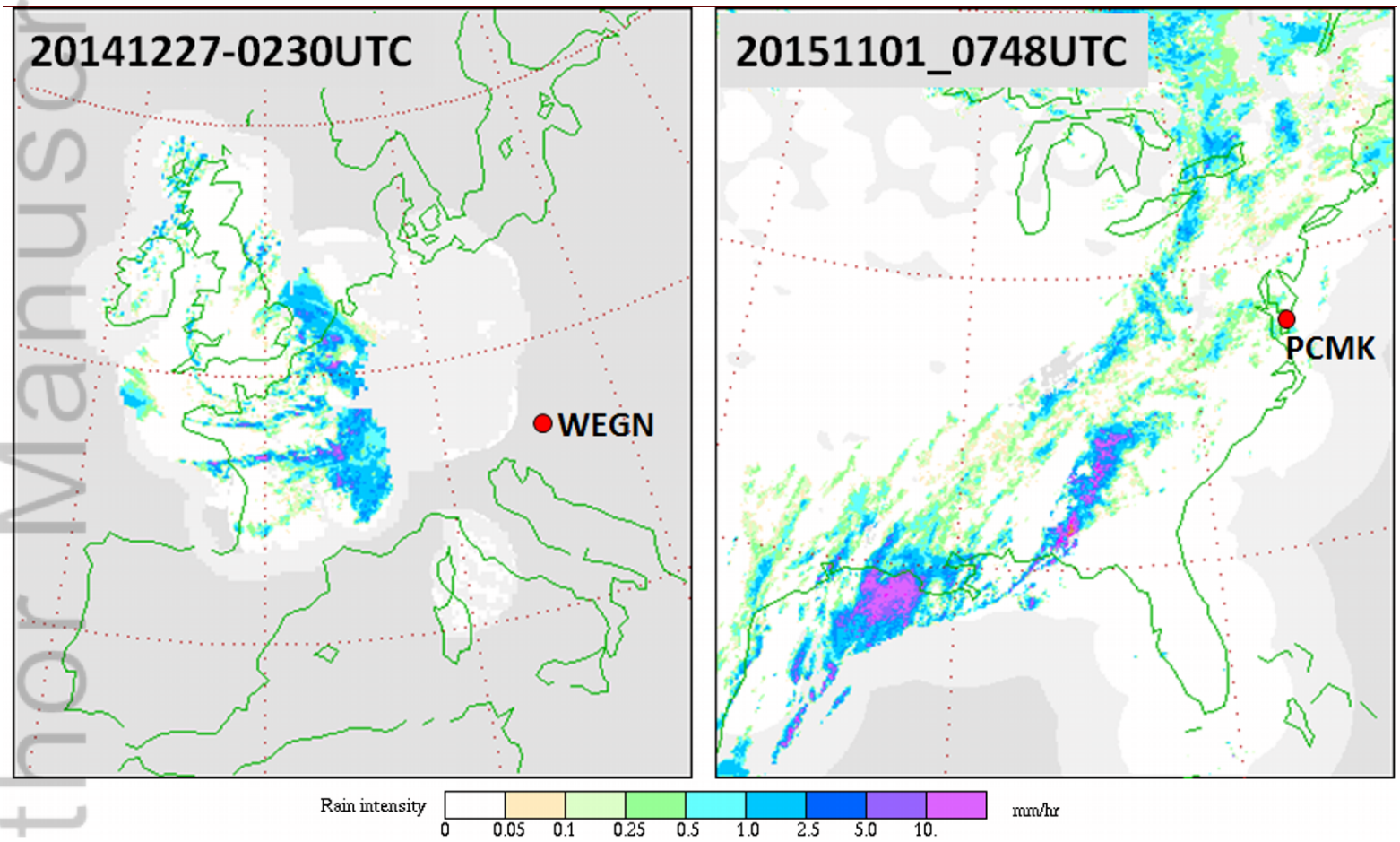
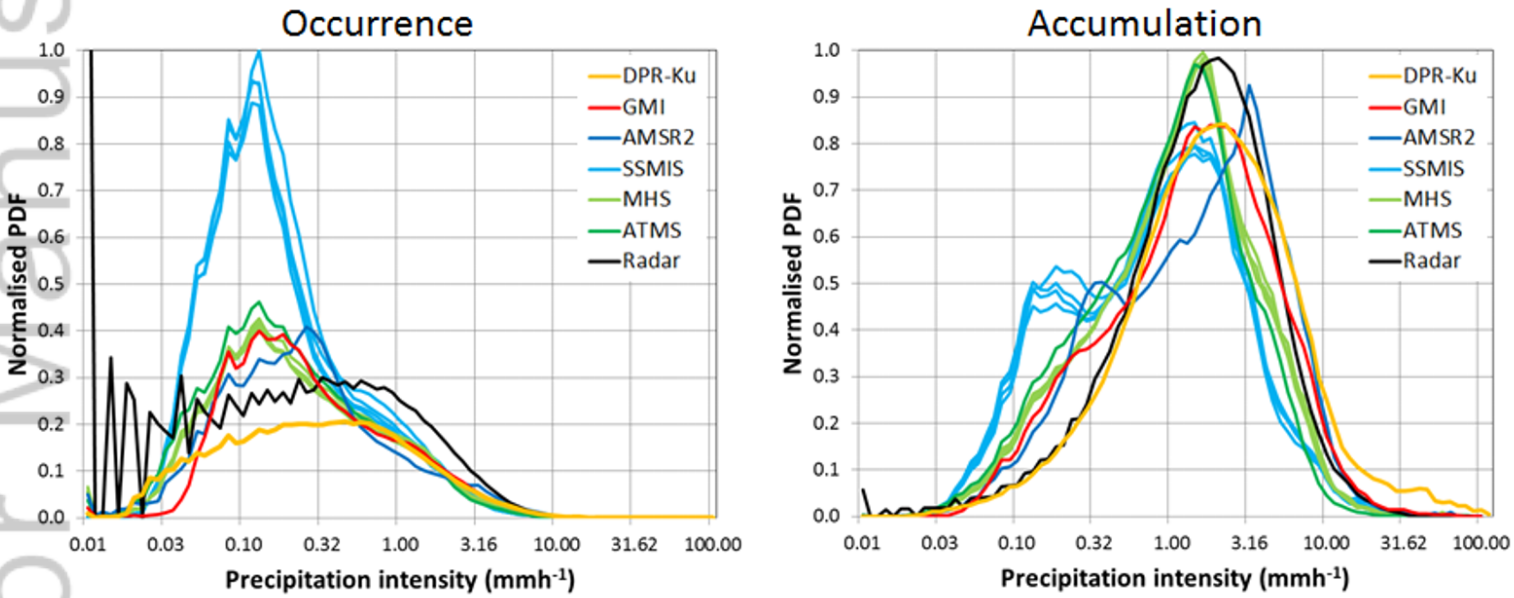


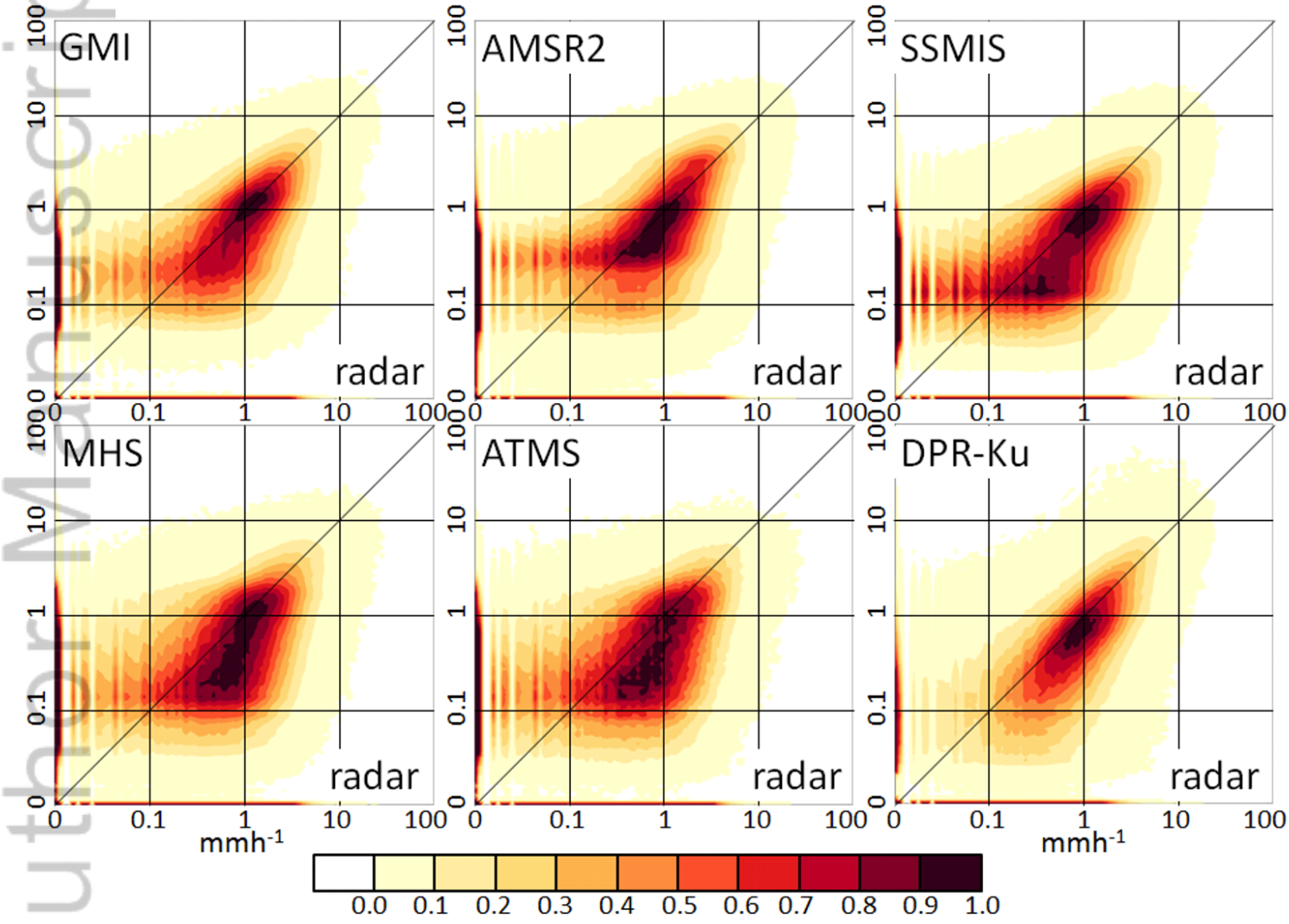
Figure 10. Statistical performance of the retrievals over the United States for (top) the surface radar data comparisons and (bottom) the Pocomoke gauge network. Note that the DPR-Ku is excluded from the statistical analysis against the gauge data due to paucity of observations.



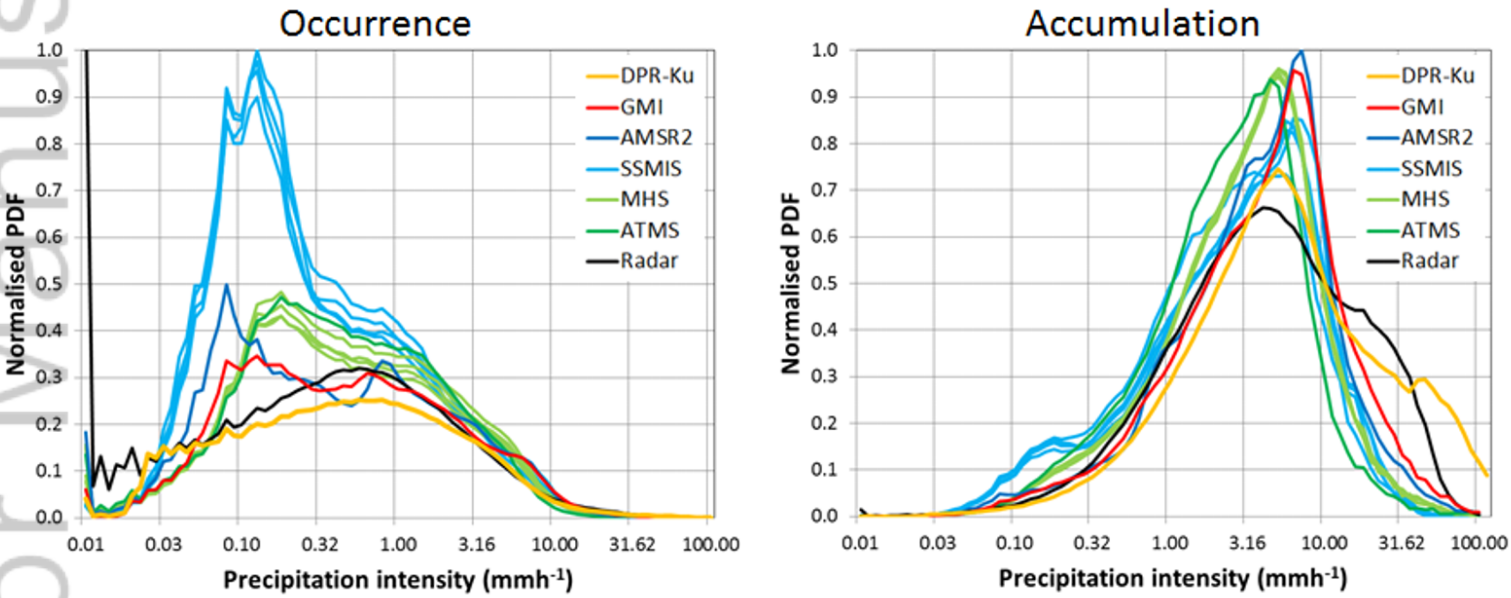
Kidd_Fig01.tif



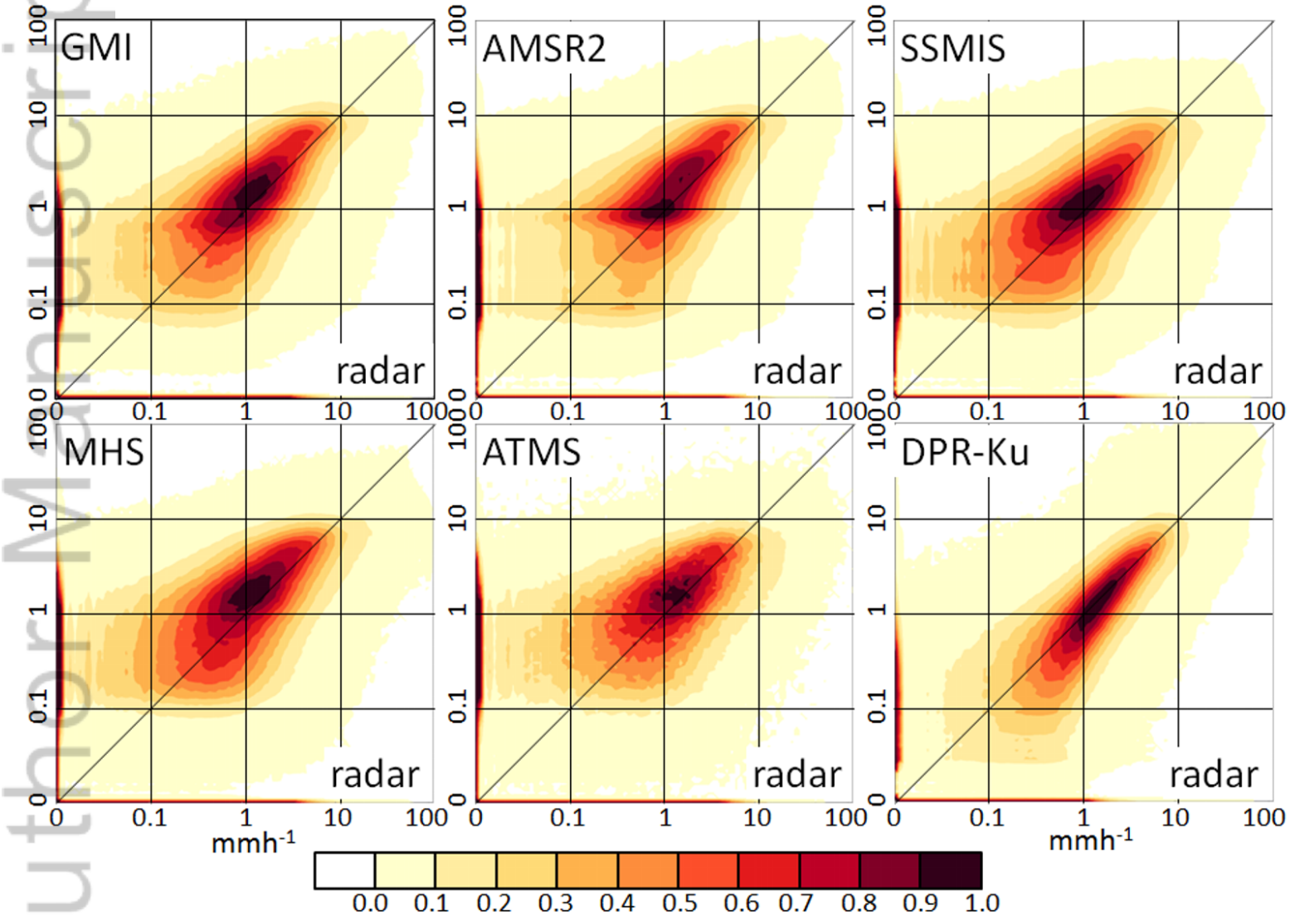
Kidd_Fig02.tif



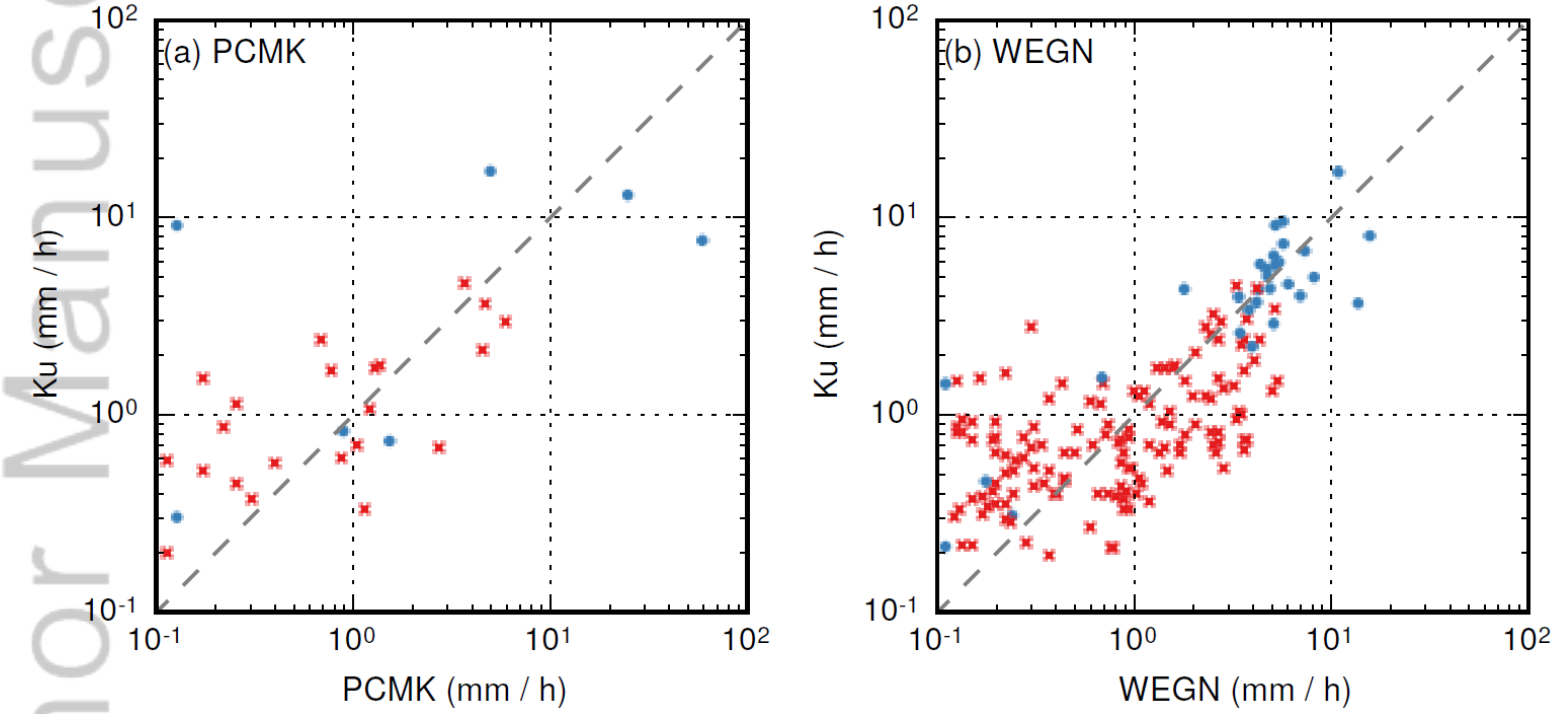
Kidd_Fig03.tif



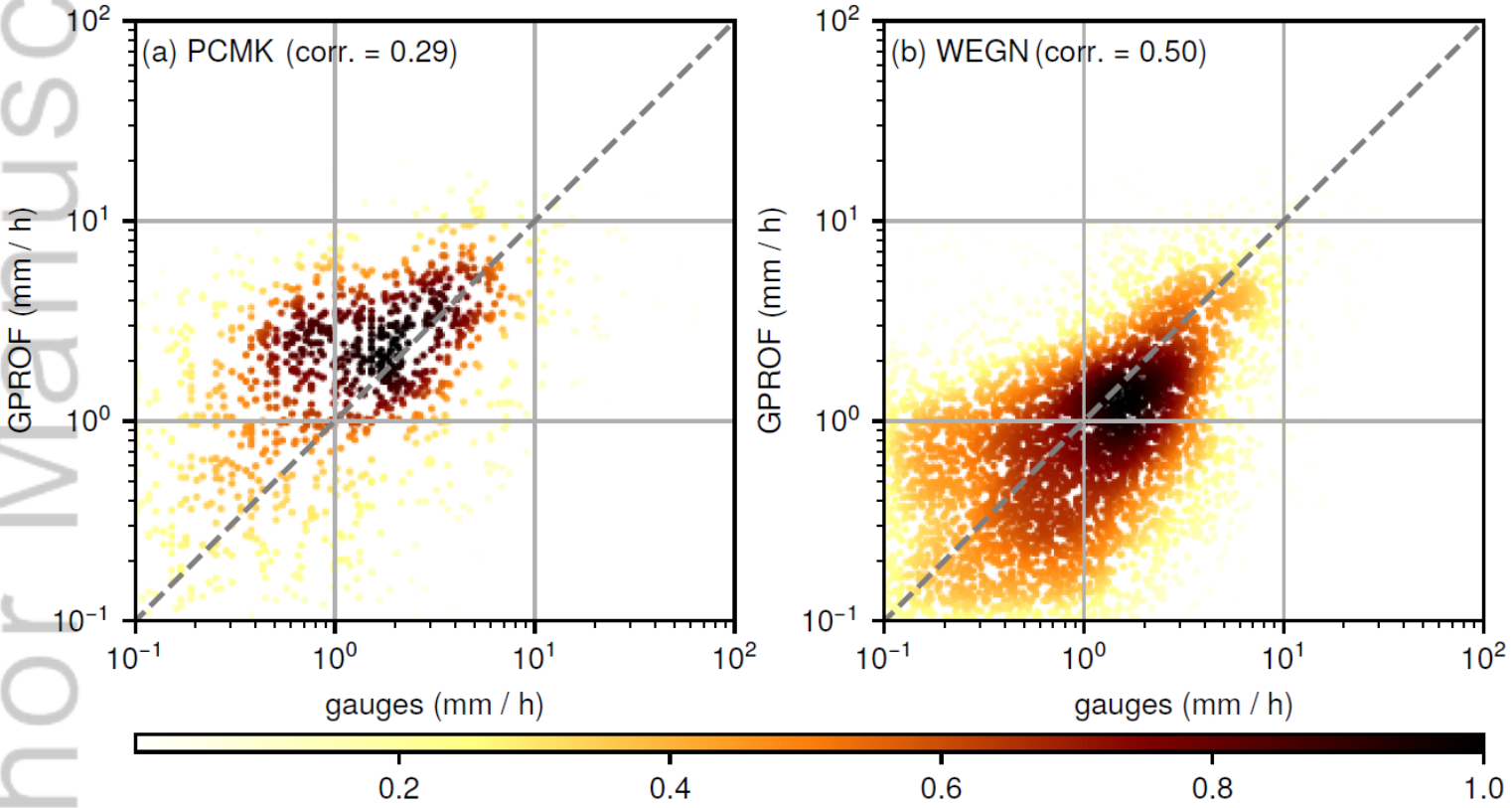
Kidd_Fig04.tif



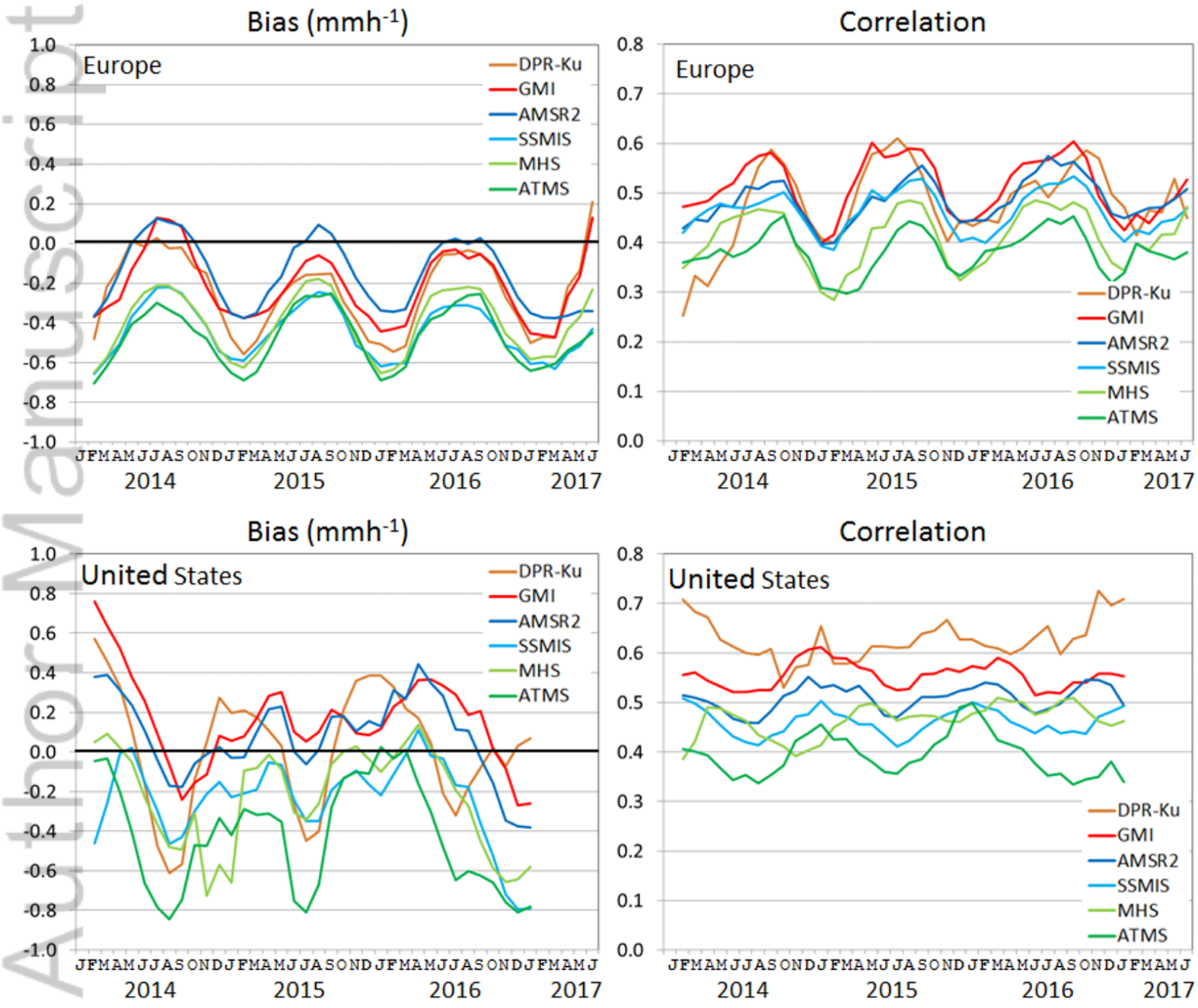
Kidd_Fig05.tif



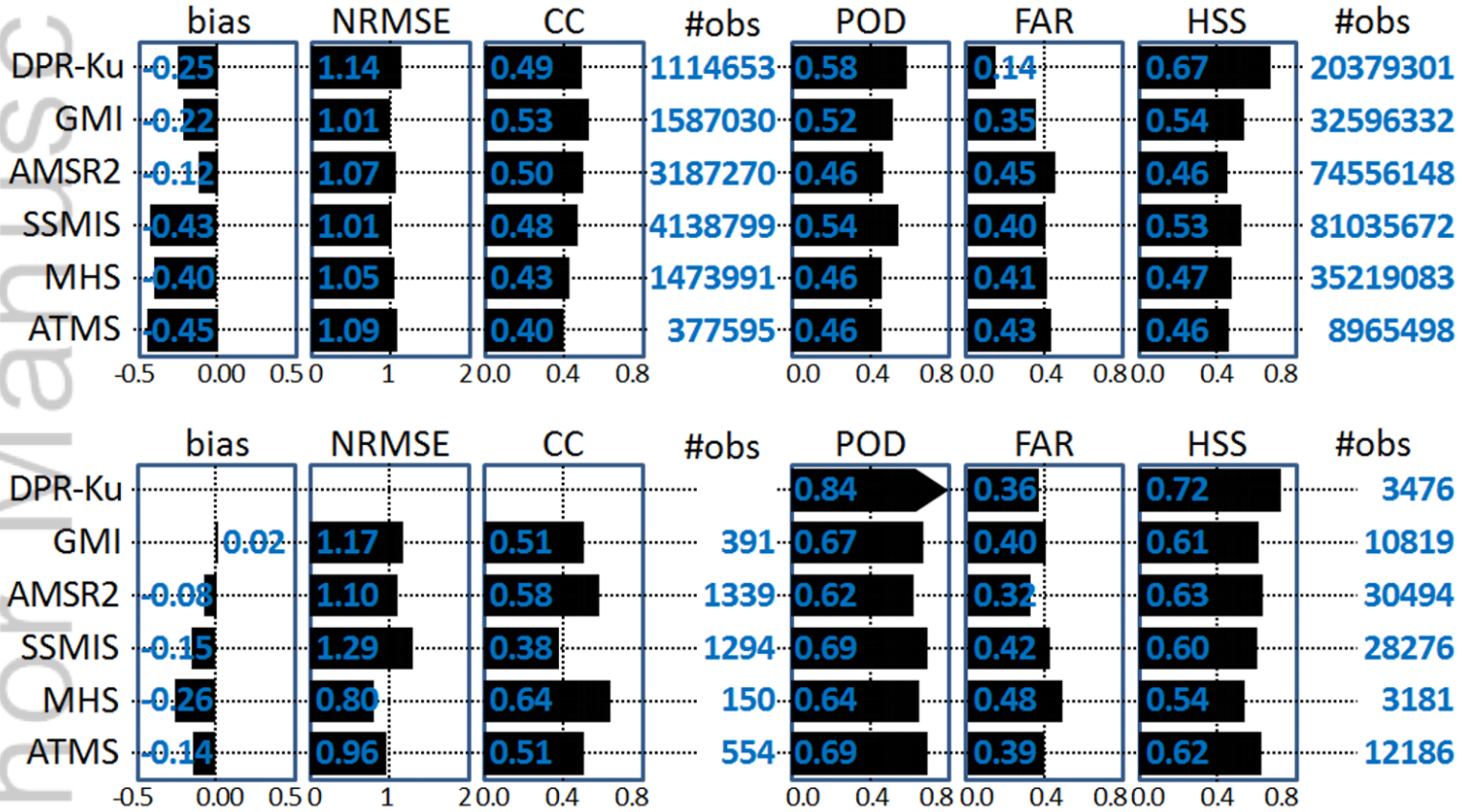
Kidd_Fig06.tif



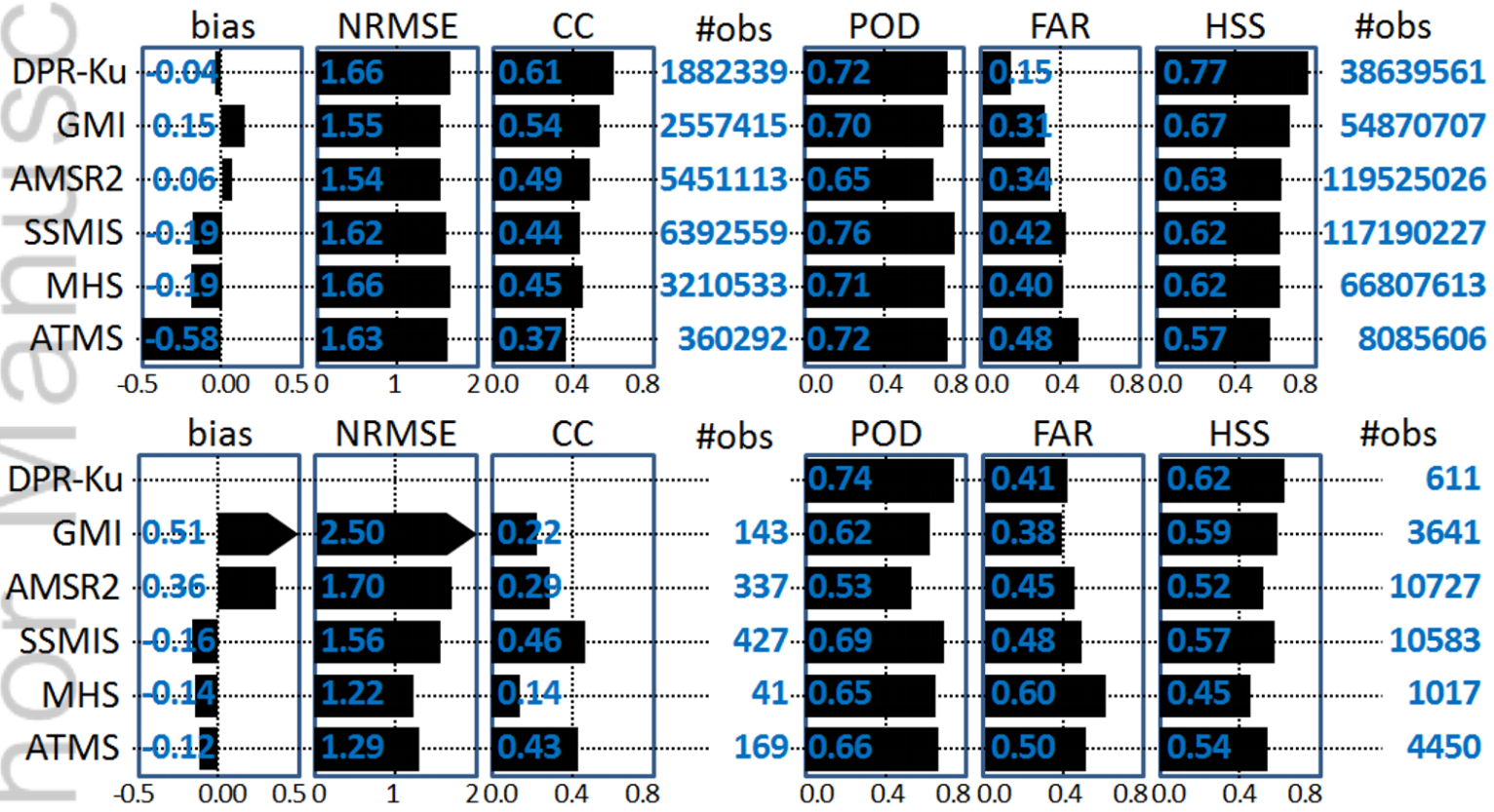
Kidd_Fig07_new.tif



Kidd_Fig08.tif



Kidd_Fig09.tif



Kidd_Fig10.tif