Calibrating ground-based radars against TRMM and GPM

Robert A. Warren*

School of Earth, Atmosphere and Environment, Monash University, Victoria, Australia

Alain Protat

Bureau of Meteorology, Melbourne, Victoria, Australia

Steven T. Siems and Hamish A. Ramsay

ARC Centre of Excellence for Climate System Science and School of Earth, Atmosphere and Environment, Monash University, Victoria, Australia

Valentin Louf and Michael J. Manton

School of Earth, Atmosphere and Environment, Monash University, Victoria, Australia

Thomas A. Kane

Bureau of Meteorology, Melbourne, Victoria, Australia

**Corresponding author address:* School of Earth, Atmosphere and Environment, Monash University, Clayton, VIC 3800, Australia

E-mail: rob.warren@monash.edu

ABSTRACT

Calibration error represents a significant source of uncertainty in quantitative applications of ground-based radar (GR) reflectivity data. Correcting it requires knowledge of the true reflectivity at well-defined locations and times during a volume scan. Previous work has demonstrated that observations from certain spaceborne radar (SR) platforms may be suitable for this purpose. Specifically, the Ku-band precipitation radars on board the Tropical Rainfall Measuring Mission (TRMM) satellite and its successor, the Global Precipitation Measurement (GPM) mission satellite, together provide nearly two decades of well-calibrated reflectivity measurements over low-latitude regions ($\pm 35°$). However, when comparing SR and GR reflectivities great care must be taken to account for differences in instrument sensitivity and frequency, as well as to ensure that the observations are spatially and temporally coincident. Here, a volume-matching method, developed as part of the ground validation network for GPM, is adapted and used to quantify historical calibration errors for three S-band radars in the vicinity of Sydney, Australia. Volume-matched GR–SR sample pairs are identified over a seven-year period and carefully filtered to isolate reflectivity differences associated with GR calibration error. These are then used in combination with radar engineering work records to derive a piecewise-constant time series of calibration error for each site. The efficacy of this approach is verified through comparisons between GR reflectivities in regions of overlapping coverage, with improved agreement when the estimated errors are removed.

1. Introduction

Since their development following the end of the Second World War, ground-based weather radars have become an indispensable tool for studying precipitation systems and associated phenomena on scales ranging from tens of meters to thousands of kilometers. Of particular value is their ability to provide quantitative information about surface rainfall intensity. Such information can be used by forecasters to monitor and warn for hazardous extreme-rain events, serves as input data to hydrological models, and allows for areal verification of quantitative precipitation forecasts. However, radar-derived rainfall estimates are subject to significant uncertainties (Villarini and Krajewski 2010). Many of these relate to assumptions that must be made regarding the drop size distribution and its evolution as hydrometeors fall from the level of observation to the surface, but perhaps the most fundamental uncertainty is that associated with errors in radar calibration.

The primary quantity measured by weather radars is the equivalent reflectivity factor Z (hereinafter reflectivity[1]), which has units of mm^6 m^{-3}. This is related to the returned power P_r from a target at range r via the radar equation:

$$Z = Cr^2 P_\mathrm{r} \tag{1}$$

Here, C is the so-called radar constant which depends on the radar-system characteristics (e.g. transmitted power, wavelength, beam width, pulse duration, antenna gain). In reality C is not constant, but varies due to degradation, maintenance, and replacement of various radar-system components, as well as due to thermal effects. Taking the common logarithm of (1) and multiplying by 10, we obtain an expression for the reflectivity measured in dBZ:

$$\hat{Z} = \hat{C} + 2\hat{r} + \hat{P}_\mathrm{r} \tag{2}$$

[1]Technically, the term reflectivity refers to the quantity $\eta = \pi^5 \lambda^{-4} \|K_\mathrm{w}\|^2 Z$, where λ is wavelength and $\|K_\mathrm{w}\|^2 = 0.93$ is the dielectric constant for liquid water. However, for brevity, and in keeping with previous studies on radar calibration, we will refer to Z as reflectivity.

where $\hat{\chi} = 10\log_{10}\chi$ for a variable χ. Hereinafter, we drop the circumflex and simply use Z to denote reflectivity, irrespective of the units. It can be seen that any error in the assumed value of $\hat{C}$ will produce an equivalent error in the reflectivity. This is referred to as a calibration error.

Maintaining a well-calibrated radar system requires regular testing and maintenance of those components which influence the true value of $\hat{C}$. Since this can be both time consuming and costly, there is great value in so-called end-to-end calibration tests which characterize the system as a whole. These tests typically involve the measurement of a target (or targets) with well-defined scattering properties, such as a standard reflector or metal sphere (Atlas 2002; Chandrasekar et al. 2015). An alternative approach is to compare reflectivity measurements with those from an independent well-calibrated radar system. The Ku-band precipitation radar (PR) on the Tropical Rainfall Measuring Mission (TRMM; Simpson et al. 1996) satellite, operational from 1997 to 2014, represented one such system. Internal and external calibration checks showed that, in the absence of attenuation, PR reflectivity measurements were accurate to within 1 dB (Kawanishi et al. 2000; Takahashi et al. 2003). The Ku-band component of the dual-frequency precipitation radar (KuPR) on board the Global Precipitation Measurement mission (GPM; Hou et al. 2014) Core Observatory satellite, which has now superseded TRMM, is anticipated to be equally accurate.

The task of comparing reflectivities observed by spaceborne and ground-based radars (hereinafter SRs and GRs, respectively) is complicated by the wildly different sampling characteristics of the two instruments. Operational GRs typically perform volume scans at regular intervals of 5–10 minutes. These scans consist of 360° radial sweeps performed at multiple elevation angles, ranging from near zero to around 20–30°. Samples of reflectivity are recorded every 0.5–1° in azimuth and every 250 m–1 km in range, out to maximum ranges of 150–300 km. By comparison, the TRMM PR and GPM KuPR measure quasi-vertical profiles of reflectivity within $\sim$ 250 km-wide orbital swaths, with horizontal and vertical sampling intervals of 5 km and 125–250 m, respectively. Sun-asynchronous orbits give rise to quasi-periodic observations at all locations within the satellite's latitudinal range ($\pm 35°$ for TRMM, $\pm 65°$ for GPM) with typical overpass frequencies of 1–2 day^{-1}. Another important difference between GR and SR measurements relates to the atmospheric volume sampled by each radar pulse. This volume is proportional to the angular beam width and increases with the square of range due to beam broadening. As a consequence, GR sample volumes vary by approximately five orders of magnitude within the instrument's field of view. In contrast, for SRs the extent of measurements in the range (vertical) direction is limited to the first 20 km above the surface and thus the relative variation in sample volume is small.

To quantitatively compare SR and GR reflectivities, measurements must be associated both in time and space. The ground speed of the satellites is sufficiently high that measurements across a typical GR field of view can be treated as instantaneous. Temporal association is thus achieved simply by identifying the GR volume scan closest in time to a given SR overpass. Due to the different sampling geometries, spatial association is much more challenging. Many researchers have taken the fairly simple approach of remapping both observation sets to a common three-dimensional Cartesian grid, using nearest-neighbour or linear interpolation (e.g. Anagnostou et al. 2001; Liao et al. 2001; Bolen and Chandrasekar 2003; Liao and Meneghini 2009b; Wang and Wolff 2009; Park et al. 2015). However, such procedures necessarily introduce errors which may swamp systematic differences in reflectivity associated with GR miscalibration. To overcome this issue, Schwaller and Morris (2011, hereinafter SM11) introduced what will herein be referred to as the volume-matching method (VMM). In this approach, intersections between individual SR beams and GR elevation sweeps are identified and the reflectivity values from both instruments are averaged within a spatial neighbourhood around the intersection. Specifically, SR data are averaged in range over the width of the GR beam at the GR range of the intersection while GR data are averaged in the range–azimuth plane within the footprint of the SR beam. The result is a pair of reflectivity measurements corresponding to approximately the same volume of atmosphere.

The VMM was originally developed as part of ground-validation efforts in support of the GPM mission. While the potential of the method as a means to track GR calibration was immediately apparent to the developers, its use in this context has thus far been very limited. Kim et al. (2014) applied the VMM to four GRs in the Korean Peninsula for the period 2006–2010, finding time-averaged calibration errors of between -2 and $+1$ dB. However, they were unable to identify shorter-timescale variations in GR calibration due to the noisiness of the GR–SR comparisons. This characteristic was also noted by SM11 and is believed to result from a combination of factors, including imperfect spatial and temporal matching, differences in radar frequency, and errors in SR attenuation correction.

The present study summarizes our efforts using the VMM to quantify and correct historical calibration errors for three GRs in the vicinity of Sydney, Australia. We first explore how systematic variations in GR–SR reflectivity difference can be related to certain characteristics of the volume-matched sample pair. By isolating samples which are least influenced by these artifacts it is possible to significantly reduce the noise in GR bias estimates. We then demonstrate how VMM results can be used in combination with radar engineering maintenance records to identify variations in GR calibration on inter- and intra-annual timescales. Finally, we present a simple method of comparing GR observations in regions of overlapping coverage as a means to validate the estimated bias corrections.

2. Methodology

a. Data

The Australian Bureau of Meteorology (BoM) operates a diverse network of over 60 single-polarization GRs comprising a mixture of C- and S-band systems of varying age and make. This study uses data from three S-band radars located close to the cities of Sydney (SYD), Wollongong (WOL), and Newcastle (NEW) in the state of New South Wales (Fig. 1). Together, these sites provide coverage of a densely populated stretch of coastline which is frequently affected by high-impact weather, including damaging hail and extreme precipitation. The characteristics of the radar systems are listed in Table 1. For each of the GRs, volume-scan data for the period 15 May 2009 (the start of operational monitoring at SYD) to 31 December 2015 were extracted from BoM archives and converted from the in-house Radar Picture (RAPIC) format to OPERA (Operational Program on the Exchange of Weather Radar Information; Köck et al. 2000) Data Information Model–Hierarchical Data Format version 5 (ODIM-HDF5; Michelson et al. 2014) for processing. Note that all data are subject to on-site processing to mitigate ground clutter and noise (Rennie 2012). No additional quality control was applied for the present analysis.

BoM engineering staff perform regular (approximately once every 6 months) maintenance works at all GR sites. Relevant to the radar calibration are checks on the transmitted peak power, frequency, and pulse duration, and on the receiver gain. Where necessary, these settings are adjusted and the radar constant is updated accordingly. In addition to these routine activities, unscheduled maintenance is sometimes required to deal with system failures or suspected faults. Records of all sites visits are maintained on an internal database called SitesDB. While the information contained in these records is minimal, with only a date and brief description of what was done (e.g. “02/12/2009: 6 monthly maintenance carried out”), it is sufficient to identify dates of *possible* calibration changes. In theory, calibration accuracy should improve following all maintenance works; however, as we shall demonstrate, this is often not the case.

The characteristics of the TRMM PR and GPM KuPR are listed in Table 2. TRMM operated almost continuously from December 1997 to April 2015, with the PR providing reliable measurements up to September 2014. This study uses data from Version 7 of the Level 2 products 2A23 and 2A25 (Table 3). These consist of orbital swaths made up of a large number of individual PR scans which in turn comprise 49 individual rays. Each scan has a unique time-stamp and rays are georeferenced by the latitude–longitude coordinates of their intersection with the Earth ellipsoid. The 2A23 product contains information on precipitation type and the characteristics of the radar brightband[2] (where present) for each ray, while 2A25 contains the vertical profiles of attenuation-corrected reflectivity. Precipitation type is determined based on the horizontal and vertical echo structure (Awaka et al. 2007), with three basic classifications: stratiform, convective, and other. The brightband is identified as outlined in Awaka et al. (2009). A hybrid method (Meneghini et al. 2004), combining the approaches of Hitschfeld and Bordan (1954) and Meneghini et al. (2000), is used to correct for attenuation of the SR beam, which can be significant in heavy rainfall. For the GPM KuPR, data are available from March 2014 onwards. Version 4 of the 2AKu product is used which contains the same basic variables as the 2A23 and 2A25 TRMM products (Table 3). All SR data were obtained using the STORM online data-access interface to NASA's precipitation processing system archive (`https://storm.pps.eosdis.nasa.gov`). To reduce data volumes, only those sections of orbital swaths corresponding to GR site overpasses were extracted.

It is noted that, at the time of writing, new product versions (Version 8 for TRMM and Version 5 for GPM) are in the process of being released. These include changes to the SR calibrations,

[2] The brightband is a layer of locally enhanced reflectivities around the melting level which occurs due to changes in the scattering properties of snow as it melts.

corresponding to reflectivity increases of 1.1 and 1.3 dB for the TRMM PR and GPM KuPR, respectively (NASA 2017; Iguchi et al. 2017). It remains to be seen whether these will be the final adjustments, but for now it must be assumed that the GR calibration errors derived herein are biased low by a little over 1 dB. This serves to illustrate the main limitation of using radar intercomparisons to assess calibration: even the most carefully monitored systems can be in error.

b. Volume-matching method

The VMM allows for quantitative comparison of SR and GR reflectivities with minimal spatial processing of the two datasets. Intersections between an SR beam and a GR elevation sweep are identified and the reflectivities from both instruments are averaged to roughly equate the sample volumes. SR reflectivities are averaged along the SR beam (approximately vertically) between the half-power points of the GR sweep. GR reflectivities are averaged in the range–azimuth plane (approximately horizontally) within the footprint of the SR beam. Figure 2 illustrates these averaging procedures for idealized cases at GR ranges of 50 and 100 km. Full details of the procedure are provided in the appendix. Here we only note the key differences between our implementation of the method and the original algorithm as described by SM11 and Morris and Schwaller (2009).

1) Minimum and maximum range

As previously discussed, the volume of atmosphere sampled by a GR varies significantly across the instrument's field of view due to beam broadening. This means that samples considered in the VMM also increase in volume with GR range. Given the limited vertical extent of many precipitating systems it is appropriate to define a maximum range, $r_{\max}$, for volume-matching to proceed. SM11 specified $r_{\max} = 100$ km, while we use a slightly higher value of 115 km. For the WOL radar which has an angular beamwidth $\omega = 2°$ this corresponds to a maximum beam diameter of

4 km. Since all the GRs considered by SM11 had $\omega \approx 1^\circ$, their maximum beam diameters were < 2 km. However, as we shall show, the GR–SR reflectivity difference displays relatively little sensitivity to range (and thus beam diameter). Unlike SM11, we additionally specify a minimum range in order to exclude samples where the GR beam width is smaller than the SR gate spacing. Specifically, $r_{\text{min}} = 15$ km which for $\omega = 1^\circ$ corresponds to a beam diameter of just over 250 m, the gate spacing of the TRMM PR.

2) Frequency correction

The different frequencies used by the SR and GR systems promotes systematic differences between the reflectivity measured by the two instruments which vary in both sign and magnitude depending on the scattering characteristics of particles within the sample volume. Scattering simulations can be used to quantify these differences and derive empirical relationships for converting reflectivity measurements from one frequency to another. SM11 used the equations from Liao and Meneghini (2009a) to convert their GR reflectivities from S to Ku band, applying the equations for snow and rain above and below the brightband, respectively. Since we are interested in quantifying GR errors it is desirable to instead convert the SR reflectivities from Ku to S band. We therefore use equations from Cao et al. (2013) which have the following form:

$$Z(\mathrm{S}) = Z(\mathrm{Ku}) + \sum_{i=0}^{4} a_i \left[Z(\mathrm{Ku})\right]^i \tag{3}$$

The coefficients a_i (given in Table 1 of Cao et al. 2013) are specified for rain, dry snow, and dry hail, and for snow and hail at varying stages of melting (from 10 to 90 % in 10 % increments). The melting layer (ML) is defined as extending from $z_\mathrm{b} - \Delta z_\mathrm{b}/2$ to $z_\mathrm{b} + \Delta z_\mathrm{b}/2$, where z_b and Δz_b are the SR-derived brightband height and width, respectively. To deal with the fact that a brightband is only present in stratiform precipitation, both quantities are computed as the median value across all stratiform SR rays that intercept the Earth ellipsoid between r_{min} and r_{max}. Overpasses where there are fewer than 10 such rays are excluded from further analysis.

3) Reflectivity thresholds

The TRMM PR and GPM KuPR both have nominal sensitivities of around 18 dBZ (Hou et al. 2014), although pre-launch tests showed that the KuPR may detect reflectivities as low as 14.5 dBZ (Toyoshima et al. 2015). In the VMM, only SR bins for which $Z_{\mathrm{s}} \geq Z_{\mathrm{s}}^{*} = 18$ dBZ are included in the calculation of the average SR reflectivity. For each volume-matched sample, the fraction of SR bins within the volume which meet this criterion, f_{s}, is recorded. A similar approach is taken with the GR using a different reflectivity threshold, Z_{g}^{*}, with the fraction of GR bins where $Z_{\mathrm{g}} \geq Z_{\mathrm{g}}^{*}$ denoted as f_{g}. When analysing the GR reflectivity bias, effects associated with nonuniform beam filling and the low PR sensitivity can be mitigated by excluding samples with f_{s} and f_{g} less than some threshold f_{min}. Based on analysis presented below, we set $f_{\mathrm{min}} = 0.7$, while SM11 used the more stringent criterion $f_{\mathrm{min}} = 0.95$. As discussed by Morris and Schwaller (2011) and illustrated below, GR–SR reflectivity differences derived using the VMM can vary substantially depending on the value of this threshold.

Another key difference is in our choice of the GR reflectivity threshold. SM11 set $Z_{\mathrm{g}}^{*} = 15$ dBZ to match the SR sensitivity with allowance for a -3 dB GR calibration error. While it is necessary to match the sensitivity of the two instruments when using one to quantify bias in the other, we argue that this should be done at a later stage in the analysis, namely when comparing the spatially averaged reflectivities from the volume-matched samples. As detailed in section 3c, this allows for the implementation of an iterative bias correction procedure where GR samples are filtered according to their bias-corrected reflectivity at the nth iteration (Protat et al. 2011). In the volume matching we therefore employ a much lower GR reflectivity threshold, $Z_{\mathrm{g}}^{*} = 0$ dBZ.

4) Reflectivity averaging

In the original VMM implementation, reflectivities for GR bins within the SR footprint are averaged using a Barnes Gaussian inverse-distance weighting, where distance is measured horizontally from the centre of the SR footprint to the centre of the GR bin (Morris and Schwaller 2009). This weighting is designed to account for the nonuniform distribution of power within the SR beam. The algorithm has since been updated to also include a linear weighting based on the volume of the GR bins, so that larger volumes are weighted more heavily (K. Morris 2015, personal communication). This is justified by the fact that GR bin volumes can vary by up to a factor of two within the PR footprint. Our VMM implementation uses this modified weighting scheme. As in the original algorithm, no weighting is applied in averaging the SR reflectivities due to uncertainties in the GR beam height associated with nonstandard refraction.

3. Results

a. Comparison examples

Figures 3 and 4 show examples of GR–SR comparisons for the SYD radar. The former shows a comparison with TRMM on 22/11/2013 while the latter shows a comparison with GPM a little over a year later on 27/01/2015. The top row in each figure shows plan views at a particular elevation angle of the (frequency-corrected) SR and GR reflectivities and their difference. The middle row shows vertical cross-sections along a particular SR scan of the same fields. The bottom row presents a statistical comparison of the reflectivities from the two instruments across all volume-matched samples. Note that samples with $f_s < f_{min}$ or $f_g < f_{min}$ have been excluded from this analysis.

From the plan views of reflectivity it appears that the VMM produces good spatial agreement between the reflectivity measurements from the two instruments. This is confirmed by high values of the Pearson correlation coefficient (0.95 and 0.87 for the first and second comparisons, respectively). This agreement allows us to estimate the GR calibration error. In the first case, the error is close to zero (Fig. 3i); however, in the second the GR shows a substantial negative bias of around 4 dB (Fig. 4i). It is thus apparent that the calibration of the SYD radar changed some time between late 2013 and early 2015.

It is noteworthy that, on a point-by-point basis, the GR–SR reflectivity difference displays a large degree of scatter. For example, for the first case, the difference varies by more than 10 dB (from < -5 to > 5 dB) across the 1.3° elevation sweep (Fig. 3c). Part of this variation will be associated with imperfect spatial matching of the data due to a combination of advection and evolution of the precipitation features during the time between measurements (200 s in this case) and beam propagation effects (e.g. non-standard refraction of the GR beam). However, as we shall demonstrate in the next section, other factors including the Ku-to-S-band frequency correction and the reflectivity value itself also strongly influence GR–SR reflectivity differences.

b. Comparison sensitivities

In this section, we investigate the sensitivity of the GR–PR reflectivity difference, ΔZ, to various characteristics of the volume-matched samples. To eliminate effects associated with the time-varying GR calibration errors we have applied the corrections derived in the next section to all GR data. We begin by examining the relationship between ΔZ and f_{min}, the minimum fraction of SR and GR bins within the sample volume with reflectivities above the respective thresholds, $Z_{\mathrm{s}}^{*} = 18$ dBZ and $Z_{\mathrm{g}}^{*} = 0$ dBZ. This is illustrated in Fig. 5 for each of the GRs. The data are binned using f_{min} values from 1 to 0 in increments of 0.1, with the median and interquartile range (IQR) of the ΔZ distribution in each bin plotted together with the number of volume-matched sample pairs.

For the most restrictive case of $f_{\text{min}} = 1$, all SR and GR bins comprising a sample must satisfy the reflectivity criteria. This ensures good volume matching but severely limits the number of valid samples. In contrast, for $f_{\text{min}} = 0$ only a single bin for each radar needs to exceed the respective reflectivity thresholds. This gives many more valid samples but can lead to very poor volume matching. As f_{min} is decreased, we thus observed an increase in both the number of samples and the variability in ΔZ (Fig. 5). The change in sample size is more pronounced for the NEW and WOL radars due to their larger beam widths; at a given GR range, more SR bins are included in each sample so the probability that $f_{\text{s}} < f_{\text{min}}$ is higher. For all three GRs, there is a pronounced decrease in the median ΔZ with decreasing f_{min}, with the total change being around 1–1.5 dB. This trend, also noted by Morris and Schwaller (2011, their Figs. 2–5), results from the low sensitivity of the SRs. As f_{min} is reduced, an increasing number of samples comprise bins with $Z < 18$ dBZ which the GR can observe but the SR cannot. Thus the average volume-matched GR reflectivity decreases while the corresponding SR reflectivity remains approximately constant.

Clearly, it is important to exclude samples with low values of f_{s} or f_{g}. Ideally we would set $f_{\text{min}} = 1$; however, testing showed that the associated reduction in sample size severely limits our ability to derive a complete time series of calibration error (not shown). As a compromise we therefore set $f_{\text{min}} = 0.7$. In doing so, Figure 5 suggests that we will introduce a slight negative bias in our calibration error estimates. However, it turns out that that this bias is largely mitigated by the reflectivity thresholding described below (not shown).

We now examine how ΔZ varies with precipitation type and height together with the impact of the Ku-to-S-band frequency correction which is applied to SR reflectivities. This information is summarized using box-and-whisker diagrams in Fig. 6. Here, samples for each GR are divided according to the SR precipitation type classification (stratiform or convective) and based on their height with respect to the ML (below, within, or above). For both precipitation types the frequency corrections for dry and melting snow have been used above and within the ML, respectively. The relationships for hail were initially used in convective precipitation but were found to worsen the agreement between ΔZ above and below the ML (not shown). Samples with precipitation type "other" accounted for a very small proportion (< 1 %) of the total for all radars and are therefore excluded from this analysis.

The frequency correction results in an increase in ΔZ (via a decrease in Z_{S}) below the ML and a decrease in ΔZ (via an increase in Z_{S}) within and above the ML. Changes are more pronounced in convective than stratiform precipitation because the former is characterized by higher reflectivities. For all three GRs, we observe good agreement between the frequency-corrected ΔZ distributions above and below the ML in stratiform precipitation. However, within the ML the distributions are shifted upwards, suggesting that the frequency correction for melting snow is underestimated. This layer also shows higher variability in ΔZ due to the fact that it includes all samples whose volume overlaps the brightband. For convective precipitation, the frequency correction clearly increases the discrepancy between the different vertical layers, promoting a systematic decrease in ΔZ with height. We speculate that this is associated with undercorrection of SR beam attenuation in convective precipitation (leading to underestimation of Z_{S} and thus overestimation of ΔZ); however, errors in the frequency correction may also contribute. In addition to a disagreement between the layers, we note that the convective samples feature larger spread in ΔZ, consistent with higher spatial variability in the precipitation field and associated poorer volume matching.

Based on these results we exclude convective precipitation samples and stratiform samples within the ML from all subsequent analysis. This reduces the SYD radar sample size by approximately 62 % and the WOL and NEW radar sample sizes by approximately 77 % (the larger beam widths of these radars mean that more samples overlap with the ML). It should be noted that, in order to mitigate potential biases associated with the SR attenuation correction, only stratiform samples *above* the ML are used in ground validation of the GPM DPR (Walt Petersen 2017, personal communication). However, when combined with the reflectivity criteria introduced below, the exclusion of samples below the ML was found to excessively limit the total number of samples. Testing reveals a slight (typically < 0.5 dB) but systematic increase in calibration error estimates when only samples above the ML are used (not shown). This may be indicative of excessive attenuation correction in stratiform precipitation (c.f. Wang and Wolff 2009) and/or undercorrection for the Ku–S band frequency difference in snow.

Figure 7 summarises the influence of two further sample characteristics, GR range, r_{g}, and GR–SR time difference, Δt, on ΔZ. The data are plotted as bivariate histograms with the median and IQR of ΔZ overlaid for each r_{g} and Δt bin. As one might expect, there is little dependence for either variable. At ranges beyond ~ 60 km, ΔZ shows a weak decreasing trend with increasing r_{g} for the WOL and NEW radars which is not present for the SYD radar. This is likely due to the fact that the beam widths of the WOL and NEW radars are around twice the angular beam spacing, $\Delta\phi$, whereas for SYD $\omega = \Delta\phi$. For $\omega > \Delta\phi$, the GR reflectivity of the volume-matched sample will tend to represent a larger area (in the range–azimuth plane) than observed by the SR, giving rise to a slight negative bias in Z_g and thus ΔZ, particularly at long ranges where the absolute difference in area is large. It should be noted that Morris and Schwaller (2011) found the same trend (increasing SR $-$ GR reflectivity in their case) despite the fact that the radar they considered (the WSR-88D in Melbourne, Florida) had a 1° beam. This probably reflects their use of a higher GR reflectivity threshold which will have reduced the number of samples with low Z_{g} (and thus low ΔZ; see below) at short range where many GR bins are averaged.

Turning to the lower row of Fig. 7, it is clear that there is no systematic variation in ΔZ with Δt; however, larger time differences are associated with higher variability, as seen from the IQRs. This again is consistent with the findings of Morris and Schwaller (2011) and makes intuitive sense: larger Δt implies a greater spatial mismatch between the SR and GR volumes, leading to larger random errors in ΔZ. These errors could potentially be reduced by applying an advection correction to each GR sweep; however, we do not attempt this here.

The final sensitivity we consider is to the reflectivity itself. Of course, we have two measures of this quantity and it is important to consider both. The top row of Fig. 8 shows how ΔZ varies with SR reflectivity, Z_s, for the three GRs, using the same format as Fig. 7. For SYD and WOL, ΔZ shows a slight increasing trend for $Z_\mathrm{s} <\sim 27$ dBZ while for all three radars there is a similarly weak decreasing trend for $Z_\mathrm{s} >\sim 33$ dBZ. The origin of the first of these trends is unclear; however, the second may be associated with the Ku-to-S-band frequency correction. Without this correction, the trend is much more pronounced (not shown), suggesting that with larger corrections it would disappear altogether. It is quite possible that the Cao et al. (2013) method underestimates the frequency correction at high reflectivities; however, given the other sources of uncertainty it is difficult to be sure. In any case, the associated variation in ΔZ is small (< 1 dB in the median).

The variations in ΔZ with Z_g are much more substantial (bottom row of Fig. 8). For all three radars, there are three distinct portions of the parameter space. For $Z_\mathrm{g} <\sim 24$ dBZ, ΔZ is negative and shows a strong positive trend. This is a direct consequence of the low sensitivity of the SRs. For $Z_\mathrm{g} < 18$ dBZ, the GR reflectivity is constrained to be lower than the SR reflectivity and thus ΔZ is constrained to be negative; similarly, if Z_g only slightly exceeds 18 dBZ, ΔZ can only be slightly positive. Effectively, the top-left portion of the histogram has been cut off. The trend only disappears once the reflectivity is large enough that the distribution of ΔZ becomes roughly symmetric, which occurs around $Z_\mathrm{g} = 24$ dBZ. Beyond this point, ΔZ remains almost constant up to around $Z_g = 36$ dBZ when it begins to rapidly increase again. We believe the latter trend to be associated with attenuation of the SR beam in regions of intense stratiform precipitation. This would be consistent with Liao and Meneghini (2009b) and SM11, who both noted an undercorrection of attenuation in version 6 of the TRMM 2A25 product, as well as several studies (Wolff and Fisher 2008; Amitai et al. 2009; Chen et al. 2013; Kirstetter et al. 2013; Rasmussen et al. 2013) which identified negative biases in PR rainfall estimates at high rain rates.

Summarizing the results of this section, we have identified several factors which strongly influence the GR–SR reflectivity difference estimates obtained using the VMM; namely, the percentage of above-threshold reflectivity values within a sample, the height of the sample with respect to the ML, the application of a Ku-to-S-band frequency correction, the precipitation type, and the reflectivity itself. Based on these findings we extract a subset of volume-matched samples for each radar which are expected to most accurately isolate reflectivity differences associated with GR calibration errors. Specifically, samples are only included if they:

A. comprise at least 70 % SR and GR bins with reflectivities above the respective thresholds;

B. are located entirely above or below the ML in stratiform precipitation;

C. have volume-averaged SR and GR reflectivity values between 24 and 36 dBZ.

Table 4 shows how the sample size, mean ΔZ and its standard deviation vary with the application of these criteria. Consistent with the discussion above, criteria A and C both produce a pronounced positive shift in mean ΔZ while criterion B produces a smaller negative shift. All three criteria act to reduce variability, with C having by far the biggest impact. This is almost entirely due to the lower reflectivity threshold; the impact of the higher threshold is much smaller because there are far fewer samples with high reflectivities. Applying all three criteria together results in a 2–2.4 dB reduction in the standard deviation of ΔZ; this despite the fact that the sample size decreases by more than 90 % for each radar.

c. Correcting calibration errors

Figure 9 shows the complete seven-year time series of GR–SR comparisons for the SYD radar. Plotted are the mean reflectivity difference (symbols, colored according to the number of samples) and its standard deviation (vertical lines) for each SR overpass. It is apparent that, even with the filtering criteria detailed above, there is considerable variability in ΔZ values (c.f. Table 4), particularly for those comparisons with fewer than 100 samples (white and light grey symbols). This is most likely associated with residual volume-matching errors in the presence of rapidly moving/evolving precipitation features and/or non-standard GR beam refraction. Nevertheless, it is possible to identify the basic temporal evolution of GR calibration.

From the start of operations in May 2009 until the middle of 2014 the calibration appears to be quite accurate and stable, with mean errors generally less than 2 dB. A possible exception is September/October 2012 where several comparisons suggests a negative offset of around 4–5 dB, although sample sizes for these are small. The period August 2014 to May 2015 shows more significant GR errors, with positive offsets of 3–4 dB during the first three months and negative offsets of 3–5 dB thereafter. There are no comparisons during June and July 2015 and only one each in August and September; however, towards the end of the year errors return to near zero.

While the VMM does not provide sufficiently precise estimates of GR reflectivity error to identify gradual changes in calibration associated with the degradation of radar hardware, it can pick out sudden jumps which may result from component failures or engineering activities. The problem is that suitable SR site overpasses are rarely frequent enough to determine the exact date of these changes. Fortunately, as discussed in section 2a, the BoM maintains records of all operational GR maintenance works. From these records, the dates of possible calibration changes were identified and used to group the GR–SR comparisons into periods ranging in length from a few weeks to around 18 months. The calibration error, ε, during each period is assumed to be constant and is calculated as follows:

1. Valid samples (i.e. those meeting criteria A–C, above) comprising all GR–SR comparisons during the period are grouped and the mean ΔZ is computed as an initial estimate of ε.

2. The set of valid samples is recomputed incorporating the estimated calibration error (i.e. with ε subtracted from the GR reflectivities) and a new value of ε is computed as the mean of the uncorrected ΔZ values.

3. Step 2 is repeated iteratively until a stable estimate of ε is obtained (to the nearest 0.1 dB). Typically, this takes fewer than five iterations.

As discussed by Protat et al. (2011), an iterative calculation is required when thresholding the reflectivity to account for the fact that, given a non-zero calibration error, samples will be incorrectly included/excluded from the calculation of ε. For example, consider a situation where the true ε is -3 dB. In the initial estimation (step 1, above), samples with uncorrected reflectivites of 21–24 dBZ (true reflectivities of 24–27 dBZ) will be incorrectly excluded while those with uncorrected reflectivities of 33–36 dBZ (true reflectivities of 36–39 dBZ) will be incorrectly included. Similarly, if the true ε is $+3$ dB, samples with uncorrected reflectivites of 24–27 dBZ (true reflectivities of 21–24 dBZ) will be incorrectly included while those with uncorrected reflectivities of 36–39 dBZ (true reflectivities of 33–36 dBZ) will be incorrectly excluded. In either case, the magnitude of ε will be underestimated. By subsetting samples according to the corrected GR reflectivities and recomputing ε iteratively this bias can be eliminated. Figure 10 illustrates the procedure for two consecutive periods (one with positive ε, one with negative ε) from the

SYD radar time series. In both cases, iteration increases the the magnitude of the calibration error estimate by 0.6 dB.

Not every single maintenance event will be associated with a change in radar calibration. For example, checks may show the transmitter and receiver settings to be stable with respect to the previous site visit. We therefore test whether the calibration error during each period is statistically distinct from the one which preceded it. Specifically, a difference of means test is performed using the error-adjusted samples from each period. If the difference is significant at the 5 % level[3] and ≥ 0.5 dB then both periods are retained; otherwise, the two are combined and the GR bias estimate is recomputed. Periods are also combined if one contains fewer than two comparisons comprising at least 50 samples each; we consider this the minimum requirement for a robust error estimate. The choice of 0.5 dB as a minimum difference is somewhat arbitrary but reflects the remaining uncertainty in the GR–SR comparisons (i.e. we do not expect the method to reliably detect changes in calibration of less than 0.5 dB).

Figure 11 shows the time series of GR–SR reflectivity difference for the SYD radar following the calculation of calibration error. The dates of possible calibration changes and the mean ΔZ (estimated ε) and its standard deviation for each intervening period are also indicated. Comparing with Fig. 9, it can be seen that the sample size and mean values for each comparison have changed, particularly where ε is large in magnitude (e.g. in September and October 2012), due to the use of bias-corrected GR reflectivities in the filtering of samples. Overall, the method appears to work very well. It is able to identify the above-noted major calibration changes in 2012, 2014 and 2015, as well as more subtle changes, for example in December 2011. Values of ε range from -5.3 to $+3.5$ dB with the average over the entire seven-year period being -0.6 dB. The same analysis

[3] Other significance levels (10 % and 1 %) were tested with almost no change in the results.

for the WOL and NEW radars (not shown) reveals similar maximum error magnitudes but more negative values on average with means of -1.4 and -1.7 dB, respectively.

Two aspects of these results must be remarked upon. The first is the large magnitude of the calibrations errors, with values frequently > 1 dB and occasionally > 5 dB. These have the potential to mislead forecasters (by suggesting that storms are more/less intense than they really are) and significantly impact radar-derived products, particularly when values are integrated in time (e.g. precipitation accumulations) or space (e.g. vertically integrated liquid water content). The second aspect to remark upon is the change in calibration associated with radar maintenance activities. One would hope that system checks and modifications always act either to maintain an existing good calibration or improve a poor one. However, our results show that this is often not the case. For example, following preventative maintenance of the SYD radar in July 2014, the calibration error was increased from $+0.8$ dB to around $+3.5$ dB (Fig. 11). Further works later that year saw the introduction of an error of roughly the same magnitude but opposite sign (-3.7 dB). It is difficult to ascertain the reason for these changes from the limited textual information contained in SitesDB; however, human error and miscalibrated test equipment may both play a role. Clearly, there is a need for more careful monitoring of radar calibration during operations, an issue we discuss further in section 4.

d. Verification

By combining filtered GR–SR comparisons with radar engineering records, we have been able to quantify historical calibration errors for three GRs in the vicinity of Sydney. We now seek to evaluate the benefits achieved by accounting for these errors. Comparison against ground truth such as rain gauges is theoretically one means to achieve this goal; however, as noted in the introduction, radar rainfall estimates are subject to many additional sources of uncertainty. We therefore instead investigate how the consistency of our three radars differs with and without calibration adjustments. Agreement between neighbouring GRs is important both from an operational perspective (e.g. forecasters viewing a storm using different radars should obtain the same impression of its intensity) and for the production of multi-radar products such as regional and national rainfall maps. Since we are adjusting the GRs relative to the same SR reference we would expect the agreement between them to improve.

Following the rationale behind the VMM, we minimize spatial processing (interpolation and averaging) of the measured reflectivities and associated errors by directly matching sample volumes in space and time. The only spatial processing we apply is the averaging of reflectivities in range to achieve a consistent gate spacing ($\Delta r_0 = 1000$ m) across all three radars. For all possible radar pairs (SYD–WOL, SYD–NEW, and WOL–NEW), we then identify bins which are (a) close in space (centres < 500 m apart), (b) close in time (elevation sweeps < 2 minutes apart), and (c) similar in size (difference in volume < 10 %). Spatial association is achieved by mapping data to a common Cartesian grid using an azimuthal equidistant projection centred half way between the sites. For simplicity, we model the volume of atmosphere sampled by each bin as a cuboid with dimensions of Δr_0, $r\left(\omega + \Delta\phi \cos\theta\right)$, and $r\omega$ in the range, azimuth, and elevation directions, respectively. Here, we account for the fact that each azimuthal sector comprises multiple rays which overlap by an increasing degree with increasing elevation angle (the $\Delta\phi \cos\theta$ term). The fractional volume difference between radar bins i and j is computed as $\left|V_i - V_j\right| / \left[\frac{1}{2}\left(V_i + V_j\right)\right]$. To reduce computational expense, only days with widespread rainfall in the area of overlapping coverage are processed. Specifically, we use gridded rain gauge data (Jones et al. 2009) to identify days with at least 1 mm of rain over two-thirds of the land portion of the overlap area. For each pair of temporally matched scans, the reflectivities and volumes of each bin pair are stored together with their spatial and temporal offsets.

This GR–GR comparison method is very similar to that used in the original version of the Radar Reflectivity Comparison Tool (RRCT; Gourley et al. 2003), which was developed for monitoring the relative calibration of radars in the US WSR-88D network. Tolerances in the RRCT were 500 m in horizontal distance, 50 m in vertical distance, 5 % in volume, and 3 minutes in time (between volume scans). Note, however, that the method was subsequently modified to use less stringent tolerances (750 m in distance and 6 minutes in time) while only considering bins within a rectangular region (120 km in length, 20 km in width, and 20 km in height) centred equidistant between the radars to ensure comparable bin volumes (http://rrct.nwc.ou.edu/). The latter approach would not work here because, unlike the WSR-88Ds, our three radars all have different beam widths (Table 1) and thus different bin volumes at a given range.

Figure 12 summarises the GR–GR comparison results using smoothed kernel density estimation violin plots (Hintze and Nelson 1998). Shown are the distributions of reflectivity difference computed with and without calibration adjustments, together with the sample size and Pearson correlation coefficients. Note that there are over a million samples for each radar pair, with nearly 10 million for the SYD–WOL comparison due to the close proximity of these sites (Fig. 1). It can be seen that the agreement of both the WOL and NEW radars with the SYD radar is improved, with smaller values of median difference, smaller IQRs, and higher correlation coefficients for the calibrated reflectivities (Fig. 12a,b). While the latter two changes are also seen in the WOL–NEW comparison, the median difference in this case actually increases in an absolute sense, from -0.1 to $+0.5$ dB. Taken alone, this would suggest that NEW reflectivities are being overcorrected and/or WOL reflectivities are being undercorrected. However, based on the SYD comparisons, we would expect a difference of only around 0.1 dB between these two radars after calibration adjustments. This discrepancy may be indicative of poorer volume matching between the WOL and NEW radars due to the large distance between them (Fig. 1).

4. Summary and outlook

In this paper, we have presented a method for estimating ground-based radar (GR) calibration errors through comparisons with spaceborne radar (SR) measurements from the TRMM and GPM satellites. This has been developed and tested using data from three Bureau of Meteorology (BoM) operational GRs in the vicinity of Sydney, Australia, for the period 2009–2015.

Spatially and temporally coincident GR and SR observations are first obtained using the volume-matching method (VMM) of SM11, which was originally developed to support ground validation efforts for GPM. Following Cao et al. (2013), a precipitation phase–dependent reflectivity correction is applied to the SR data to account for differences in measurement frequency (S-band for GRs, Ku-band for SRs). The resulting sample pairs are then filtered to isolate reflectivity differences associated with GR calibration error. Specifically, samples are only retained if they (a) predominantly comprise bins with reflectivities above the respective instrument sensitivities (18 dBZ for the SRs, 0 dBZ for the GRs), (b) are located in stratiform precipitation outside of the melting layer, and (c) have moderate reflectivities (24–36 dBZ) which are largely unaffected by the low SR sensitivity or attenuation of the SR beam. It was shown that the application of these criteria reduces the standard deviation of GR–SR reflectivity difference by around 2 dB.

Time series of the filtered GR–SR comparisons show periods of relatively stable GR calibration separated by sudden jumps of several dB. However, it is not possible to determine the precise date of these changes due to the low frequency of suitable satellite overpasses. In addition, residual noise in the comparisons, resulting from imperfect volume matching, makes it difficult to detect more subtle changes in calibration. To address these issues we make use of radar engineering work records maintained by the BoM. Dates of possible calibration changes are identified, between which the GR error is assumed to be constant. The calibration error for each period is then computed as the mean GR–SR reflectivity difference across all contemporaneous samples, using an iterative procedure to account for biases introduced by the reflectivity thresholding (Protat et al. 2011). This method produces results which are consistent with a subjective assessment of the time series while providing precise estimates of calibration error.

Since no ground truth exists to verify the accuracy of our calibration error estimates, we have examined the impact of correcting for these errors on the agreement between the three radars. Following the rationale behind the VMM, a method has been developed where spatially and temporally coincident GR sample volumes are identified and their reflectivities compared (c.f. Gourley et al. 2003). It was found that the calibration corrections in general lead to a robust improvement in the agreement between GRs, with an increase in correlation coefficients and a narrowing and shift towards zero of reflectivity difference distributions.

In the future it would be valuable to explore ways to further reduce the variability in GR–SR comparisons. One method, currently being investigated is to use quality indices to screen out GR samples that may be contaminated by ground clutter, anomalous propagation, or beam blockage (Crisologo et al. 2017). Screening could also be applied in cases where the orientation of the two radar beams leads to poor volume matching (e.g. Fig. 2c). The accuracy of the VMM would likely be further improved by accounting for non-standard GR beam refraction and the movement of precipitation features between SR and GR scans. Future work could also explore refinements to the Cao et al. (2013) frequency correction in the melting layer, with a view to eliminating the need to filter out volume-matched samples which fall within this layer.

In theory, the approach presented in this paper could be applied to any radar that falls within the coverage of the SRs ($\pm 35^\circ$ and $\pm 65^\circ$ during the TRMM and GPM eras, respectively). In practise, however, its potential is limited by the requirement for reliable engineering records, as these may not be available for many GR networks. Furthermore, since several months can pass between suitable satellite overpasses, the approach cannot be used for operational calibration monitoring. An alternative method which does not suffer from these issues is the relative calibration adjustment (RCA) technique (Silberstein et al. 2008; Wolff et al. 2015). This uses the statistical properties of ground clutter to provide a precise ($\pm$0.5 dB) measure of day-to-day variations in GR calibration relative to some baseline. The problem in this case is identifying an accurate baseline.

Clearly, the two techniques—SR comparison and RCA—are complementary. We are thus exploring the potential of applying them in tandem: using SR comparisons to set and periodically check the baseline reflectivity and the RCA to monitor and correct day-to-day fluctuations in calibration. This approach has already been successfully applied to 16 years of observations from the CPOL research radar in Darwin, Australia, and work is ongoing to incorporate it into operational radar quality control procedures at the BoM (Louf et al. 2017). Given the near-global coverage of GPM and the ubiquity of ground clutter, we believe that this approach has the potential to improve the accuracy and stability of GR calibration the world over.

Acknowledgments. RAW was funded by an Australian Research Council Linkage Project grant (LP130100679). We are grateful to Walt Petersen for insightful discussions on this work and to Bob Morris for providing details of the current VMM implementation. We also thank Mark Curtis for assistance with extracting the radar maintenance records and two anonymous reviewers for their comments.

APPENDIX

The VMM algorithm was coded up using Interactive Data Language (IDL) based on the descriptions in SM11 and Morris and Schwaller (2009), with modifications as detailed in section 2b. At the time of writing, work is ongoing to incorporate it into the wradlib radar analysis and visualization library for Python (Heistermann et al. 2013). Here we summarize the steps involved in creating the GR–SR comparison for a single SR overpass. The geometry of the two measurements is illustrated in Fig. A1.

The first step is to determine the location of each SR bin with respect to the GR. For this, we use an azimuthal equidistant projection centred on the GR. Each SR ray has an associated longitude–latitude pair corresponding to its intersection with the Earth ellipsoid (TRMM and GPM both use the WGS 84 ellipsoid). These are easily converted to Cartesian coordinates using standard map projection routines. To determine the full three-dimensional coordinates of each SR bin we must apply a parallax correction. The magnitude of the parallax error is

$$\Delta S = r_0 \sin\alpha, \tag{A1}$$

where r_0 is the range of the bin from the Earth ellipsoid and α is the local zenith angle of the ray (Fig. A1a). The parallax-corrected horizontal coordinates are then

$$x_\mathrm{s} = x_0 - \Delta S \cos\gamma, \tag{A2a}$$

$$y_\mathrm{s} = y_0 - \Delta S \sin\gamma. \tag{A2b}$$

Here x_0 and y_0 are the coordinates of the ellipsoid intersection ($z = 0$) and γ is the angle of the SR scan line (Fig. A1a). Finally, the height of the bin is computed as

$$z_\mathrm{s} = r_0 \cos\alpha \tag{A3}$$

Note that we do not account for the curvature of the Earth in these calculations. This is a reasonable approximation because ΔS is small (typically < 5 km).

In addition to the coordinates of each SR bin, we calculate their horizontal and vertical dimensions. The radius of a bin projected onto the horizontal plane is computed as the average of the projected radii the in along-track and cross-track directions (the latter varies with $\cos\alpha$):

$$R_\mathrm{s} = \tfrac{1}{2}\left(1+\cos\alpha\right) r_\mathrm{s} \tan\left(\omega_\mathrm{s}/2\right) \tag{A4}$$

where r_s is the SR range of the bin (Fig. A1a) and ω_s is the SR angular beam width (0.71° for both TRMM and GPM). The vertical depth of a bin is given by

$$D_s = \Delta r_s / \cos\alpha \tag{A5}$$

where Δr_s is the SR gate spacing (250 m for TRMM, 125 m for GPM).

The next step is to identify the nearest GR volume scan in time. Each SR scan has a unique timestamp; however, since it takes less than a minute for the satellite to traverse the GR field of view, a single time may be reasonably applied to all scans in the overpass. Specifically, we use the time corresponding to the closest point of approach to the GR, t_{cpa}. For the BoM radars, volume scans have a timestamp for every elevation sweep (corresponding to the start of that sweep), t_θ, but are named according to the start time of the entire scan, t_{vol}. Preliminary work indicated that the largest number of GR–SR matched volumes occurred around the third or forth elevation sweep or about $\delta t = 90$ seconds into the scan. Thus, to ensure the best temporal matching, we identify the scan which minimizes $|\Delta t_{\mathrm{vol}}| = \left|t_{\mathrm{vol}} + \delta t - t_{\mathrm{cpa}}\right|$ and only proceed if $|\Delta t_{\mathrm{vol}}| \leq 5$ min.

The Cartesian coordinates of the GR bins are next determined under the assumption of standard refraction; i.e. modelling the Earth as a sphere of equivalent radius $a_e = k_e a$, where $k_e = 4/3$ and a is the geocentric Earth radius at the latitude of the GR. The geometry illustrated in Fig. A1b leads to the follow simultaneous equations:

$$\left(a_e + z_g\right)\cos\left(S_g/a_e\right) = r_g \sin\theta_g + a_e + h, \tag{A6a}$$

$$\left(a_e + z_g\right)\sin\left(S_g/a_e\right) = r_g \cos\theta_g, \tag{A6b}$$

where h is the height of the GR antenna above the Earth ellipsoid, z_g is the height of the GR bin, S_g its horizontal distance from the radar, and r_g and θ_g are the GR range and elevation angle, respectively. Solving for S_{g}, we obtain

$$S_{\mathrm{g}} = a_{\mathrm{e}} \tan^{-1}\left(\frac{r_{\mathrm{g}}\cos\theta_{\mathrm{g}}}{r_{\mathrm{g}}\sin\theta_{\mathrm{g}} + a_{\mathrm{e}} + h}\right) \tag{A7}$$

from which the x and y coordinates can be determined as

$$x_{\mathrm{g}} = S_{\mathrm{g}}\cos\left(\pi/2 - \phi_{\mathrm{g}}\right), \tag{A8a}$$

$$y_{\mathrm{g}} = S_{\mathrm{g}}\sin\left(\pi/2 - \phi_{\mathrm{g}}\right), \tag{A8b}$$

where ϕ_{g} is the GR azimuth angle (Fig. A1b). Returning to A6 and solving for z_{g}, we find

$$z_{\mathrm{g}} = \sqrt{r_{\mathrm{g}}^2 + (a_{\mathrm{e}} + h)^2 + 2r_{\mathrm{g}}\,(a_{\mathrm{e}} + h)\sin\theta_{\mathrm{g}}} - a_{\mathrm{e}}, \tag{A9}$$

At this point, we have the coordinates of every SR and GR bin in a common reference frame. We now compute the median brightband height and width and apply a Ku-to-S band frequency correction to the SR data as described in section 2b. The volume matching then proceeds by looping first over SR rays and then over GR elevation sweeps. SR rays are only considered if they (a) contain precipitation (rainFlag = 20 for TRMM and flagPrecip = 1 for GPM; Table 3) and (b) are located between GR ranges of r_{min} and r_{max}. GR sweeps are only considered if $\Delta t = t_\theta - t_{\mathrm{cpa}} \leq$ 5 min. The steps involved in identifying a volume-matched GR–SR sample pair are as follows:

1. Calculate the GR elevation angle of each SR bin (using A6) as

$$\theta_{\mathrm{s}} = \tan^{-1}\left[\frac{\cos\left(S_{\mathrm{s}}/a_{\mathrm{e}}\right) - \left(a_{\mathrm{e}} + h\right)/\left(a_{\mathrm{e}} + z_{\mathrm{s}}\right)}{\sin\left(S_{\mathrm{s}}/a_{\mathrm{e}}\right)}\right] \tag{A10}$$

   where $S_{\mathrm{s}} = \sqrt{x_{\mathrm{s}}^2 + y_{\mathrm{s}}^2}$ is the horizontal distance of the SR bin from the GR.

2. Identify the SR bins that fall within the GR beam; i.e. for which $\theta_{\mathrm{g}} - \omega_{\mathrm{g}}/2 \leq \theta_{\mathrm{s}} \leq \theta_{\mathrm{g}} + \omega_{\mathrm{g}}/2$, where ω_{g} is the GR's angular beam width. Note the fraction of these, f_{s}, for which $Z_{\mathrm{s}} \geq Z_{\mathrm{s}}^*$.

3. Average the values of x_{s}, y_{s}, and z_{s} to get the coordinates of the sample centroid ($\bar{x}$, $\bar{y}$, and $\bar{z}$) and approximate its horizontal and vertical dimensions ($\bar{R}$ and $\bar{D}$) by the maximum R_{s} and

total D_s, respectively. Also determine the GR range of the sample (again using A6) as

$$\bar{r}_\mathrm{g} = \sqrt{(a_\mathrm{e}+\bar{z})^2 + (a_\mathrm{e}+h)^2 - 2\,(a_\mathrm{e}+\bar{z})\,(a_\mathrm{e}+h)\cos\left(\bar{S}/a_\mathrm{e}\right)}, \tag{A11}$$

where $\bar{S} = \sqrt{\bar{x}^2+\bar{y}^2}$ is the horizontal distance of the sample from the GR.

4. Linearly average the reflectivity values (in linear units, mm^6 m^{-3}) for which $Z_\mathrm{s} \geq Z_\mathrm{s}^*$ to get the SR reflectivity of the matched volume, $\bar{Z}_\mathrm{s}$. Do this for both raw and frequency-corrected reflectivities.

5. Identify the GR bins that fall within the footprint of the SR beam; i.e. for which $d \leq \bar{R}$, where $d = \sqrt{\left(x_\mathrm{g}-\bar{x}\right)^2 + \left(y_\mathrm{g}-\bar{y}\right)^2}$. Note the fraction of these, f_g, for which $Z_\mathrm{g} \geq Z_\mathrm{g}^*$.

6. Average the reflectivity values (in linear units, mm^6 m^{-3}) for which $Z_\mathrm{g} \geq Z_\mathrm{g}^*$, weighting bins inversely by d (using a Barnes Gaussian function with radius $\bar{R}$) and linearly by r_g^2 (proportional to the bin volume), to get the GR reflectivity of the matched volume, $\bar{Z}_\mathrm{g}$.

For every SR overpass, a single file is produced containing data for all volume-matched samples. The variables stored for each sample are as follows:

- Cartesian coordinates ($\bar{x}$, $\bar{y}$, $\bar{z}$);
- volume dimensions ($\bar{R}$, $\bar{D}$);
- GR range ($\bar{r}_\mathrm{g}$);
- averaged SR and GR reflectivities ($\bar{Z}_\mathrm{s}(\mathrm{Ku})$, $\bar{Z}_\mathrm{s}(\mathrm{S})$, $\bar{Z}_\mathrm{g}$);
- fraction of SR and GR bins above the respective minimum reflectivity thresholds (f_s, f_g);
- precipitation type index ($P = 1$ for stratiform, $P = 2$ for convective, $P = 3$ for other);
- GR–SR time difference (Δt).

The file also contains the median brightband height and width for the overpass.

It should be noted that the quality parameters listed in Table 3 are used at various stages of the algorithm to ensure that all matched samples are accurate. Specifically, SR scans are rejected if dataQuality $\neq 0$ and SR rays are rejected if status ≥ 100 for TRMM and if qualityBB > 1 and/or qualityTypePrecip > 1 for GPM.

References

Amitai, E., X. Llort, and D. Sempere-Torres, 2009: Comparison of TRMM radar rainfall estimates with NOAA next-generation QPE. *J. Meteor. Soc. Japan*, **87**, 109–118.

Anagnostou, E. N., C. A. Morales, and T. Dinku, 2001: The use of TRMM precipitation radar observations in determining ground radar calibration biases. *J. Atmos. Oceanic Technol.*, **18**, 616–628.

Atlas, D., 2002: Radar calibration: Some simple approaches. *Bull. Amer. Meteor. Soc.*, **83**, 1313–1316.

Awaka, J., T. Iguchi, and K. Okamoto, 2007: Rain type classification algorithm. *Measuring Precipitation From Space*, V. Levizzani, P. Bauer, and F. J. Turk, Eds., Springer, 213–224.

Awaka, J., T. Iguchi, and K. Okamoto, 2009: TRMM PR standard algorithm 2A23 and its performance on bright band detection. *J. Meteor. Soc. Japan*, **87**, 31–52.

Bolen, S. M., and V. Chandrasekar, 2003: Methodology for aligning and comparing spaceborne radar and ground-based radar observations. *J. Atmos. Oceanic Technol.*, **20**, 647–659.

Cao, Q., Y. Hong, Y. Qi, Y. Wen, J. Zhang, J. J. Gourley, and L. Liao, 2013: Empirical conversion of the vertical profile of reflectivity from Ku-band to S-band frequency. *J. Geophys. Res.*, **118**, 1814–1825.

Chandrasekar, V., L. Baldini, N. Bharadwaj, and P. L. Smith, 2015: Calibration procedures for Global Precipitation-Measurement ground-validation radars. *Radio Science Bull.*, **88**, 45–73.

Chen, S., and Coauthors, 2013: Evaluation of spatial errors of precipitation rates and types from TRMM spaceborne radar over the southern CONUS. *J. Hydrometeor.*, **14**, 1884–1896.

Crisologo, I., R. A. Warren, K. Muehlbauer, and M. Heistermann, 2017: Using data quality to improve comparison between GPM measurements and ground radars. *38th Conf. on Radar Meteorology, Chicago, IL*, Amer. Meteor. Soc.

Gourley, J. J., B. Kaney, and R. A. Maddox, 2003: Evaluating the calibration of radars: A software approach. *38th Conf. on Radar Meteorology, Seattle, WA*, Amer. Meteor. Soc., available online at https://ams.confex.com/ams/32BC31R5C/techprogram/paper_64171.htm.

Heistermann, M., S. Jacobi, and T. Pfaff, 2013: An open source library for processing weather radar data (wradlib). *Hydrology and Earth System Sciences*, **17**, 863–871.

Hintze, J. L., and R. D. Nelson, 1998: Violin plots: A box plot–density trace synergism. *The American Statistician*, **52**, 181–184.

Hitschfeld, W., and J. Bordan, 1954: Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *J. Meteor.*, **11**, 58–67.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722.

Iguchi, T., S. Seto, R. Meneghini, N. Yoshida, J. Awaka, M. Le, V. Chandrasekar, and T. Kubota, 2017: *GPM/DPR Level-2 Algorithm Theoretical Basis Document*. Avaliable online at http://www.eorc.jaxa.jp/GPM/doc/algorithm/ATBD_DPR_201708_whole_1.pdf.

Jones, D. A., W. Wang, and R. Fawcett, 2009: High-quality spatial climate data-sets for Australia. *Aust. Meteor. and Oceanogr. J.*, **58**, 233–248.

Kawanishi, T., and Coauthors, 2000: TRMM precipitation radar. *Adv. Space Res.*, **25**, 969–972.

Kim, J.-H., M.-L. Ou, J.-D. Park, K. R. Morris, M. R. Schwaller, and D. B. Wolff, 2014: Global precipitation measurement (GPM) ground validation (GV) prototype in the Korean Peninsula. *J. Atmos. Oceanic Technol.*, **31**, 1902–1921.

Kirstetter, P.-E., Y. Hong, J. Gourley, M. Schwaller, W. Petersen, and J. Zhang, 2013: Comparison of TRMM 2A25 products, version 6 and version 7, with NOAA/NSSL ground radar–based national mosaic QPE. *J. Hydrometeor.*, **14**, 661–669.

Köck, K., T. Leltne, W. Randeu, M. Divjak, and K.-J. Schrelber, 2000: OPERA: operational programme for the exchange of weather radar information. First results and outlook for the future. *Phys. Chem. Earth*, **25**, 1147–1151.

Liao, L., and R. Meneghini, 2009a: Changes in the TRMM version-5 and version-6 precipitation radar products due to orbit boost. *J. Meteor. Soc. Japan*, **87**, 93–107.

Liao, L., and R. Meneghini, 2009b: Validation of TRMM precipitation radar through comparison of its multiyear measurements with ground-based radar. *J. Appl. Meteor. Climatol.*, **48**, 804–817.

Liao, L., R. Meneghini, and T. Iguchi, 2001: Comparisons of rain rate and reflectivity factor derived from the TRMM precipitation radar and the WSR-88D over the Melbourne, Florida, site. *J. Atmos. Oceanic Technol.*, **18**, 1959–1974.

Louf, V., A. Protat, C. Jakob, S. Rauniyar, and R. A. Warren, 2017: The relative calibration adjustment technique for calibrating Australian operational radars in near real-time. *38th Conf. on Radar Meteorology, Chicago, IL*, Amer. Meteor. Soc.

Meneghini, R., T. Iguchi, T. Kozu, L. Liao, K. Okamoto, J. A. Jones, and J. Kwiatkowski, 2000: Use of the surface reference technique for path attenuation estimates from the TRMM precipitation radar. *J. Appl. Meteor.*, **39**, 2053–2070.

Meneghini, R., J. Jones, T. Iguchi, K. Okamoto, and J. Kwiatkowski, 2004: A hybrid surface reference technique and its application to the TRMM precipitation radar. *J. Atmos. Oceanic Technol.*, **21**, 1645–1658.

Michelson, D. B., R. Lewandowski, M. Szewczykowski, H. Beekhuis, and G. Haase, 2014: EUMETNET OPERA weather radar information model for implementation with the HDF5 file format, version 2.2. Tech. rep., EUMETNET OPERA, 42 pp. Available online at http://eumetnet.eu/wp-content/uploads/2017/01/OPER_hdf_description_2014.pdf.

Morris, K. R., and M. R. Schwaller, 2009: An enhanced Global Precipitation Measurement (GPM) validation network prototype. *34th Conf. on Radar Meteorology, Williamsburg, VA*, Amer. Meteor. Soc., available online at https://ams.confex.com/ams/pdfpapers/155254.pdf.

Morris, K. R., and M. R. Schwaller, 2011: Sensitivity of spaceborne and ground radar comparison results to data analysis methods and constraints. *35th Conf. on Radar Meteorology, Pittsburgh, PA*, Amer. Meteor. Soc., avalaible online at https://ams.confex.com/ams/35Radar/webprogram/Manuscript/Paper191729/AMS_35th_Radar_paper_final.pdf.

NASA, 2014: *Tropical Rainfall Measuring Mission Precipitation Processing System File Specification 2A23 (Version 7)*. Avaliable online at https://storm-pps.gsfc.nasa.gov/storm/data/docs/filespec.TRMM.V7.2A23.pdf.

NASA, 2015: *Tropical Rainfall Measuring Mission Precipitation Processing System File Specification 2A25 (Version 7)*. Avaliable online at https://storm-pps.gsfc.nasa.gov/storm/data/docs/

filespec.TRMM.V7.2A25.pdf.

NASA, 2016: *Global Precipitation Measurement Precipitation Processing System File Specification 2AKu (Preliminary Version)*. Avaliable online at https://storm.pps.eosdis.nasa.gov/storm/data/docs/filespec.GPM.V1.2AKu.pdf.

NASA, 2017: *Release notes for the PR Level 1 products*. Avaliable online at https://pps.gsfc.nasa.gov/Documents/ReleaseNote_PU1_productV05.pdf.

Park, S., S.-H. Jung, and G. Lee, 2015: Cross validation of TRMM PR reflectivity profiles using 3D reflectivity composite from the ground-based radar network over the Korean Peninsula. *J. Hydrometeor.*, **16**, 668–687.

Protat, A., D. Bouniol, E. OConnor, H. Klein Baltink, J. Verlinde, and K. Widener, 2011: CloudSat as a global radar calibrator. *J. Atmos. Oceanic Technol.*, **28**, 445–452.

Rasmussen, K. L., S. L. Choi, M. D. Zuluaga, and R. A. Houze, 2013: TRMM precipitation bias in extreme storms in South America. *Geophys. Res. Lett.*, **40**, 3457–3461.

Rennie, S., 2012: Doppler weather radar in Australia. Tech. Rep. 055, Centre for Australian Weather and Climate Research, 52 pp. Available online at http://www.cawcr.gov.au/technical-reports/CTR_055.pdf.

Schwaller, M. R., and K. R. Morris, 2011: A ground validation network for the Global Precipitation Measurement mission. *J. Atmos. Oceanic Technol.*, **28**, 301–319.

Silberstein, D. S., D. B. Wolff, D. A. Marks, D. Atlas, and J. L. Pippitt, 2008: Ground clutter as a monitor of radar stability at Kwajalein, RMI. *J. Atmos. Oceanic Technol.*, **25**, 2037–2045.

Simpson, J., C. Kummerow, W.-K. Tao, and R. F. Adler, 1996: On the tropical rainfall measuring mission (TRMM). *Meteor. Atmos. Phys.*, **60**, 19–36.

Takahashi, N., H. Kuroiwa, and T. Kawanishi, 2003: Four-year result of external calibration for Precipitation Radar (PR) of the Tropical Rainfall Measuring Mission (TRMM) satellite. *IEEE Trans. Geosci. Remote Sens.*, **41**, 2398–2403.

Toyoshima, K., H. Masunaga, and F. A. Furuzawa, 2015: Early evaluation of Ku- and Ka-band sensitivities for the Global Precipitation Measurement (GPM) dual-frequency precipitation radar (DPR). *SOLA*, **11**, 14–17.

Villarini, G., and W. F. Krajewski, 2010: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surv. Geophys.*, **31**, 107–129.

Wang, J., and D. B. Wolff, 2009: Comparisons of reflectivities from the TRMM precipitation radar and ground-based radars. *J. Atmos. Oceanic Technol.*, **26**, 857–875.

Wolff, D. B., and B. L. Fisher, 2008: Comparisons of instantaneous TRMM ground validation and satellite rain-rate estimates at different spatial scales. *J. Appl. Meteor. Climatol.*, **47**, 2215–2237.

Wolff, D. B., D. A. Marks, and W. A. Petersen, 2015: General application of the relative calibration adjustment (RCA) technique for monitoring and correcting radar reflectivity calibration. *J. Atmos. Oceanic Technol.*, **32**, 496–506.

LIST OF TABLES

TABLE 1. Characteristics of the three radars used in this study. Symbols have the following meaning: λ = wavelength, ω = angular beam width, Δr = range gate spacing, $\Delta\phi$ = angular beam spacing, and n_θ = number of elevation angles. Note that the WOL radar was replaced in early 2011, with associated changes in λ, ω, and Δr. The volume scan pattern was also updated at this time, making it consistent with the SYD radar. The same change was applied to the NEW radar in mid 2013 following an upgrade from an analogue to a digital receiver.

Site	Make	λ (cm)	ω (°)	Δr (m)	$\Delta\phi$ (°)	n_θ	Volume scan elevation angles (°)
SYD	Meteor-1500S	10.0	1.0	250	1.0	14	0.5, 0.9, 1.3, 1.8, 2.4, 3.1, 4.2, 5.6, 7.4, 10.0, 13.3, 17.9, 23.9 32.0
WOL	WSR-74S	10.4	1.9	1000	1.0	15	0.5, 1.2, 1.9, 2.7, 3.5, 4.7, 6.0, 7.5, 9.2, 11.0, 13.0, 16.0, 20.0, 25.0, 32.0
	DWSR-8502S	10.0	2.0	500	1.0	14	0.5, 0.9, 1.3, 1.8, 2.4, 3.1, 4.2, 5.6, 7.4, 10.0, 13.3, 17.9, 23.9 32.0
NEW	WSR-74S	10.4	1.9	500	1.0	15	0.5, 0.8, 1.1, 1.4, 1.9, 2.5, 3.3, 4.4, 5.8, 7.7, 10.3, 13.6, 18.1, 24.1, 32.0
	DWSR-74S	10.4	1.9	500	1.0	14	0.5, 0.9, 1.3, 1.8, 2.4, 3.1, 4.2, 5.6, 7.4, 10.0, 13.3, 17.9, 23.9 32.0

TABLE 2. Characteristics of the TRMM PR and GPM KuPR. Symbols have the following meaning: z_o = orbital altitude, λ = wavelength, ω = angular beam width, Δr = range gate spacing, $\Delta\Phi$ = angular beam spacing, and Φ_{max} = maximum off-nadir scan angle. Note that prior to August 2001, the TRMM orbital altitude was 350 km.

Instrument	z_o (km)	λ (cm)	ω (°)	Δr (m)	$\Delta\Phi$ (°)	Φ_{max} (°)
TRMM PR	402.5	2.2	0.71	250	0.71	17.04
GPM KuPR	407.0	2.2	0.71	125	0.71	17.04

TABLE 3. TRMM and GPM parameters extracted for our analysis. Detailed descriptions of these parameters can be found in the file specification documents for the Version 7 TRMM 2A23 and 2A25 products (NASA 2014, 2015) and Version 5 GPM 2AKu products (NASA 2016). The array dimensions n_{scan}, n_{ray}, and n_{bin} correspond to the along-track, cross-track, and range directions, respectively. For both radars $n_{\text{ray}} = 49$, while $n_{\text{bin}} = 80$ for TRMM and 176 for GPM. The value of n_{scan} varies depending on the distance and angle of the GR site overpass.

Satellite	Product	Parameter	Description	Dimensions
TRMM	2A23	dataQuality	Quality index for scan data	n_{scan}
		rainFlag	Flag indicating likelihood of surface precipitation in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		rainType	Classification of precipitation in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		HBB	Height of bright band (if present) in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		BBwidth	Width of bright band (if present) in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		status	Quality index for 2A23 products	$n_{\text{scan}} \times n_{\text{ray}}$
	2A25	scLocalZenith	Zenith angle of ray at Earth ellipsoid	$n_{\text{scan}} \times n_{\text{ray}} \times n_{\text{bin}}$
		correctZFactor	Attenuation-corrected reflectivity	$n_{\text{scan}} \times n_{\text{ray}} \times n_{\text{bin}}$
GPM	2AKu	dataQuality	Quality index for scan data	n_{scan}
		localZenithAngle	Zenith angle of ray at Earth ellipsoid	$n_{\text{scan}} \times n_{\text{ray}}$
		flagPrecip	Flag indicating presence of precipitation in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		heightBB	Height of bright band (if present) in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		widthBB	Width of bright band (if present) in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		qualityBB	Quality information for bright band products	$n_{\text{scan}} \times n_{\text{ray}}$
		typePrecip	Classification of precipitation in ray	$n_{\text{scan}} \times n_{\text{ray}}$
		qualityTypePrecip	Quality index for precipitation type product	$n_{\text{scan}} \times n_{\text{ray}}$
		zFactorCorrected	Attenuation-corrected reflectivity	$n_{\text{scan}} \times n_{\text{ray}} \times n_{\text{bin}}$

TABLE 4. Statistics of GR–SR comparisons under different sample filtering criteria (see text for definitions): n = sample size, $\overline{\Delta Z}$ = mean reflectivity difference (dB), and $\sigma_{\Delta Z}$ = standard deviation of reflectivity difference (dB).

Site	Stat.	None	A	B	C	All
SYD	n	922407	544081	405144	306244	75941
	$\overline{\Delta Z}$	−1.9	−0.8	−2.2	+0.1	0.0
	$\sigma_{\Delta Z}$	4.5	3.7	4.2	2.6	2.1
WOL	n	901591	372661	262531	299575	33608
	$\overline{\Delta Z}$	−2.2	−0.7	−2.7	−0.2	0.0
	$\sigma_{\Delta Z}$	4.4	3.3	4.4	2.7	2.1
NEW	n	851335	347115	271858	305187	33997
	$\overline{\Delta Z}$	−1.6	−0.5	−1.7	0.0	0.0
	$\sigma_{\Delta Z}$	4.0	3.1	3.8	2.6	2.0

LIST OF FIGURES

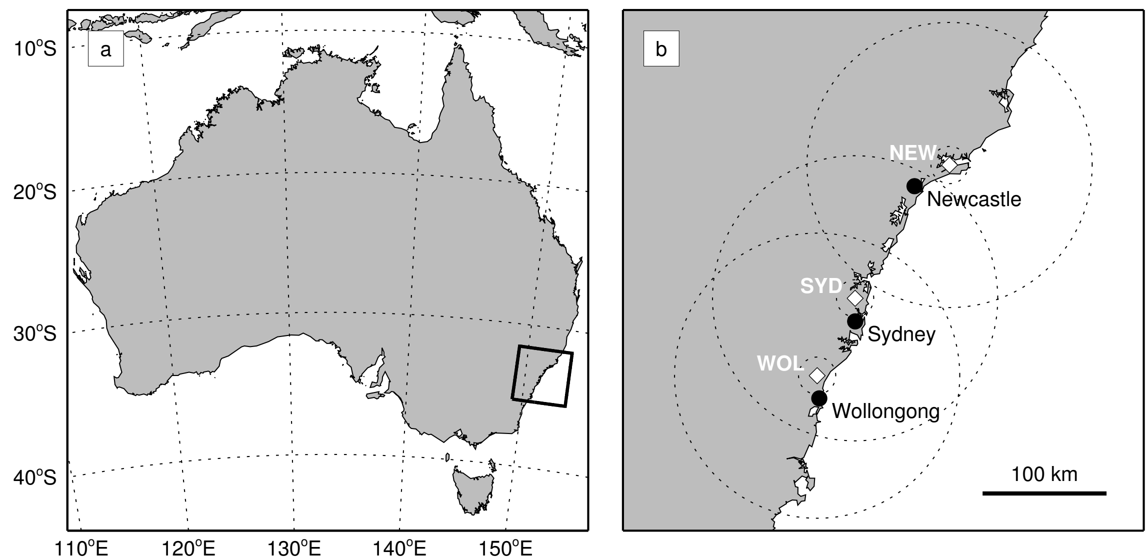


FIG. 1. (a) Map of Australia showing the location of the region of interest (black box). (b) Map of the region of interest showing the locations of the three radars (white diamonds) and nearby major cities (black circles). Dotted lines in (b) show the 15 km and 115 km range rings around each radar.

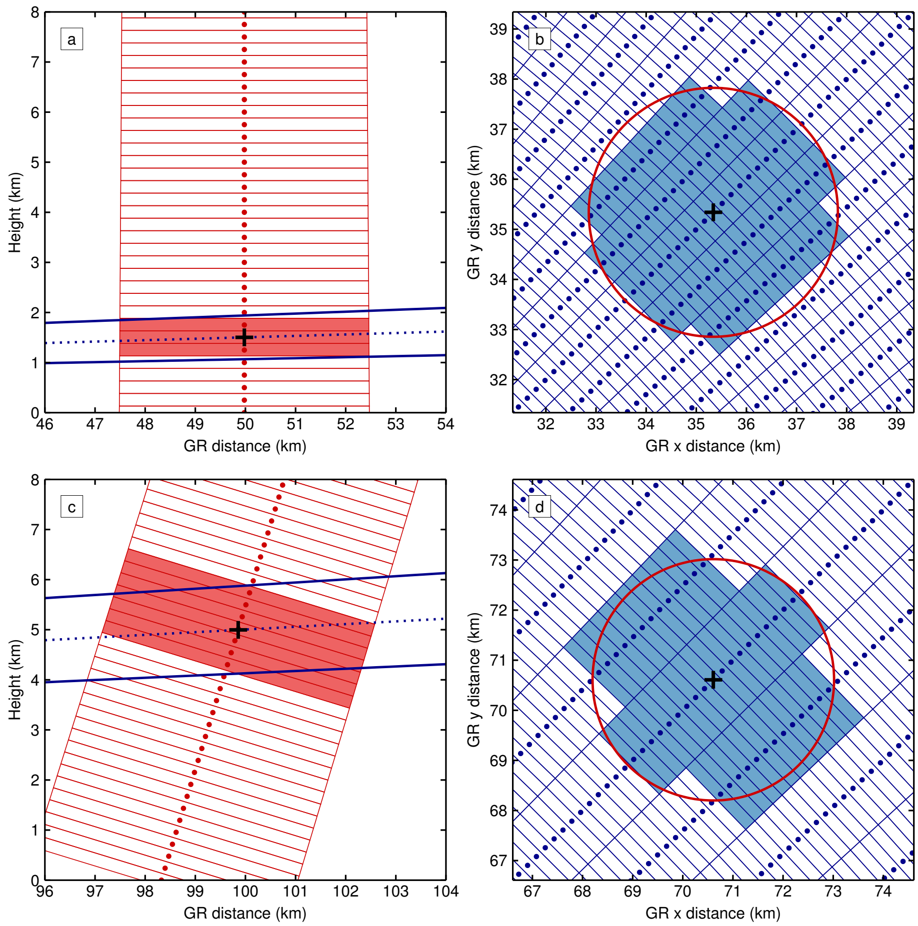


FIG. 2. Illustration of the volume-matching method. Top panels (a,b) show a GR–SR intersection at a GR range of 50 km, azimuth angle of 45°, and elevation angle of 1.3°, and an SR scan angle of 0° (nadir scan). Bottom panels (c,d) show a GR–SR intersection at a GR range of 100 km, azimuth angle of 45°, and elevation angle of 2.4°, and an SR scan angle of 17.04° (maximum off-nadir scan). In both cases, the GR has a range gate spacing of 250 m and an angular beam width of 1.0°, the SR has a range gate spacing of 250 m, and the SR scan line is perfectly aligned with the intersecting GR azimuth. Left panels (a,c) show vertical cross sections along the SR scan line, with the GR beam boundaries and centre indicated by thick solid and dotted blue lines, respectively, and the SR bin boundaries and centres indicated by thin red lines and red dots, respectively. SR bins contributing to the volume-matched reflectivity are shaded. Right panels (b,d) show horizontal cross sections at the height of the intersection, with the GR bin boundaries and centres indicated by thin blue lines and blue dots, respectively, and the horizontal projection of the SR beam boundary approximated by the red circle. GR bins contributing to the volume-matched reflectivity are shaded. In all panels the point of intersection is indicated with a black cross.

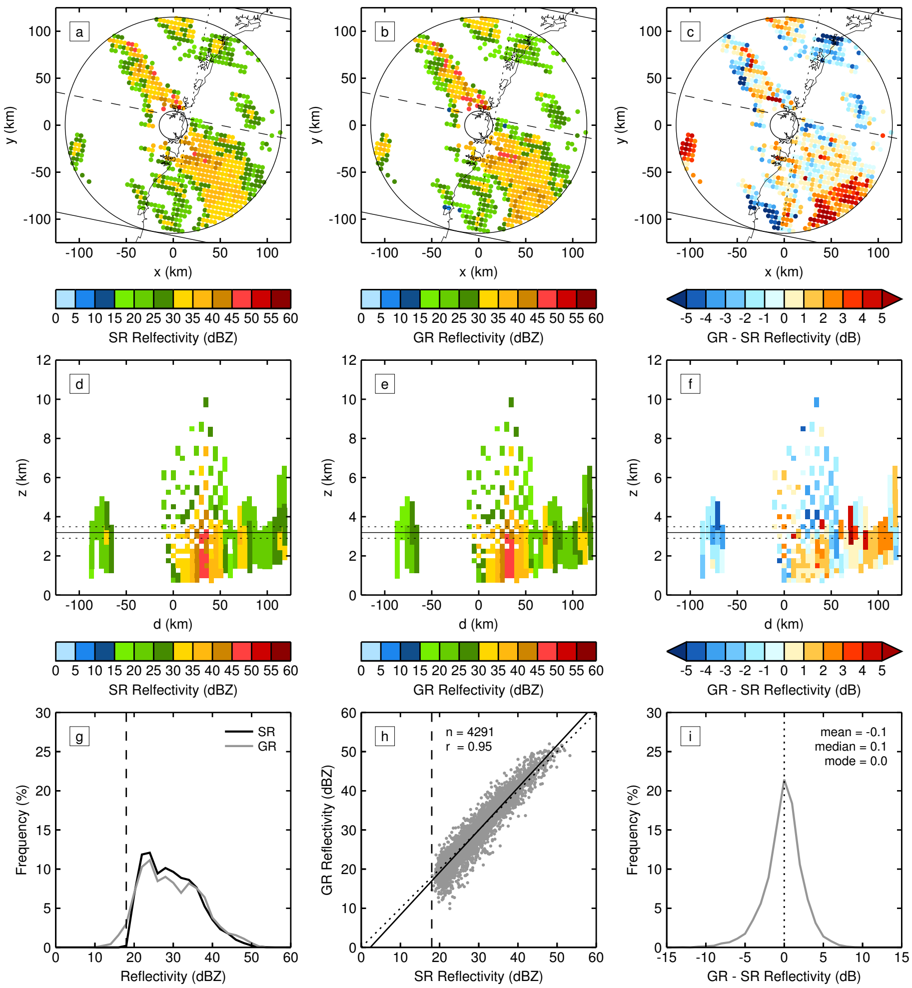
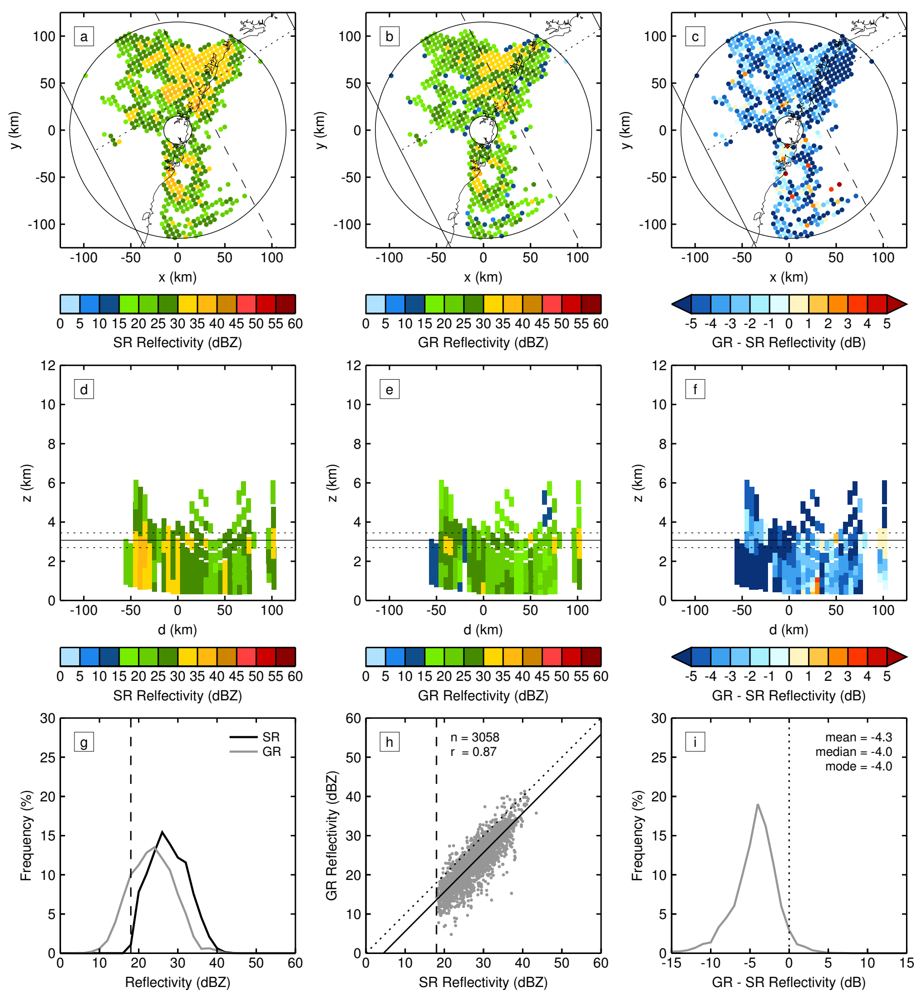


FIG. 3. Example of GR–SR comparison between TRMM and the SYD radar on 22/10/2013. The top row shows plan views of (a) SR reflectivity, (b) GR reflectivity, and (c) GR – SR reflectivity for the 1.3° elevation sweep. Each filled circle shows an individual volume-matched sample. Rings show the minimum and maximum GR range, while solid and dashed lines show, respectively, the boundaries and centre of the SR swath. The middle row shows vertical cross sections taken across the SR swath (location indicated by dotted lines in top panels) of (d) SR reflectivity, (e) GR reflectivity, and (f) GR – SR reflectivity. Here volume-matched samples are shown as columns of varying depth (note that these overlap for low elevation sweeps). Solid and dotted lines show, respectively, the centre and boundaries of the brightband. The bottom row summarises the statistics of the volume-matched sample pairs: (g) histogram of SR (black) and GR (grey) reflectivities (2 dB bins); (h) scatter plot of paired SR and GR reflectivities; (i) histogram of GR – SR reflectivity (1 dB bins). Dashed vertical lines in (g) and (h) show the minimum SR reflectivity. Solid and dotted lines in (h) show the line of best fit and one-to-one line, respectively, with the sample size, *n*, and Pearson correlation coefficient, *r*, given at the top. Mean, median, and modal GR bias are indicated in the top-right corner of (i).



957 FIG. 4. As in Fig. 3, but showing a comparison between GPM and the SYD radar on 27/01/2015. In this case
958 the 0.5° elevation sweep is shown in the top row.

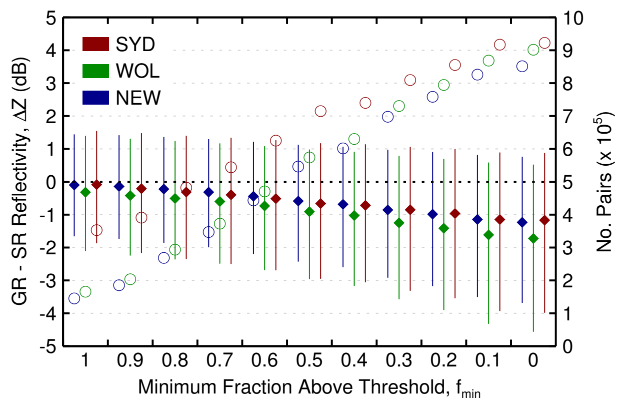


FIG. 5. GR–SR reflectivity difference plotted as a function of the minimum fraction of SR and GR bins with reflectivity values above their respective thresholds. Filled diamonds and vertical lines show, respectively, the median and interquartile range for each $f_{\min}$ bin, while open circles indicate the sample size (scale on right y axis). Values for the SYD, WOL, and NEW radars are shown in red, green, and blue, respectively.

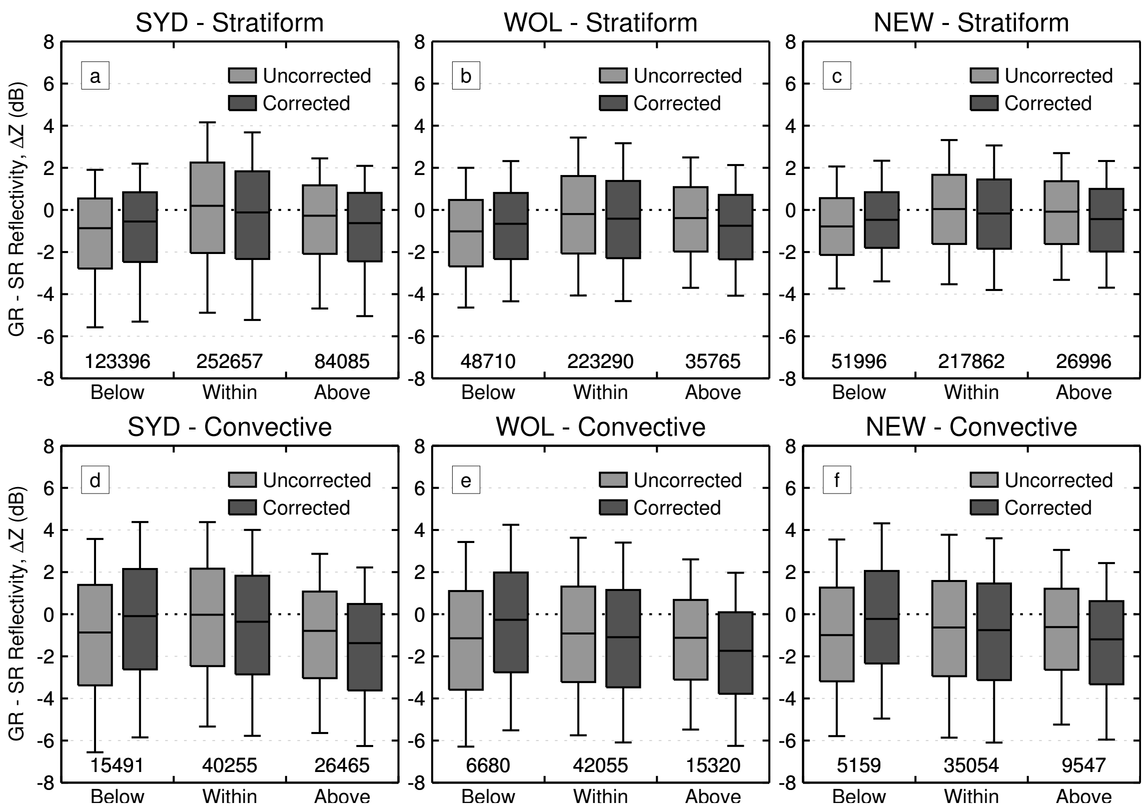


FIG. 6. Box-and-whisker plots showing GR–SR reflectivity difference for samples below, within, and above the ML in (top) stratiform and (bottom) convective precipitation for the (left) SYD, (middle) WOL, and (right) NEW radars. Boxes show the median and interquartile range of the distribution; whiskers show the 10th and 90th percentiles. Dark and light grey boxes show values computed using SR data with and without the Ku-to-S-band frequency correction, respectively. The number of samples in each layer is given at the bottom of the panels. Note that all samples whose volume overlaps vertically with the ML are classified as being within it.

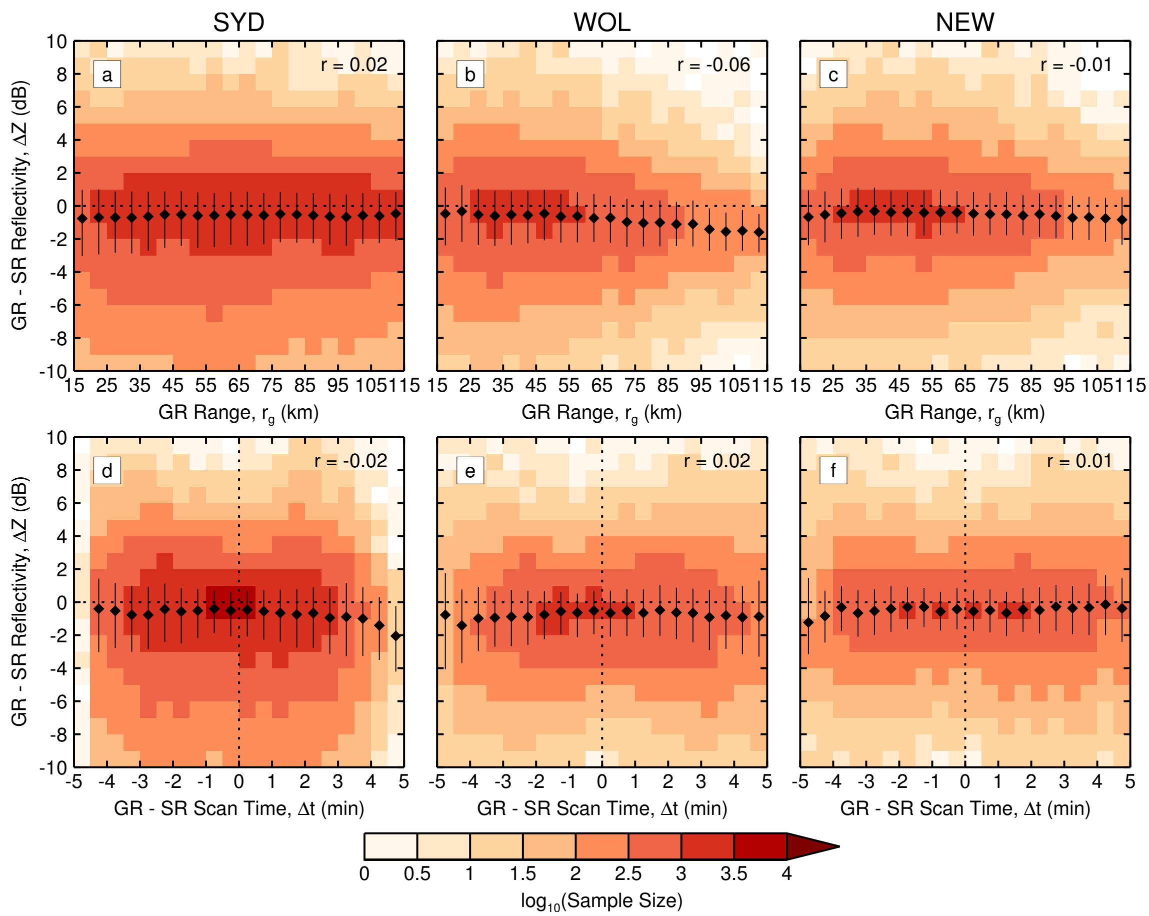
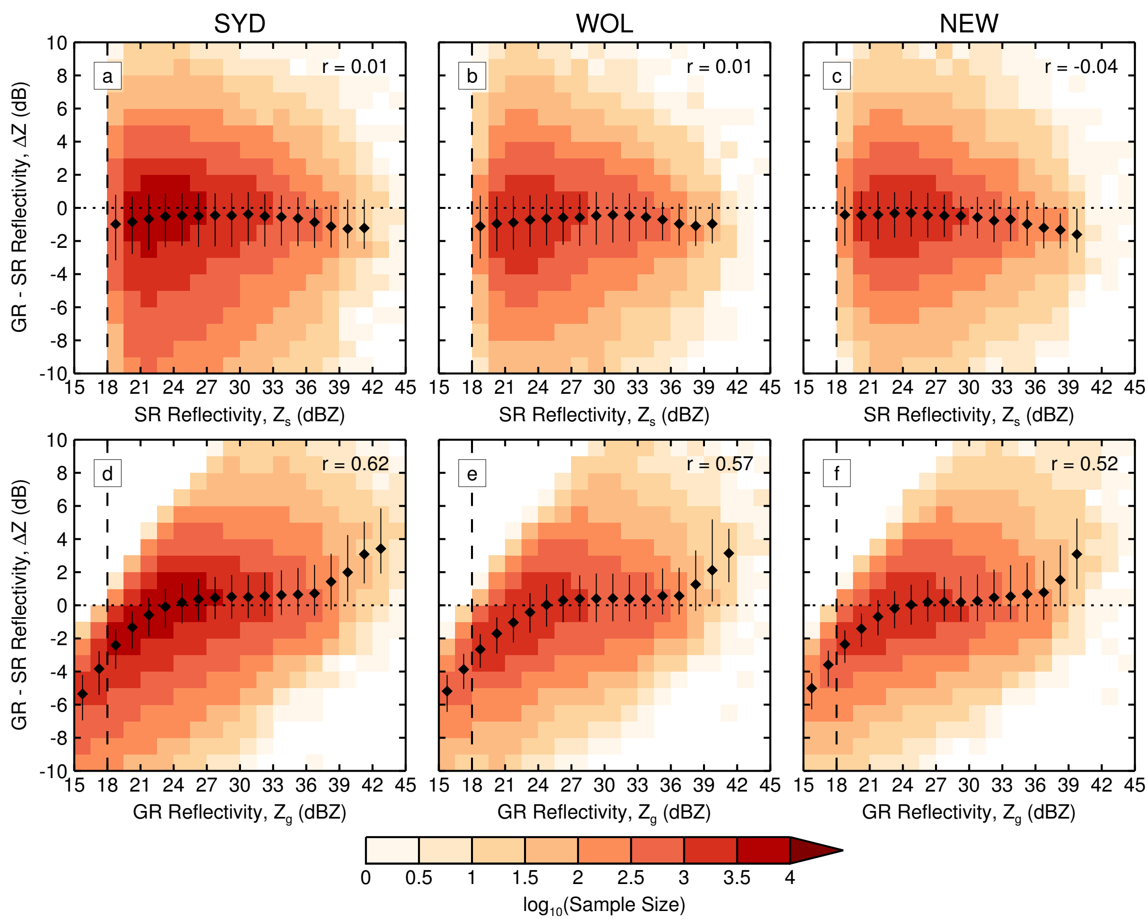
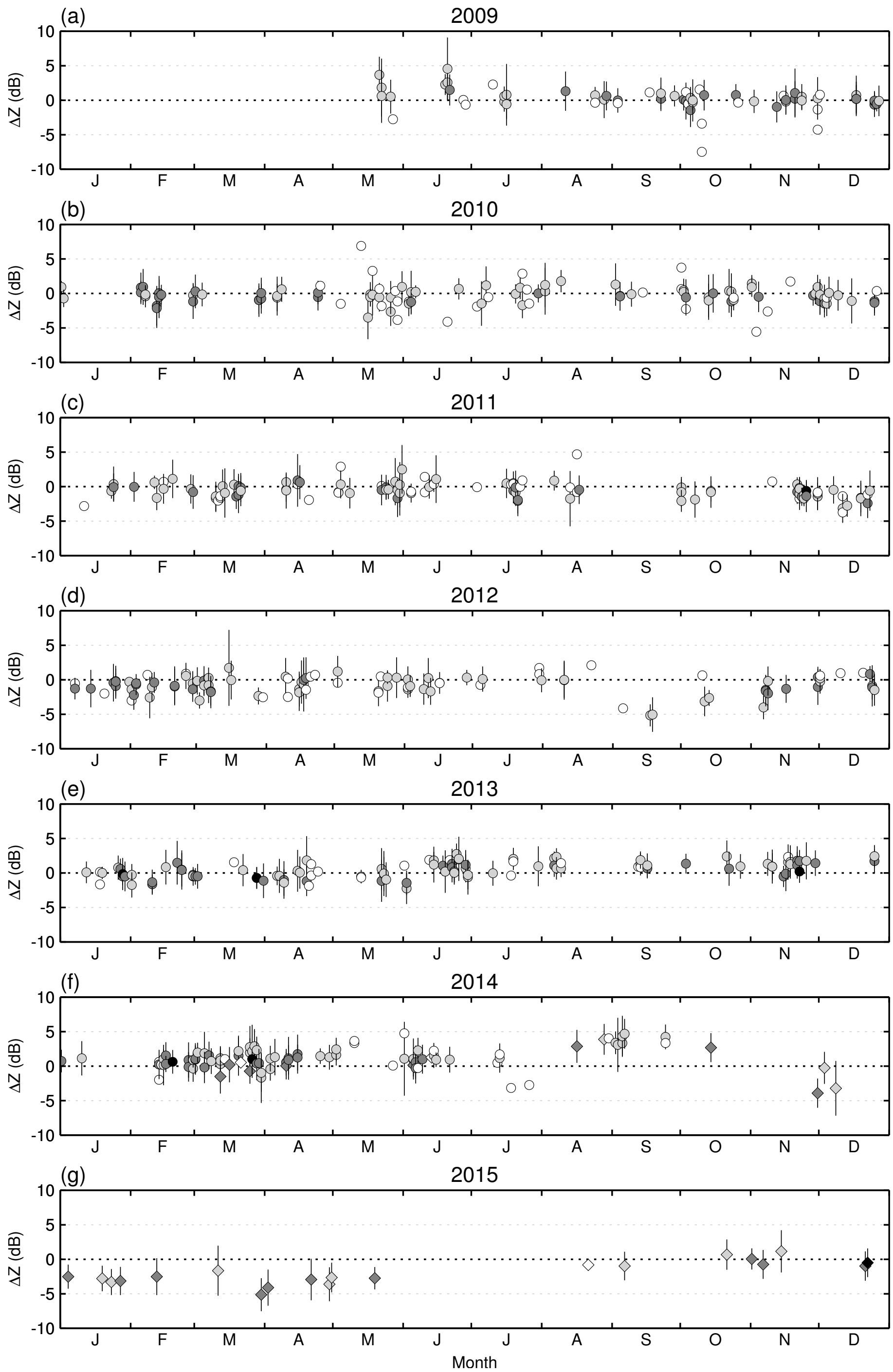


FIG. 7. Bivariate histograms showing GR–SR reflectivity difference as a function of (top) GR range and (bottom) GR–SR time difference for the (left) SYD, (middle) WOL, and (right) NEW radars. Colors show the sample size for each bin on a logarithmic scale. Black diamonds and vertical lines show, respectively, the median and interquartile range for each x-axis bin (not shown for sample sizes < 100). Pearson correlation coefficients, r, are given in the top-right corner of each panel.



974 FIG. 8. As in Fig. 7 but showing GR–SR reflectivity difference as a function of (top) SR reflectivity and

975 (bottom) GR reflectivity. Dashed lines indicate the SR sensitivity.



976 FIG. 9. Annual time series of GR–SR reflectivity difference, ΔZ, for the SYD radar. Symbols and thin vertical lines show, respectively, the mean GR bias and its standard deviation for each SR overpass. Circles and diamonds indicate comparisons with TRMM and GPM, respectively. Symbols are colored according to the number of volume-matched sample pairs on a logarithmic scale: white = 1–9, light grey = 10–99, dark grey = 100-999, and black = 1000+. Standard deviations are not shown for sample sizes < 10.

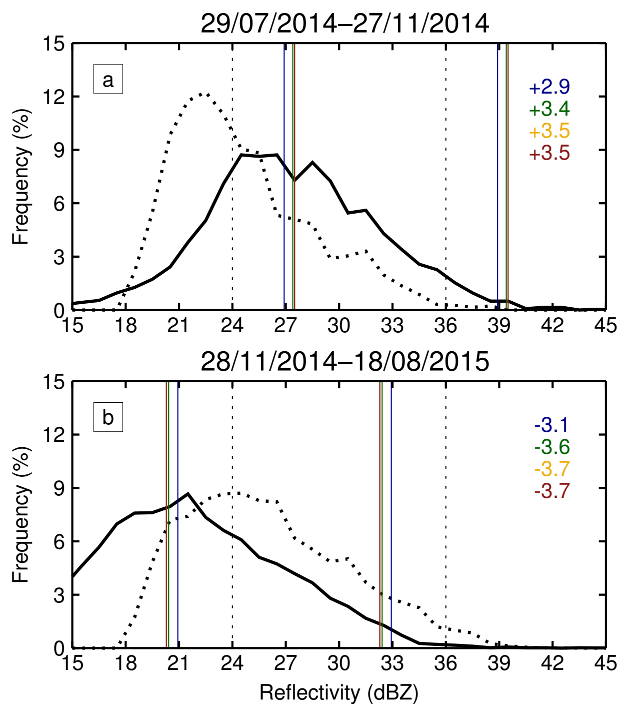


FIG. 10. Histograms showing SR (thick dotted line) and SYD GR (thick solid line) reflectivity distributions for (a) 29/07/2014–27/11/2014 and (b) 28/11/2014–18/01/2015. Thin vertical lines bound the portion of each histogram used in the comparison; dotted for the SR and solid for the GR with colors indicating the iteration step (blue = 1, green = 2, yellow = 3, and red = 4). The corresponding calibration error estimates are given in the top right corner of each panel.

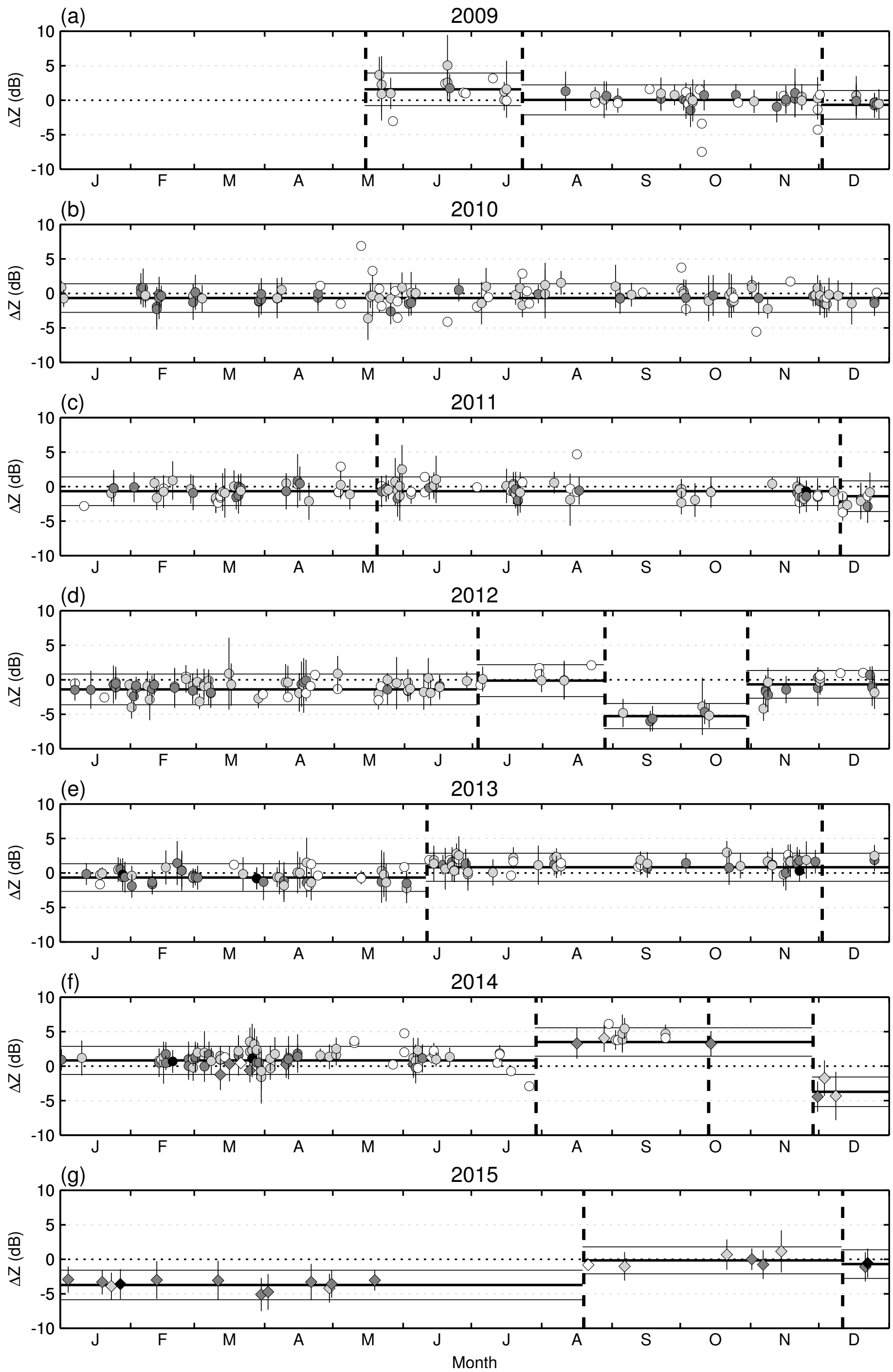


FIG. 11. As in Fig. 9 but following the iterative calculation of calibration error. Thick dashed vertical lines show the dates of possible calibration changes; thick and thin horizontal lines show, respectively, the mean calibration error and its standard deviation during the intervening periods.

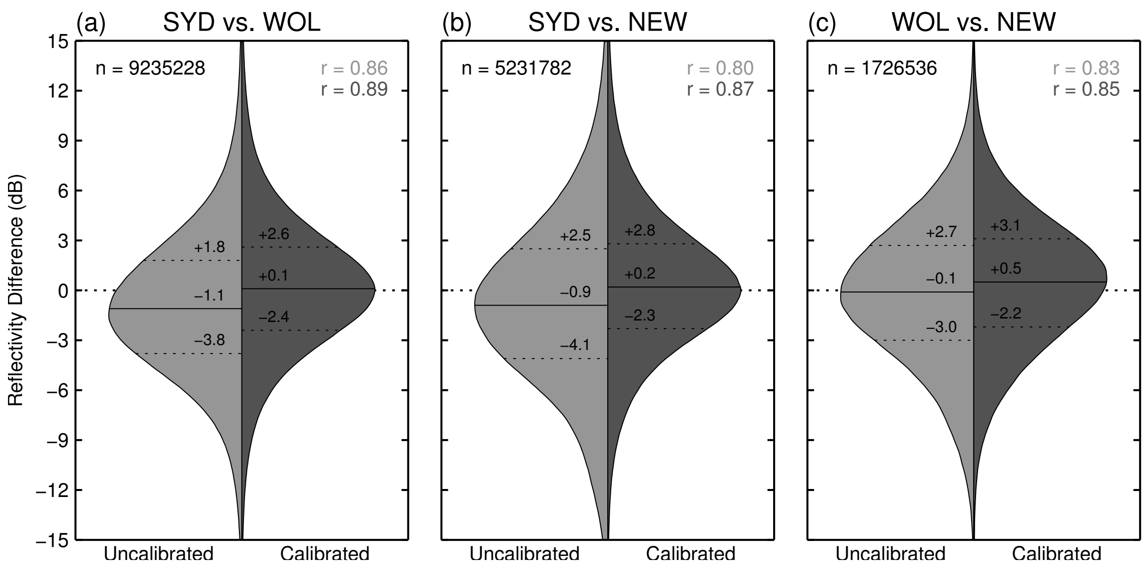
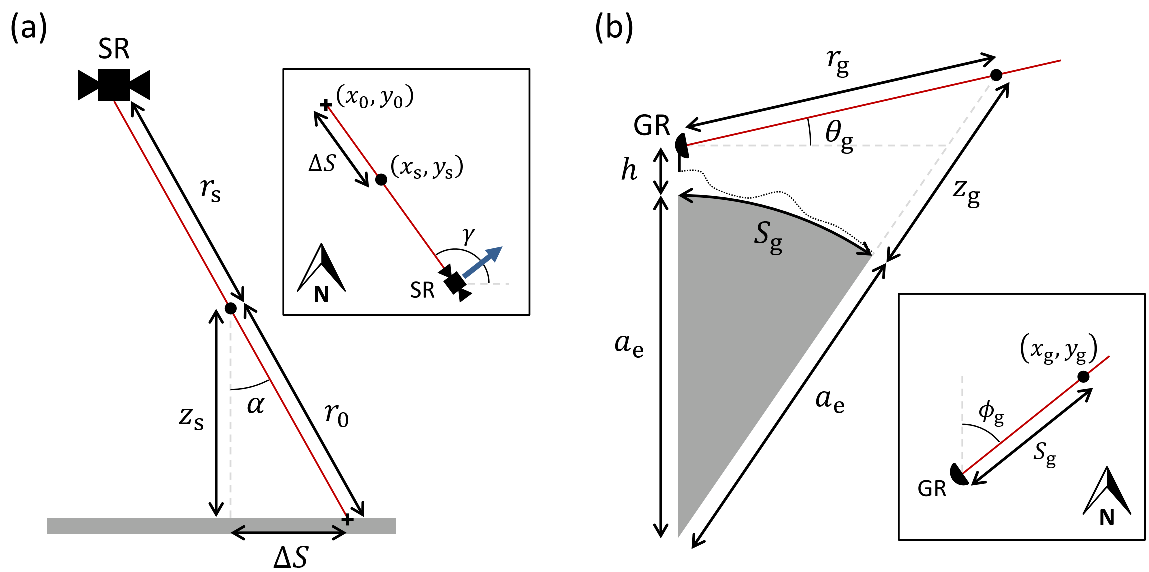


FIG. 12. Smoothed kernel density estimation violin plots showing bin-matched reflectivity differences between (a) SYD and WOL, (b) SYD and NEW, and (c) WOL and NEW radars before (light grey) and after (dark grey) the application of the calibration corrections derived herein. Thin horizontal solid and dotted lines show, respectively, the median and interquartile range of the distribution. Sample sizes, *n*, and Pearson correlation coefficients (before and after calibration corrections), *r*, are given at the top of each panel.



994 Fig. A1. Diagrams showing the geometry of (a) SR and (b) GR measurements (not to scale). In both panels, the main view is from the side in the plane parallel to the radar beam, while the inset shows a plan view. Red lines indicate the radar beam and black circles indicate the location of the bin. Crosses in (a) mark the intersection of the SR beam with the Earth ellipsoid and the blue arrow shows the direction of travel of the satellite. The thin dotted black line in (b) indicates the height of the surface relative to the Earth ellipsoid (i.e. surface orography). See text for details.