4.4 TEMPORAL COMPARISON OF THE COMPREHENSIVE PACIFIC RAINFALL DATABASE (PACRAIN) WITH SATELLITE RAINFALL ESTIMATES

Michael D. Klatt*, Mark L. Morrissey, and J. Scott Greene
Environmental Verification and Analysis Center
University of Oklahoma, Norman, Oklahoma, USA

1. INTRODUCTION

The Comprehensive Pacific Rainfall Database (PACRAIN) is a database of rainfall observations from throughout the tropical Pacific region (Morrissey et al., 1995). PACRAIN merges rainfall records from various weather agencies and other sources into a unified format and makes them available via the World Wide Web (http://pacrain.evac.ou.edu). The database currently contains more than 1.8 million records from close to 800 sites. The PACRAIN project is also part of the Pacific Island Global Climate Observing System (PI-GCOS) initiative to expand and enhance observation networks in the tropical Pacific.

Using rainfall data can be problematic because it does not represent an instantaneous quantity. The greatest opportunity for confusion is with daily rainfall as there is no single definition of what "daily" means. For most daily totals this does not mean a calendar day because the observations are not taken at midnight. Even for midnight observations there is the question of local time versus UTC. Daily does not even mean 24 hours unless observations are made at the same time every day.

PACRAIN attempts to remove the ambiguity inherent in daily rainfall data by explicitly defining the time period of every accumulation. The latest version of the database provides a complete UTC timestamp (date and time of day) for every record, and all timestamps designate the beginning of an accumulation period. In some cases time of day is not present in the source data and has to be approximated. Also, some data sources report the end of the accumulation period (observation time) instead of the beginning, so the timestamps for these records must be adjusted to conform to the PACRAIN convention.

Data accuracy is a top priority for PACRAIN. In April 2005 a PACRAIN user reported that rainfall occurrence at several sites appeared to be advanced by one day relative to the 3B-42 satellite rainfall product. Further analysis confirmed this offset, and a more comprehensive PACRAIN/3B-42 temporal comparison was initiated (David Bolvin, personal communication). The first objective of the comparison was to discover all temporal inconsistencies between the two data sets. The next objective was to determine if these were isolated anomalies or systematic errors in either the PACRAIN or 3B-42 data. The ultimate goal of the comparison was to correct any PACRAIN timestamp errors.

The 3B-42 satellite rainfall product is produced by the Tropical Rainfall Measuring Mission (GESDAAC, 2005). The 3B-42 algorithm blends microwave and infrared (IR) sensor data from multiple sources to give estimates of instantaneous rainfall rate (TRMM, 2005). This technique combines the accuracy of microwave rainfall estimates with the continuous coverage of IR estimates (Huffman et al., 1995). Version 6 of the product was used for the comparison. This version has a temporal resolution of 3 hours, a spatial resolution of 0.25º, and global coverage between 50ºS and 50ºN (GESDISC, 2005).

2. COMPARISON PROCEDURE

The basic methodology for the comparison was to determine which arbitrary 3B-42 daily time series gave the best fit with a co-located PACRAIN time series. If there was a time offset between these two time series further investigation was warranted. The comparison was performed for the available 3B-42 period of record: January 1998 through May 2003, with the exception of two missing months in 2000. All 459 PACRAIN sites which had any data during this period were included in the comparison.

The first step of the comparison for a given PACRAIN site was to produce "daily" (24 hour) rainfall totals from the 3B-42 rainfall rates at the grid point closest to the site. These totals were calculated using a moving 24-hour window in 3-hour steps: Day 1/00Z to Day 2/00Z, Day 1/03Z to Day 2/03Z, etc. The average rainfall rate (mm/hour) was calculated for each 24 hour

*Corresponding author address: Michael D. Klatt, EVAC, University of Oklahoma, 3200 Marshall Ave, Suite 150, Norman, OK, USA 73072; e-mail: mdklatt@ou.edu

period and then multiplied by 24 to obtain an accumulation (mm). When calculating an average rate the boundary values at the beginning and end of the period were weighted by 0.5 (David Bolvin, personal communication). If a rainfall rate within a period was missing that period was excluded.

The next step was to construct the PACRAIN time series and its corresponding set of 3B-42 time series. The PACRAIN time series (T_{PAC}) consisted of all daily totals for the site. Multi-day accumulations were excluded, and trace amounts were considered to be 0.01 mm. Each 3B-42 time series (T_{3B42}) was offset from the PACRAIN time series by a given number of hours (Δ):

$$(1)\quad T_{3B42} = \left\{ x(t_{3B42}) \,|\, t_{3B42} = t_{PAC} + \Delta \right\}$$

The full set of 3B-42 time series was generated by varying the offset from –48 hours to +48 hours in 3-hour increments.

Once the time series had been created for the site the product-moment correlation (*r*) was calculated between the PACRAIN time series and each 3B-42 time series. The result was a set of correlation coefficients (*R)*:

$$(2)\quad r(x,y) = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n\sum x^2 - \left(\sum x\right)^2\right]\left[n\sum y^2 - \left(\sum y\right)^2\right]}}$$

$$(3)\quad R = \left\{ r\left(T_{PAC}, T_{3B42}\right) \,|\, T_{3B42}(\Delta), -48 \le \Delta \le 48 \right\}$$

At this point 91 sites had to be excluded from the comparison because there were not enough non-zero rainfall amounts to calculate a correlation coefficient.

The final step of the comparison was to determine the time offset for the site. This was defined as the offset of the 3B-42 time series that had the maximum correlation with the PACRAIN time series. A zero offset meant the PACRAIN and 3B-42 data were coincident for the site, a positive offset meant there was a time lag in the 3B-42 data, and a negative offset meant there was a time lag in the PACRAIN data.

3. COMPARISON RESULTS

Before a complete analysis of the comparison results could be done it was necessary to determine if there was a relationship between the PACRAIN and 3B-42 data in the first place. Figure 1 shows the overall correlation at each time offset, calculated as the average correlation coefficient for that offset for all sites. There is a peak value of 0.408 at an offset of +3 hours, and the correlations decrease smoothly and symmetrically towards larger magnitude offsets. This suggests that any further analysis would be meaningful.

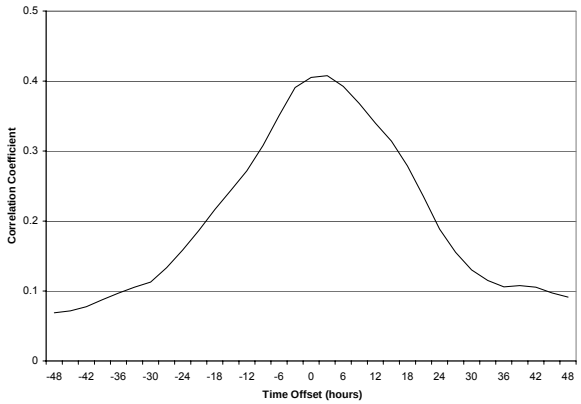

Figure 1: Correlation by time offset for all sites.

If there is no uniform time bias in the PACRAIN data the frequency distribution of site offsets should be symmetric and centered at 0 hours, and Figure 2 shows this to be generally true. The two predominant site offsets are +3 hours (66 sites) and 0 hours (63 sites). The median offset is +3 hours, and half of all sites have an offset between –3 and +9 hours. While these values might suggest a +3 hour bias, it should be noted that the resolution of the comparison data itself is only 3 hours.

Once it was determined that there was not a database-wide time bias, attention was focused on the individual sites with significant time offsets. This led to the discovery of two systematic errors. One error effected all records obtained from the French Polynesia meteorological service, and another error effected some records obtained from the US National Climatic Data Center (NCDC). These records were corrected, and the temporal comparison was repeated to verify the corrections.

When the French Polynesia (FR) sites were examined individually they all had a time offset of approximately +12 hours (Figure 3). The documentation for these data said that all observations were made at 0600 UTC, or 8 pm local time (LT). As most manual observations are made in the morning this seemed unlikely, but there was no evidence to contradict it. Therefore,

0600 UTC was initially used as the time of day for all FR records. Because the temporal comparison confirmed the earlier suspicion, all timestamps for these sites were advanced forward by 12 hours to 1800 UTC (8 am LT) to eliminate the offset.

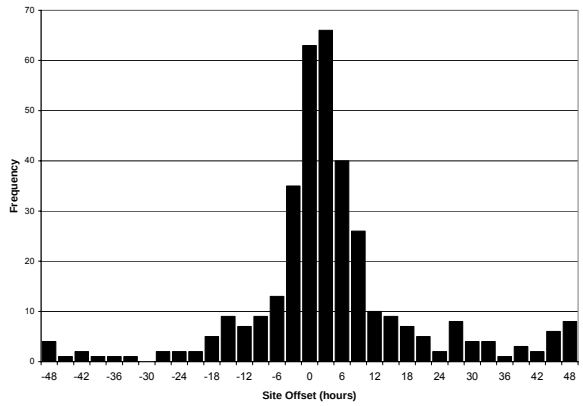

Figure 2: Frequency distribution of site offsets for all sites.

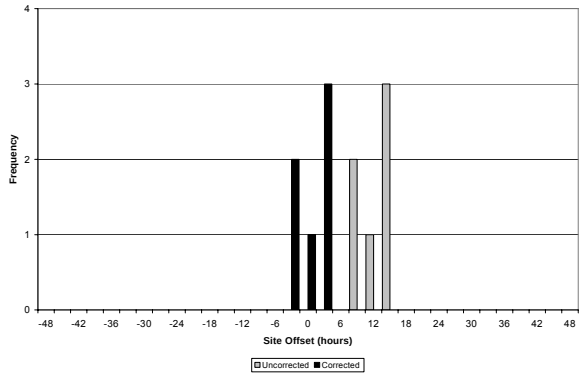

Figure 3: Frequency distribution of site offsets for corrected and uncorrected FR sites.

The individual analysis of US sites showed that a significant number of them had offsets on the order of +24 hours (Figure 4). A review of the original source data for these sites revealed a discrepancy with the data documentation regarding records with midnight observation times (NCDC, 2003). Discussion with NCDC led to the conclusion that when midnight records were ingested into the PACRAIN database they were being attributed to the previous date, causing a 1 day offset (Sam McCown, personal communication). The timestamps for all such records were advanced forward by one day to correct the problem.

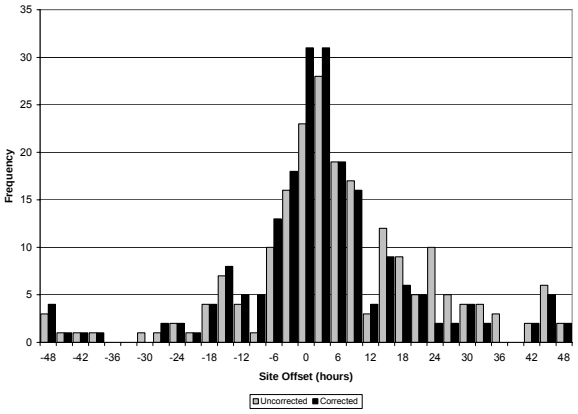

Figure 4: Frequency distribution of site offsets for corrected and uncorrected US sites.

Even after the FR and US corrections there are still many sites with significant offsets (see Figure 2). There are many possible explanations for the remaining offsets. Some of the sites are SPaRCE (Postawko et al., 1994) sites for which the observation time has been estimated. Some of the sites did not correlate well with the 3B-42 data so the resulting site offset is unreliable. Also, human error is always a possibility. In any case, the mere presence of an offset does not warrant any database modifications. Data corrections based on the results of this comparison will not be made without corroboration.

4. SUMMARY

A comparison of the PACRAIN database and the 3B-42 satellite rainfall product was performed in order to assess the temporal accuracy of PACRAIN data. The comparison was done by finding the time offset that maximized the correlation between the PACRAIN time series and an arbitrary 3B-42 time series for a given site. There was no evidence of an overall time bias in the PACRAIN data. However, the comparison did uncover two different systematic errors effecting a number of sites, and those errors were corrected. Work continues to identify the remaining errors and correct them if possible on a site by site basis.

5. REFERENCES

GESDAAC (Goddard Earth Sciences Distributed Active Archive Center), downloaded 2005: TRMM 3B-42 (V6) dataset. [ftp://lake.nascom.nasa.gov/data/TRMM/Gridded/3B42_V6/]

GESDISC (Goddard Earth Sciences Data and Information Services Center), cited 2005: Readme for TRMM Product 3B42 (V6). [http://daac.gsfc.nasa.gov/precipitation/TRMM_README/TRMM_3B42_readme.shtml]

Huffman, G. J., R. F. Adler, B. Rudolph, U. Schneider, and P. Keehn, 1995: Global Precipitation Estimates Based on a Technique for Combining Satellite-Based Estimates, Rain Gauge Analysis, and NWP Model Precipitation Information, *Journal of Climate*, **8**, 1284-1295.

Morrissey, M. L, M. A. Shafer, S. E. Postawko, and B. Gibson, 1995: The Pacific rain gage rainfall database, *Water Resources Research,* **31**, 2111-2113.

NCDC (National Climatic Data Center), 2003: Data Documentation for Data Set 3200: Surface Land Daily Cooperative Summary of the Day. 18 pp.

Postawko, S., M. L. Morrissey, and B. Gibson, 1994: The Schools of the Pacific Rainfall Climate Experiment: Combining research and education, *Bull. Amer. Meteorol. Soc.*, **75**, 1260-1266.

TRMM (Tropical Rainfall Measuring Mission), cited 2005: Algorithm 3B42 - TRMM Merged HQ/Infrared Precipitation. [http://trmm.gsfc.nasa.gov/3b42.html