### VALIDATION OF SURFACE REFERENCE DATA SETS USING SATELLITE AND MODEL INFORMATION

### Chris Kidd<sup>1</sup>

<sup>1</sup>Earth System Science Interdisciplinary Center, University of Maryland, College Park & NASA/Goddard Space Flight Center, Greenbelt, MD, USA E-mail: Chris.kidd@nasa.gov

### ABSTRACT

The use of surface reference data sets is fundamental to the calibration, verification and validation of satellite-derived precipitation estimates. It is therefore imperative that these data sets are of the highest quality. This paper outlines the use of multi-source information to evaluate and quantify the quality of surface radar data, frequently used for satellite calibrations/validation studies. An initial study focuses on the inter-comparison of radar data over the southern United States versus precipitation retrievals from the precipitation radar (PR) onboard the Tropical Rainfall Measuring Mission (TRMM). This study is then expanded to include the rest of the contiguous United States through comparison of the surface radar, geostationary infrared and model data. Results show that the inter-comparison can be usefully employed to highlight regions of over/under-estimation, while use of skill score statistics can be used to provide a quantitative quality index.

### **1. INTRODUCTION**

Surface reference data sets (SRDs) are an integral part of any precipitation retrieval scheme, whether this is through the initial calibration of the retrieval technique, or through the verification/validation of the results. Such SRDs may encompass measurements from a number of different instrument types, the most common of which are surface radars and gauges, but may also include distrometers and other specific instrumentation. Important for the comparison with satellite precipitation estimates is the ability of the SRDs to represent an areal or volumetric measure of precipitation, consequently, radars are considered to be one of the best sources of data where available, particularly when cross-calibrated with gauge data sets.

Despite extensive work to reduce or mitigate inherent errors within SRDs, errors still exist, particularly at local scales. Within radar data sets the most common errors include range effects, caused by the lifting of the beam above the Earth's surface with increasing range, and anomalous propagation (anaprop) errors that are caused by false radar returns off terrain or structures; these errors lead to under/over-estimation respectively. Although techniques are employed to control such errors they often persist and manifest themselves in radar-derived rainfall accumulation maps. Bias-correction by surface gauge measurements may only partially resolve such errors.

In order for the SRDs to be usefully employed as calibration/validation data there is a pressing need to identify regions of good (and conversely, poor) data within the available SRDs to ensure that i) where used, good quality data is used for calibration, verification and validation and; ii) the overall quality of the SRDs themselves can be improved in regions of poor performance.

### 2. BACKGROUND

The inter-comparison of satellite precipitation estimates is a crucial part of algorithm development and refinement, while their validation is vital to ensure the accuracy of the various products. A number of precipitation inter-comparisons have been carried out (see Kidd et al. 2010). Recently Kidd (2012) looked at the validation of a number of satellite precipitation products together with the precipitation product from an operational model over northwest Europe. In this study it was shown that there are significant regional variations in the performance of the different techniques. However, some of this variation could be attributed to the surface validation data, rather than the precipitation products themselves. Kidd and Hou (2012) studied the performance of these products spatially (25 km) and temporally (3-hourly) over a homogenous region of SE England with high-quality radar and gauge data. They showed that the performance of the products were consistent both spatially and temporally (e.g. the model output was consistently better than the satellite products). However, more importantly, within each of the different techniques the relative performance between different locations was consistent; i.e. the performance of the technique at one location was always better than that at another location - even if the locations were adjacent to each other. This suggests that these errors are consistent spatially, and should therefore be quantifiable.

Satellite (and model) precipitation estimates versus surface estimates (derived from the SRDs) should contain random errors particularly over small regions with similar features. However, the above studies showed that the relative performance of products at particular locations remain generally constant; these are thought to relate to local differences in the characteristics of the surface reference data that lead to differences in statistical performance. These local factors may include radar range (*beam height above ground*), beam blockage (*terrain/buildings*) and anaprop errors (*terrain/buildings, shipping/aircraft*); many of these factors are subtle and may not be immediately obvious when viewing or quality-controlling the SRDs.

Kidd (1997) compared the performance of the polarization-corrected temperate precipitation estimates with the surface radar over the TOGA-COARE region used during the AIP-3 study (see Ebert 1996). This study found that the radar range had significant effects on the relationship between the satellite and surface rainfall (see Figure 1). Furthermore, through the mapping of the discrepancies in the relationship it was possible to spatially analyse and attribute the source of the errors.



Figure 1. Relationship between the polarization-corrected temperature (PCT) and surface radar over the TOGA-COARE region (from Kidd 1997).

One of the techniques used to compare the performance of precipitation data sets is the contingency table (see Figure 2). Through the analysis of the agreement (R:R and NR:NR) and disagreement (R:NR and NR:R) it is possible to assess the detection of precipitation or falsely-reported precipitation. Of importance here are the disagreement elements of the contingency table; if these errors are mapped spatially, the locations of these errors can be identified. By varying the rain/no-rain threshold the sensitivity to different rain intensities can also be evaluated.

		Observed	
		R	NR
Estimated	R	R:R	R:NR
	NR	NR:R	NR:NR

Figure 2. The contingency table; the green-shaded regions indicate the elements in agreement while those in yellow highlight the elements in disagreement.

### 3. Methodology

This study concentrates upon the identification of errors within the currently available radar data sets over the United States. Data from the surface radar networks are compared with satellite estimates derived from the TRMM Precipitation Radar (PR) and the Climate Predication Center's global infrared (Geo-IR) composite. In addition, the precipitation output from the ECMWF operational forecast model is used to confirm and reinforce the use of the Geo-IR data beyond the extent of the TRMM PR data. A summary of the data sets can be found below in Table 1.

Data Set	Temporal	Spatial
US NMQ surface radar	5 minute	0.01°
TRMM Precipitation Radar	Occasional	4.3 km
Global IR composite	30 minute	0.036°
ECMWF operational forecast	3 hour	15 km

Table 1: Surface, satellite and modelled data sets used in this inter-comparison.

# 3.1 Data processing

To facilitate the comparison of the different data sets the surface radar data (the NMQ and Nimrod) were used as the reference, with each of the satellite/model data sets being mapped to that reference; the radar data was then aggregated spatially/temporally. It should be noted that slight variations occur in the resolution match-ups due to the NMQ data having a resolution of  $0.01 \times 0.01^{\circ}$ , or  $1.01 \times 1.11$  km at 25°N decreasing to 0.858 x 1.111 km at 38°N. Specifics of the matching are outlined below:

*NMQ vs PR data:* Data from the PR was mapped to the NMQ latitude/longitude grid at a resolution of 0.01 x 0.01 degree for each 5-minute period (the temporal resolution of the NMQ data) when PR data was available. The NMQ data was then spatially averaged over an area  $\pm 2x/\pm 2y$  from the PR footprint centre for same 5 minute period to match PR resolution. Note that the comparison covers only the US to 38°N due to the extent of the PR coverage,

*NMQ vs IR data:* The Geo-IR data was mapped to the NMQ 0.01 x 0.01 degree grid for each 30 minute image; co-temporal NMQ data (for the time of the stated IR observation) was averaged over  $\pm 1x/\pm 1y$  from the Geo-IR footprint centre. Note that the resolution of the Geo-IR is essentially 0.036° x 0.036° rather than 0.03° x 0.03°.

*NMQ vs ECMWF output*: The ECMWF data was mapped to the NMQ 0.01 x 0.01 degree grid for each 3 hour period; the NMQ data was then accumulated for the 3 hour period of the ECMWF data and averaged over  $\pm 7x/\pm7y$  pixels from the centre of the ECMWF footprint.

## 4. Results

## 4.1 US NMQ surface radar vs TRMM PR

The first comparison was between the NMQ surface radar data and the TRMM PR data sets, spatially and temporally matched as above. The TRMM PR provides an excellent reference data set since it provides a relatively consistent data set and is not beset by the merger of many different radar systems. However, it is known that off-nadir PR scans may under-represent the precipitation due to 'lifting' of the lowest resolvable range bin off the surface. In addition, the PR has a lower rain intensity detection limit of about 0.5 mmh<sup>-1</sup>, consequently the comparison takes the latter into consideration; any

NMQ rain rates between 0.0 and 0.5 mmh<sup>-1</sup> are set to zero. The occurrence of the misclassified (R:NR and NR:R) pixels were mapped, and are shown in Figure 3 below, as well as the Heidke Skill Score.



a) PR rain vs NMQ no-rain



b) PR no-rain vs NMQ rain



c) Heidke Skill Score

Figure 3. Comparison of TRMM PR and NMQ data for co-incident (time/space) matchups at 5 km resolution (2009-2011)

Figure 3a shows the relative occurrences of where the TRMM PR identified precipitation (> 0.5 mmh<sup>-1</sup>) and where the NMQ shows no precipitation (R:NR). As expected, the main artefact which is highlighted is the range effect of the radars, particularly obvious on the right, over the Atlantic Ocean. Although the mountainous regions in the west do not appear to show any significant under-occurrence of precipitation identified by the NMQ this is probably related to the drier conditions in this region. The contrary situation is shown in Figure 3b; here the occurrence of NMQ rain/PR no-rain is shown. This map highlights regions of clutter or anaprop errors, particularly noticeable in the western half of the US. More subtly, on the right-hand side of the image, coastline effects can be observed; these are perhaps more difficult to explain since there ought to be no physical rationale for the radar to observe more precipitation over land than over the ocean.

Figure 3c maps the Heidke Skill Score; this score essentially combines the information from all the cells of the contingency table. The radar range artefact is clearly visible, with very low skill scores, while other regions have moderate skill scores relating to a number of factors. However, over much of the south-eastern region of the US the skill scores are generally high, in excess of 0.8, while some parts surpass 0.9. In these regions the agreement between the PR and NMQ data is clearly very good and therefore would be suitable for calibration/validation activities.

### 4.2 Extension to extra-TRMM regions

The TRMM PR data provides an excellent opportunity to compare satellite-based and surface-based radar estimates of precipitation, and holds the promise of being able to cross-calibrate non-contiguous radar networks. However, the orbit of TRMM limits the radar coverage to  $\pm 35^{\circ}$ N/S, therefore limiting its usefulness to the Tropical region. The Global Precipitation Measurement (GPM) mission, which carries the Dual-frequency Precipitation Radar (DPR), will provide greater global coverage, but will not be available until 2014. Consequently, how do we evaluate the regions outside the coverage of the TRMM PR?

If we make a simple premise that the surface radar data sets are inconsistently correct, but the satellite observations consistently incorrect it should be possible to use conventional satellite precipitation retrievals to identify, or at least infer, the errors within the surface data sets. These statements can be backed-up as follows. The physics behind the radar observations of precipitation are well known, as are many of the artefacts. Although correction schemes can/are implemented, they are not always effective, and when individual radars are incorporated into networks further artefacts can occur. In particular, 'stitching' between individual radars can often be seen, as well as the range effects (both near and far). Therefore, while the overall radar 'map' appears correct, there are a number of defects that are only apparent when, for instance, the radar data is aggregated over time. Satellite precipitation retrievals, on the whole, provide measures of precipitation that are consistent (particularly true of conical scanning radiometers). Unlike radar systems, a single satellite system can provide data over a large area that is spatially consistent; i.e. fundamentally the retrievals will be similar whether they are in the mid-west or east-coast of the US. However, a consistent problem faced by the satellite retrievals is that they require calibration/validation in order to achieve a good degree of accuracy. Thus the satellite retrievals may be consistent, but not necessarily correct.

A simple test of this premise is shown below in Figure 4. Here the CPC Geo-IR composite data has been processed using a threshold of 235K; <235K represents rain, while >235K represents no-rain (the actual threshold of the Geo-IR data for rainfall retrievals is known to vary). The collocated NMQ data uses a zero rain/no-rain threshold. Since the spatial variability in the Geo-IR threshold is likely to be greater than the spatial variations in the radar artefacts, these artefacts should be identifiable.



a) Darker = radar 'overestimates'; lighter = radar 'underestimates'



b) Darker = IR 'overestimates'; lighter = IR 'underestimates'

Figure 4. Comparisons between thresholded Geo-IR data and NMQ surface radar data.

In Figure 4a the lighter areas indicate where the Geo-IR data observes a greater occurrence of precipitation than the surface radar, while the darker area shows where the surface radar indicates a greater occurrence than the Geo-IR data. The effects of radar range are again obvious, reinforcing the NMQ:PR findings, as are the coastal effects around the Jackonsville radar (in the southeast) and the Brownsville radar (centre bottom). In the northwest region the effects of the mountains are clear, and are more striking than was apparent in the NMQ:PR comparison, primary due to only covering the southwest region. In the northeast region, at first glance, the NMQ radar seems to be overestimating the precipitation; however, it is likely to be the Geo-IR that is underestimating the precipitation due to regional variations in the rain/no-rain threshold. In the mid-west small isolated dark spots can be seen; these relate to surface features that cause anaprop errors in the radar data, and include surface relief and wind-farms.

The 'reverse' comparison is shown in Figure 4b; here the light areas show where the Geo-IR data is 'underestimating' the occurrence of precipitation, while the dark regions

are where it is 'overestimating'. The image contains primarily artefacts associated with the Geo-IR data. In particular, banding of the image across the southern part of the image can be related to the quantisation or temperature-correction of the geostationary data. In addition, wave-like patterns can be seen over the Rocky Mountains and are likely to be related to mountain-induced cloud formations.

### 4.3 Model information

Another source of information that can be utilised is that obtained from numerical models. Similar to the use of the satellite-retrieved precipitation data sets, models can be thought of spatially contiguous, although with some regional biases due to the accuracy of the model to represent different meteorological regimes. Data from the ECMWF operational forecast model provided precipitation estimates at 3-hourly, 15 km resolution; the NMQ data was aggregated to the same spatial/temporal resolution. The results of the comparison are shown in Figure 5 below. In Figure 5a the light regions indicate where the model observes rain and the surface radar does not, while the dark areas indicate where the radar observes more rain occurrence than the model. The overall patterns match those of the NMQ:PR and the NMQ:IR comparisons, showing the radar range effects, coastal disparities, wind-farms and mountainous terrain. Figure 5b shows the 'opposite' image; the image is generally homogeneous, although variations occur over the Appalachian Mountains in the east and over the Rockies in the west.

Figure 5c shows the NMQ:model Heidke Skill Score result; over the eastern part of the US the skill scores are generally in the range 0.2 to 0.4, although some regions have lower scores. Further west the skill scores are much more varied with some areas indicating no skill, while over the Sierra Nevada Mountains some high skill scores can be observed. These high scores may however be the result of observed persistence of precipitation over these regions; radar might overestimate the precipitation while the model might over-enhance the orographic effect, resulting in false rainfall in both products.

One key finding that can be seen in Figure 5c is that the Heidke Skill Scores for the NMQ:model is very much lower than the NMQ:PR result; this is despite the averaging of the data over 3-hour, 15 km resolution. This reduction shows that very good similarities exist between similar data (i.e. surface radar and satellite radar), while only moderate agreement is found between the surface radar and other precipitation data sets.



a) Darker = NMQ has higher occurrence; lighter = ECMWF has higher occurrence



b) Darker = ECMWF has higher occurrence; lighter = NMQ has higher occurrence



c) Heidke Skill Score

Figure 5. Comparison of ECMWF vs NMQ precipitation data over the US.

### 5. Conclusions

This study has shown the usefulness of using the cross-validation of precipitation data sets to evaluate surface data. Comparisons using simple thresholded Geo-IR data can help identify small-scale artefacts within the surface radar data sets. The artefacts found in the Geo-IR study are reinforced by those found in the comparison of modelled precipitation data with surface radar data. Critically, the use of the PR data from the TRMM satellite proved most useful, primarily due to the similarity in the observational principles, but also due to the instantaneous nature of the observations and the good spatial resolution. This study shows that there is great hope in applying the technique to the new DPR on the GPM satellite; through greater global coverage improved intercalibration of surface radar networks will be possible. In addition, the use of multi-source 'confidence' maps will allow the identification of regions of high-quality SRDs that will help in the calibration, verification and validation or satellite techniques, while allowing the originators of the SRDs to better target deficiencies in their products.

### 6. References

Ebert, E.E. 1996: Overview of the AIP-3 Project. Proceedings of the Eighth Conference on Satellite Meteorology and Climatology. 215-219. Jan 28- Feb 2 1996, Atlanta, GA.

Kidd, C., 1997: Satellite identification of radar artefacts in rainfall retrievals, *Proceedings* of the Remote Sensing Society, Reading, 1-5 September 1997. 23, 487-492.

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

Kidd, C., Ferraro, R.R. and Levizzani, V. 2010: The International Precipitation Working Group. *Bulletin of the American Meteorological Society* 8, 1095-1099. doi: 10.1175/2009BAMS2871.1

Kidd, C. and Hou, A. 2012: Combined satellite-surface precipitation products: assessment of regional and temporal characteristics and relationships. European Geophysical Union, Vienna, 22-27 April 2012.