9A.2

The impact of precipitation dataset choices on analyses and forecast verification during the HMT

Edward I. Tollerud[1], Huiling Yuan[1], Christopher J. Anderson[1,2], and John A. McGinley[1]

[1]NOAA Earth System Research Laboratory, 325 Broadway, Boulder, CO 80305
[2]Cooperative Institute for Research in the Atmosphere (CIRA), Fort Collins, CO 80523

1. Introduction

Over the continental United States, a variety of different sources of precipitation observations and estimates are available for use as verification datasets and as input to hydrologic forecast models. These range from radar and satellite estimates, generally available in continuous spatial grids, to gauge networks of varying spatial and temporal resolution. Under difficult conditions (*e.g.,* extreme terrain and heavy rainfall) the performance characteristics of different precipitation observations can vary widely. As a result, analyses, streamflow forecasts, and model verification scores based on these different datasets will not generally be identical. Understanding how the choice of data impacts precipitation analysis and verification thus becomes an important objective.

In the present work, we address this issue by evaluating several observational datasets (primarily independent gauge datasets and radar-derived estimates) that are used to compute basin-average precipitation statistics. An eventual use of these results will be to assess other national- and regional-scale precipitation analyses, and to verify numerical predictions of precipitation by high-resolution ensemble forecast systems. Here, we compare precipitation amount distributions computed from the so-called Stage IV gridded radar/gauge analyses with scores computed from independent hourly and daily gauge sites in a specific mountainous region of the western United States. We also briefly examine the effects of data quality. Finally, we discuss the most useful next steps that can increase our understanding of inherent uncertainties in precipitation analyses and numerical forecast verification.

2. The American River Basin Experiment

For two successive winter seasons (2005-6 and 2006-7), research observations and high-resolution Weather Research and Forecast (WRF)-model runs have been made during heavy precipitation events over the American River Basin (ARB) of northern California. Fig. 1 displays the region of the experiment. This effort is part of a series of planned exercises under the auspices of NOAA's Hydrometeorological Testbed (HMT) whose intent it has been to better understand and forecast high-impact rainfall events. The Earth System Research Laboratory of NOAA and the California-Nevada River Forecast Center have been principal contributors to this project. Both maintain websites that can provide further information about the HMT-ARB project.

Since a primary objective for the project has been researching toward better forecasts of river flooding, a single river basin is the focus of this first project. The choice for the American River Basin in the northern Sierras is particularly pertinent because of the orographic nature of the heavy rainfall events there, which presents a perhaps neglected but very significant remaining forecast problem. Unfortunately, the severe terrain and meteorological extremes that drive these events also have negative impact on data quality. This is particularly true for precipitation measurements and estimates. Gauges in remote areas are difficult to maintain and are subject to inaccuracies due to freezing precipitation and other effects. Furthermore, gauge density is a problem in mountainous and sparsely inhabited areas. Radar estimates, on the other hand, are often dubious due to beam blockage, which results in inhomogeneous sampling. For these reasons, it is important to compare the validity and accuracy of operational datasets employed in the forecasting and warning process. Equally important is assessing the attributes of the different datasets as those attributes determine the datasets' suitability for model verification activities.

computed from the 5 days during these IOPs.

3. Operational Precipitation Datasets

Two operational gauge datasets form the focal point of this study. The first dataset, consisting of high-quality, primarily manual 24-h (1200-1200 UTC) precipitation accumulations, is here designated as the 'RFC' set because its observations are monitored, screened, and disseminated at River Forecast Centers. The other (the Hydrometeorological Automated Data System, or HADS) is assembled from automated hourly gauge networks maintained by several agencies including the National Forest Service, the Bureau of Land Management, and the United States Geological Survey. Because they are

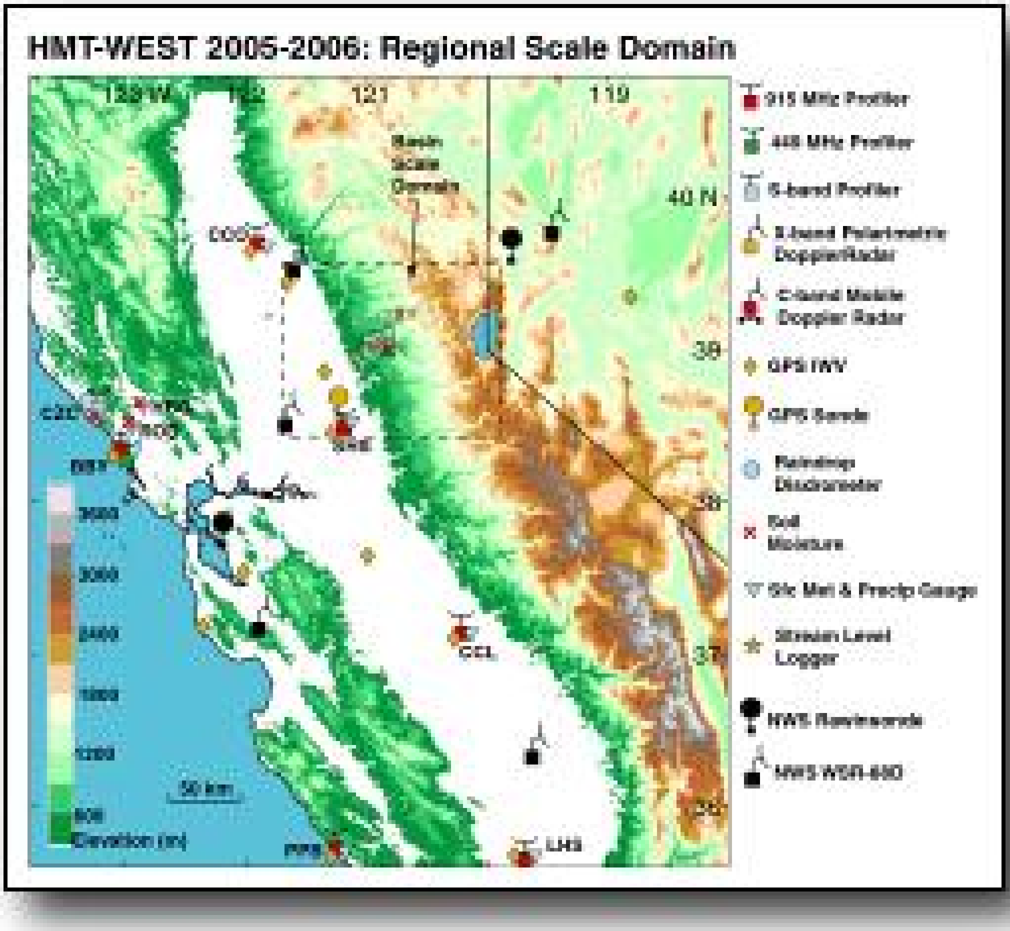


Fig. 1. The HMT American River Basin experimental domain (dashed box) with terrain and elevation features as indicated.

During the HMT-ARB, several major precipitation events were observed. A particularly severe set of storms occurred between 31 December 2005 and 4 January 2006. These storms produced very large precipitation amounts over the ARB and over the west-facing Sierra Nevada foothills, and large stream flow along the American River (Fig. 2). In the context of the HMT-ABR, they make up Intensive Operating Periods (IOPs) 4 and 5. The bulk analyses assembled and displayed here were automated and often located at remote sites, these gauges are more susceptible to large data inaccuracies. We discuss the effect of quality control procedures applied to these data in a later section.

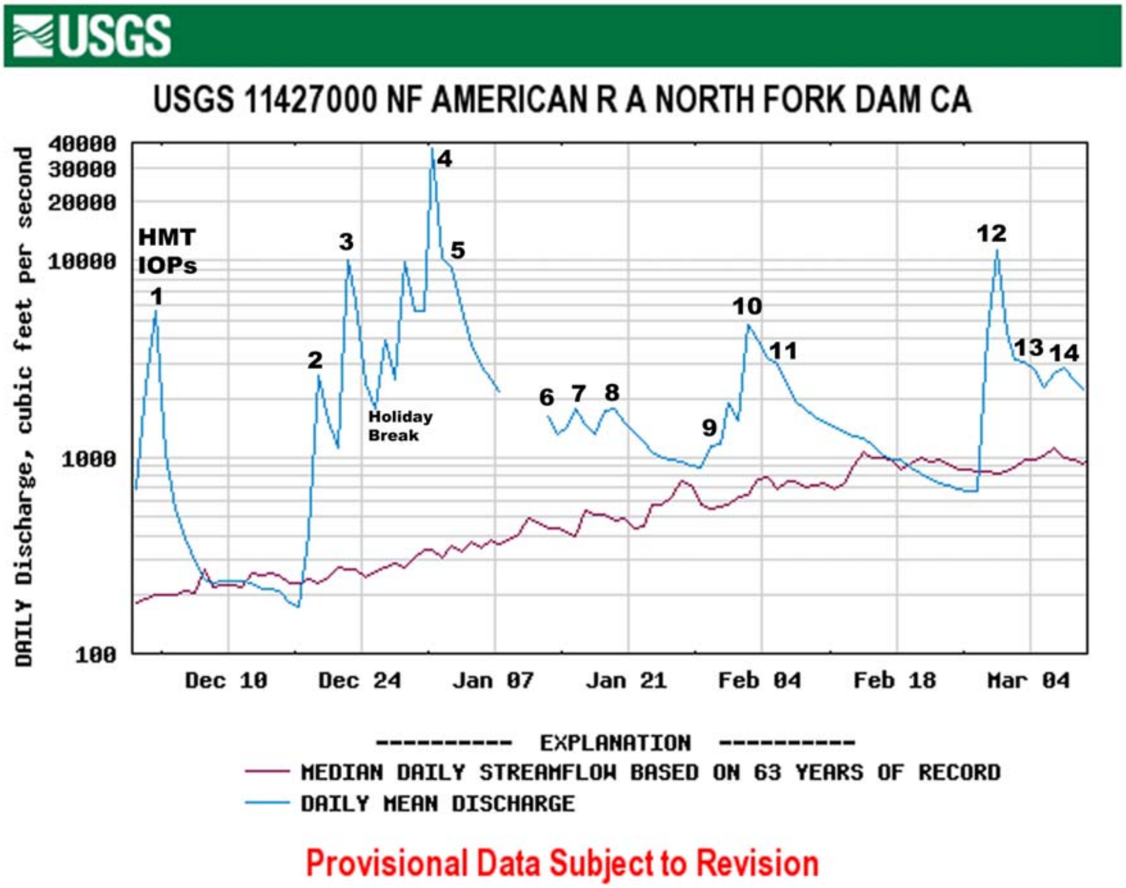


Fig. 2. Stream flow in the American River during intensive operating periods (IOPs) during 2005-6. Display is taken from the USGS operational website http://ca.water.usgs.gov/. Note in particular the peak discharge during HMT IOP 4.

The locations of the gauges in these two data streams relevant to the ARB are displayed in Figs. 3 and 4, respectively. As suggested by the figures, there are approximately equal numbers of gauges in the two sets. Gauge distributions are fairly even, and except for Nevada and other points east of the Sierra Nevada Mountains, the density is fair. We emphasize, however, that the gauge networks still cannot effectively capture the complex terrain of the region. Furthermore, there are other network limitations that affect their sampling characteristics (*e.g.,* a tendency for gauges to be located in river valleys).

A solution to sampling problems is offered by radar precipitation estimates, which are nominally present at very high density and at high-temporal resolution. As the display of radar estimates for the first day of IOP 4 in Fig. 5 demonstrates, radar can provide details of the precipitation fields that are simply impossible for feasible rain gauge networks. However, the tradeoff for spatial continuity includes systematic problems like beam blockage, range height variability, and other radar-specific complications. Comparing radar estimates in Fig. 5 with the hourly gauge observations in Fig. 4, for instance, reveals inaccurate nonzero radar rainfall estimates over much of Nevada.

4. Distributions of Rainfall During IOP 4-5

The very large RFC and HADS daily precipitation totals for 31 December on Figs.

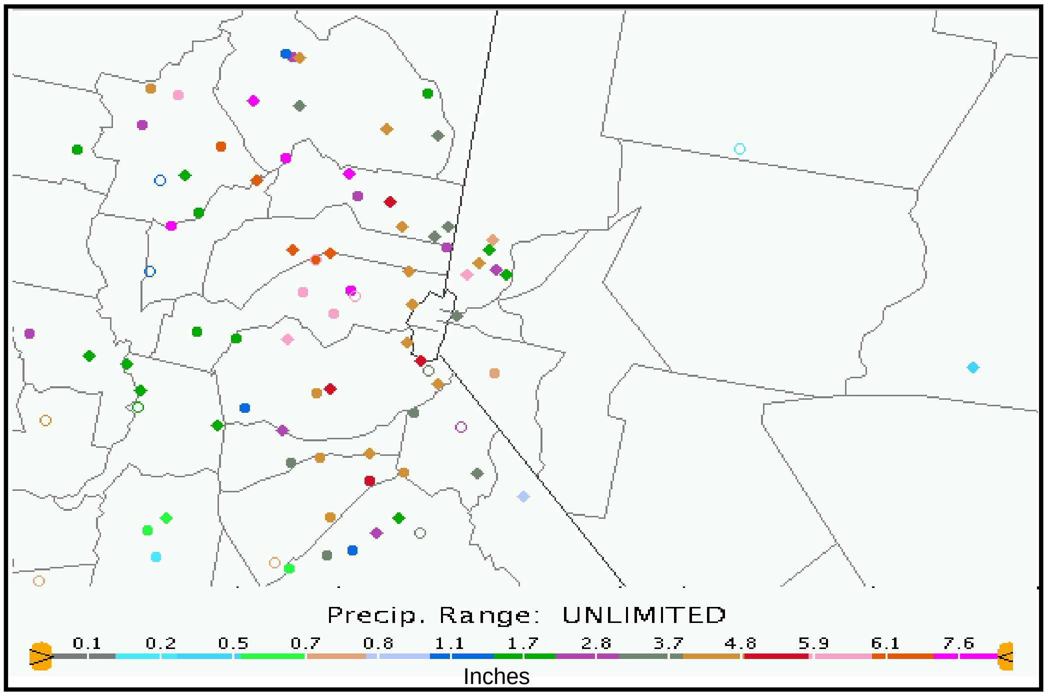


Fig. 3. Daily precipitation totals (1200-1200 UTC, 30-31 December 2005) for the American River Basin in northern California during IOP 4. Observations are accessed from the network of operational daily rain gauges monitored by the California-Nevada River Forecast Center and designated as the RFC in the text. Circular symbols indicate sites that also report hourly observations. An Interactive version of this display with additional gaugesite information and other operational observations is available at http://precip.fsl.noaa.gov/beta/precip5.html

3 and 4 graphically demonstrate the severity of the storm of IOP 4. Rainfall amounts exceeding 5 in were common, and several were well in excess of 7 in. It is interesting to note that when these amounts were initially passed through the automated quality control algorithms at ESRL's Global Systems Division (GSD) many were rejected as being far in excess of reasonable climatic extremes!

Qualitatively, rainfall maxima in the three precipitation fields described by the RFC, HADS, and Stage IV observations appear similar. Most notably, each displays a very strong rainfall maximum along the Sierra Nevada ridge just to the northwest of Lake Tahoe (located at the vertex of the western border of Nevada). Their rainfall fields all also decrease rapidly westward down the slope of the Sierra. If examined quantitatively, however, differences in the three emerge. One common type of representation for precipitation with several useful applications is the frequency distribution of rainfall amounts, as shown in the histograms of Figs. 6 and 7. One overall impression gained from these figures is that of a tendency for HADS observations to be weighted towards lesser amounts compared to the RFC. When the counts of each network are normalized to account for sample size differences (Fig. 7), this tendency is further emphasized.

More striking still on this figure is a similar but significantly stronger tendency for the Stage IV frequency distribution to also be shifted toward lower precipitation values relative to that of the RFC. There is in fact a locally large maximum at the smallest rainfall category for the Stage IV estimates. The comparison between RFC and Stage IV statistics (the 'goodu' and 'stage4' columns) shown in Fig. 8 demonstrates the strong influence of this tendency: domain averages computed with RFC data are a full 15% larger than those computed with Stage IV estimates. We surmise that algorithms estimating rainfall rates from radar

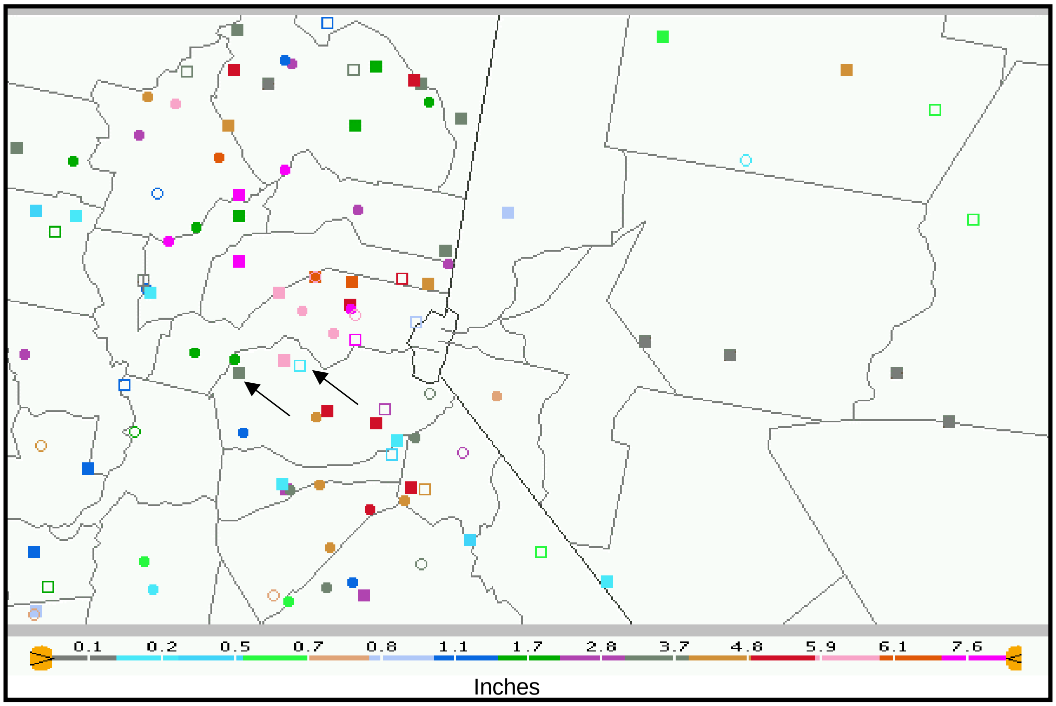


Fig. 4. As in Fig. 3 except for operational hourly gauge sites (HADS and ASOS).

reflectivity do not perform well at these small values, at least in this particular meteorological setting. If so, this offers a possible explanation for the previously mentioned (and apparently erroneous) small but nonzero Stage IV estimates in Nevada.

5. Data Quality Effects

As discussed previously, differences in observation quality from one precipitation dataset to another should also be assumed to produce differences in their use. The table of counts in rainfall categories for stations screened as questionable by QC procedures shown in Fig. 8 helps to quantify the effect of data quality. Clearly, there is a tendency for HADS gauges to erroneously report zero precipitation at a much greater rate than RFC stations (compare the ‘evalh’ column to ‘evalu’ in Fig. 8). Examples of this kind of error are identified by black arrows on Fig. 4. If several of these mistaken observations are missed by the screening process before their use in analyses, the distribution of rainfall rates could be noticeably affected and domain averages might be reduced (as Fig. 8 shows, the domain average produced from the HADS is in fact smaller than that computed using RFC stations).

6. Conclusions and Further Research

From these preliminary analyses of rainfall statistics in the HMT-ABR for a case of very heavy precipitation, we draw several conclusions about the ramifications of the choice of precipitation measurements and estimates. First, independent networks display distinct structural features that can significantly alter the perception of the character of rainfall in the basin. For instance, radar estimates, even when calibrated by available gauge reports, appear to underestimate basin total rainfall volume. Second, gauge site quality issues can affect basin statistics, especially malfunctioning gauges that are incorrectly interpreted to be reading zero or very light precipitation (the HADS data are vulnerable to this problem). Finally, in the face of finite supplies of observation sites (reality, in other words), sampling will always be an important issue for basins set in extreme terrain such as the ABR.

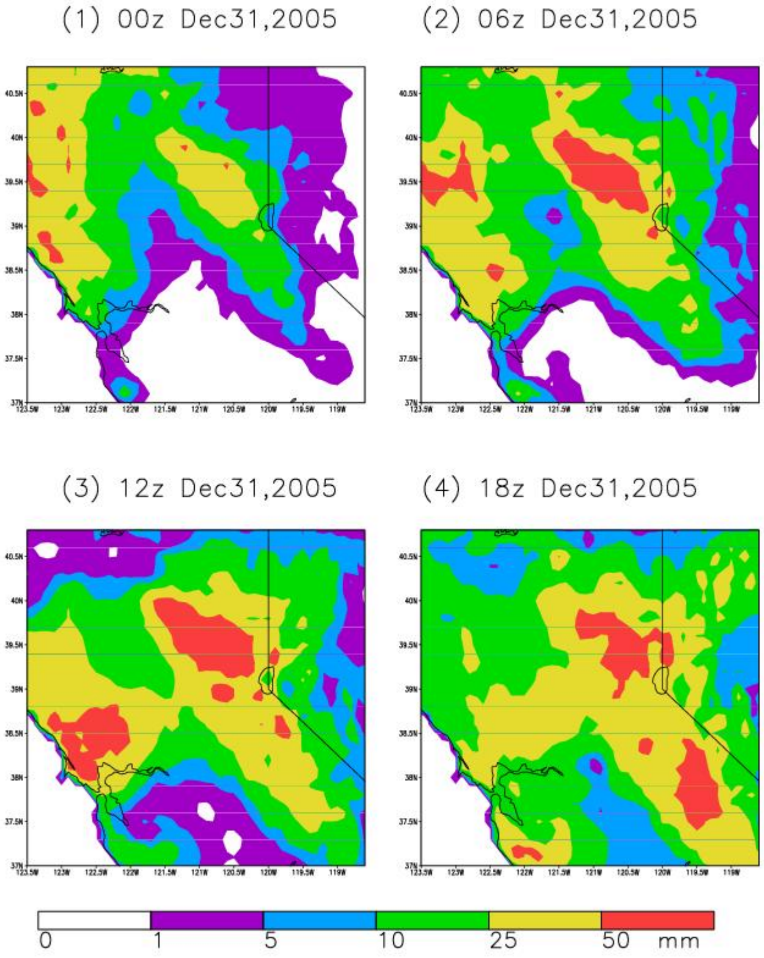


Fig. 5. Gage-corrected radar precipitation estimates for six h accumulation periods during IOP 4 over the American River Basin in northern California.

This apparent sensitivity to network characteristics suggests that gauge placement within a basin could also have a significant impact on network performance. Might it be possible, for instance, that a network of gauges intentionally placed at “hot spots” within a river basin, rather than at locations chosen for convenience or accessibility, might better represent the total rainfall that the basin receives? Determining the locations of these “hot spots” is obviously key here. Radar estimates could serve as initial proxy measures for this purpose, but it is possible that high-resolution forecast fields would be even better.

Research reported in the previous sections has focused on the potential impact of dataset differences on applications like basin or model domain averaging and frequency distribution computation. An equally important issue yet to be addressed involves the quantitative impact of data choice on QPF verification. Would it be most valuable for verification studies to merge the data sets to attain maximum observation density? Or alternatively, might we be better advised to consider the different sets of precipitation measurements as members of an observation ‘ensemble’ that can be individually applied to produce a probability distribution of possible verification scores? This latter possibility is appealing because it suggests a method to estimate verification uncertainty.

Acknowledgments

We thank Ann Reiser and Ed Szoke, both of GSD, for reviews, Randy Collander for figures, and the NOAA HMT for funding.

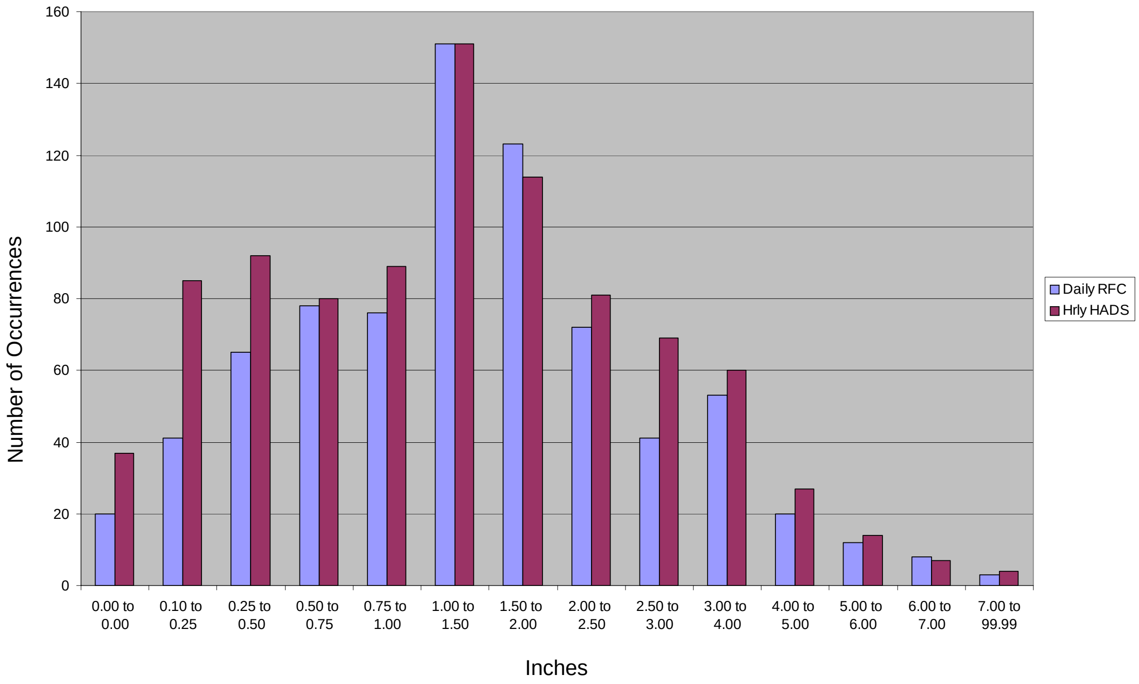


Fig. 6. Histogram of precipitation frequency for selected daily rainfall categories during IOP 4-5 for two independent gauge networks (denoted HADS and RFC; see text for explanation).

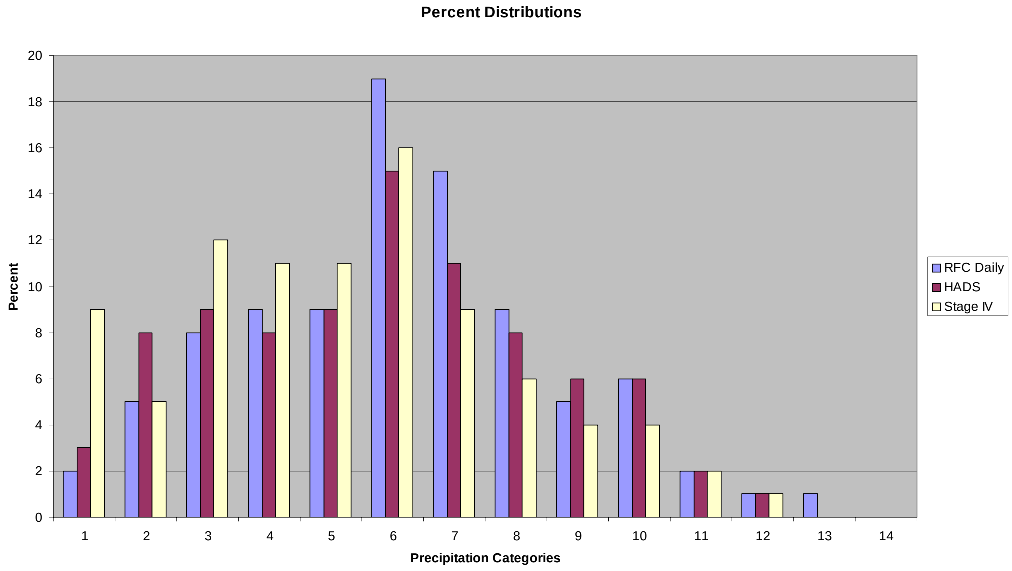
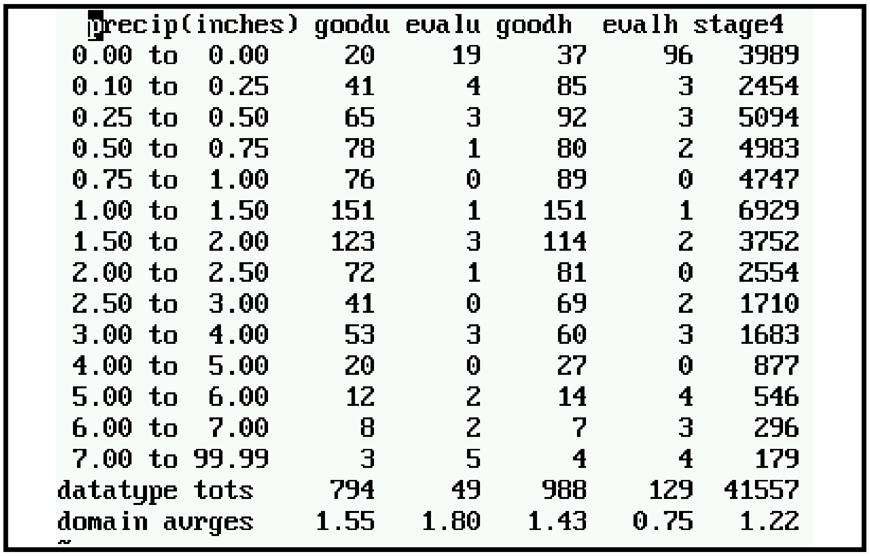


Fig. 7. Histogram of precipitation frequency (percent of total) for selected daily rainfall categories during IOP 4-5 for two independent gauge networks and national radar analysis (denoted HADS, RFC and Stage IV; see text for explanation). Rainfall amount categories are as in Fig. 6.



```
  precip(inches) goodu evalu goodh  evalh stage4
 0.00 to  0.00     20     19    37     96   3989
 0.10 to  0.25     41      4    85      3   2454
 0.25 to  0.50     65      3    92      3   5094
 0.50 to  0.75     78      1    80      2   4983
 0.75 to  1.00     76      0    89      0   4747
 1.00 to  1.50    151      1   151      1   6929
 1.50 to  2.00    123      3   114      2   3752
 2.00 to  2.50     72      1    81      0   2554
 2.50 to  3.00     41      0    69      2   1710
 3.00 to  4.00     53      3    60      3   1683
 4.00 to  5.00     20      0    27      0    877
 5.00 to  6.00     12      2    14      4    546
 6.00 to  7.00      8      2     7      3    296
 7.00 to 99.99      3      5     4      4    179
datatype tots      794     49   988    129  41557
domain avrges     1.55   1.80  1.43   0.75   1.22
~
```

Fig. 8. Comparison of precipitation occurrences for the indicated daily total rainfall categories and the accompanying averaged daily (1200-1200 UTC) observed precipitation over the WRF model domain during IOP4-5 for RFC daily amounts (goodu and evalu), HADS (goodh and evalh), and radar-based analysis (stage4). The number of stations either passed or screened out by QC processing (denoted 'good' and 'eval', respectively) are also displayed.