J8.14 UPGRADES TO AND EXPANSION OF THE COMPREHENSIVE PACIFIC RAINFALL DATABASE (PACRAIN)

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

and

Mark L. Morrissey
South Pacific Regional Environmental Programme, Apia, Samoa

1. INTRODUCTION

The Pacific Ocean plays a vital role in Earth's climate, and precipitation is one of the most important climatological variables. Of particular interest is tropical rainfall (Morrissey et al., 1995). There is an abundance of rainfall data in this region, but obtaining and using these data can be problematic. The area contains a number of independent nations and other political entities, most of which have their own meteorological services. Many of these local agencies have agreements with regional agencies for the distribution of data, but a significant number do not. Thus, finding all relevant data can be difficult. Furthermore, making use of these data is complicated by the existence of multiple formats and reporting standards.

The motivation for a single source of easily-accessible and homogeneous tropical Pacific rainfall data led to the creation of the Comprehensive Pacific Rainfall Database (PACRAIN). PACRAIN is now maintained by the Environmental Verification and Analysis Center (EVAC) at the University of Oklahoma. Over the past two years PACRAIN has undergone a number of upgrades aimed at increasing the value of the database to the research community. In addition to these upgrades the database is being expanded in its coverage, in terms of both the number of sites and the period of record.

2. DATABASE DESCRIPTION

2.1 Overview

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

As of October 2004 PACRAIN contained over 1.7 million daily and monthly rainfall records from 804 sites (see Table 1). There are monthly records for 1874 through 1972, and daily observations for 1941 to the present. Metadata are available for each site, including name, location (see Figure 1), elevation, terrain classification (see Table 2), and period of record. The database is updated with the latest available data on a monthly basis, with additional updates as required.

Table 1. Data distribution by frequency (as of October 2004).

Frequency	Records	Sites
Daily	1678169	686
Monthly	39032	118

Table 2. Data distribution by terrain classification (as of October 2004).

Terrain	Records	Sites
Atoll	595991	145
Coastal	606106	224
Coastal (Orographic)	396849	261
Undetermined	118255	174

2.2 Sources

PACRAIN is made up of data from a combination of government agencies, historical archives, and resources unique to EVAC. The primary government agencies are the National Institute for Water and Atmospheric Research (NIWA) in New Zealand, the National Climatic Data Center (NCDC) in the United States, and Météo-France in French Polynesia. Currently, all monthly observations are from Taylor's (1973) compendium of historical rainfall data. One source of data unique to EVAC is the Schools of

the Pacific Rainfall and Climate Experiment (SPaRCE; Postawko et al., 1994).

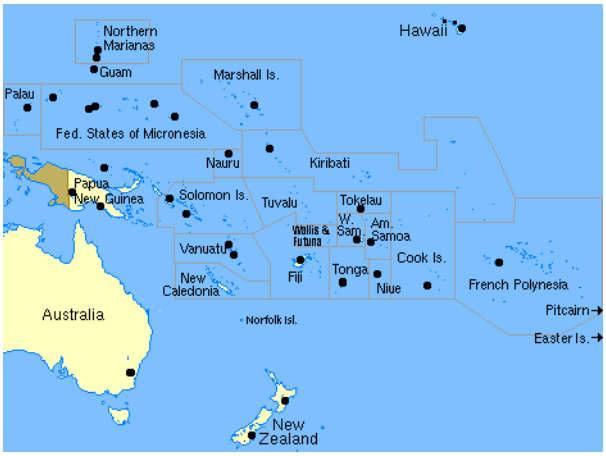

Figure 1. A map showing the general extent of PACRAIN data.

Table 3. Data distribution by source (as of October 2004).

Source	Records	Sites
Météo-France	92986	21
NCDC	775293	316
NIWA	724337	186
SPaRCE	85553	163
Taylor	39032	118

2.3 Access

Data may be obtained from the PACRAIN web site at http://pacrain.evac.ou.edu. Data selection is via an interactive form where the user is allowed to specify various criteria, such as a range of dates and locations. The selected data are presented as a downloadable archive of compressed text files. Users may subscribe to the PACRAIN mailing list to receive announcements about updates and other news concerning the database (http://pacrain.evac.ou.edu/pacusers.html).

3. UPGRADES

3.1 Infrastructure

In 2003 the transition was made from a flat file database to a true database management system (DBMS). The use of a DBMS has improved data security, integrity, retrieval, and concurrency. These improvements come at the expense of raw speed, but this is a worthwhile tradeoff. PostgreSQL is being used as the DBMS engine. PostgreSQL implements Structured Query Language (SQL), a powerful mechanism for data manipulation. Custom interfaces for both C++ and Fortran 95 applications have been developed using SQL and PostgreSQL's native C interface; these custom interfaces allow for rapid development of database applications. As part of the transition to a DBMS the database and associated web site were moved to a semi-dedicated Linux server.

3.2 Database Schema

Since the initial DBMS transition the database schema has been revised. One of the features of the new schema is metadata for each record consisting of a unique identifier, a time stamp indicating when the record was last modified, and optional footnotes. Another improvement over the original PACRAIN scheme is the ability to store multiple data flags for each record. A main consideration in the schema design is flexibility. One aspect of this is the capability to store records of arbitrary temporal resolution, i.e. not just daily or monthly. Part of this capability includes the use of a complete time stamp (date and time) for each record.

3.3 Underlying Data

Three of the most significant PACRAIN improvements concern the handling of dates and times. Previously, the date as given in the source data was used for each rainfall record, and the time was ignored. One disadvantage of this is that different data sources use different reporting conventions. For example, NCDC uses the end of the accumulation period (i.e. observation date) while NIWA uses the beginning of the accumulation period. In August 2003 dates were changed to the beginning of the accumulation period as necessary. Using the beginning of the accumulation period seems more intuitive than using the observation date, especially for lay users. This convention is also superior for storing records of arbitrary frequency. In May 2004 every record was given a complete time stamp, which is essential for daily and higher-frequency data. In August 2004 all time stamps were further converted to UTC where possible, a change that is

especially important for a domain that straddles the Date Line.

The identification and correction of inaccuracies in the data is an ongoing effort. In May 2004 all records from NCDC sites were compared against source data. Errors were corrected, and previously missing records were added to the database. The source data for the majority of the remaining records are not readily available in an electronic format, leaving more than 800,000 records to be verified manually. For these records, errors and omissions are corrected as soon as they are discovered. This review process is being applied to site records as well as rainfall records.

Future plans include the development of a quality control scheme and the capability for site versioning. In this case, quality control is the identification of suspect rainfall amounts, such as unrealistic values and statistical outliers. Sites which seem to be unreliable will also be identified. Suspect records will be marked with some combination of flags and footnotes so that the user may decide how to proceed. Site versioning refers the identification of changes in site metadata over time. Ideally, any significant changes would have prompted a new site identifier in the source data, but this is not always the case.

3.3 Interface

Database upgrades have permitted and in some cases necessitated changes to the online interface and associated applications. The new interface overcomes many of the limitations of the original PACRAIN interface. A serious shortcoming of the original interface is the inability to handle simultaneous queries, but this is not a problem with the new interface. The use of SQL for data access means that an even more powerful interface is possible in the future. For example, there will be more selection criteria, and the user will be able to determine how the data are sorted.

The most noteworthy interface changes concern the actual output. Unlike with the original interface, the output format is now the same for data of any frequency, and is designed to be easy to parse using a variety of programming languages and commercial data analysis software. Data are organized into individual text files by site, and compressed and packaged for download; the user may select either the zip or gzip/tar format. For convenience, data requests are processed offline and the user is notified by e-mail when the data are ready to download.

4. EXPANSION

Most PACRAIN changes to date have focused on usability. Now that these upgrades are in place more effort can be devoted to expanding the scope of the database. In September 2004 all available pre-1971 daily data for non-Hawaiian NCDC sites were added to the database. In time the same thing will be done for Hawaiian NCDC sites and NIWA sites. EVAC is collaborating with the South Pacific Regional Environmental Programme (SPREP) and the Pacific Islands component of the Global Climate Observing System (PI-GCOS) to search for additional sources of rainfall data. Another element of this collaboration is the expansion of contemporary observation networks through technical and financial support.

5. SUMMARY

The Comprehensive Pacific Rainfall Database is a collection of daily and monthly rainfall records from the tropical Pacific region. Data are compiled from a variety of sources into a uniform format and made publicly available via the PACRAIN web site. The database has recently undergone a number of comprehensive upgrades to increase its usability, consisting of improvements to the infrastructure, underlying data, and user interface. Future upgrades are being planned as well. In addition to these changes, the database is being expanded in its spatial and temporal extent in collaboration with the South Pacific Regional Environmental Programme and the Pacific Islands component of the Global Climate Observing System.

6. REFERENCES

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.

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.

Taylor, R. C., 1973: An atlas of Pacific island rainfall, *Dep. of Meteorol. Publ.*, **25**, University of Hawaii, Honolulu.