

Michael D. Klatt \*, Mark L. Morrissey, J. Scott Greene, Susan Postawko  
University of Oklahoma, Norman, Oklahoma

## 1. INTRODUCTION

The Pacific Ocean plays a central role in climate analysis and prediction, and precipitation is one of the most important variables. The Environmental Verification and Analysis Center (EVAC) maintains the Pacific Rainfall (PACRAIN) database, a collection of long-term daily and monthly rainfall observations in the tropical Pacific. PACRAIN data are used as both source data and validation data for gridded precipitation estimates. They are currently being used at EVAC as a component of the Surface Reference Data Center and in other research. They are also frequently used by the general public as a source of climate information. EVAC's goal is to make the PACRAIN database a comprehensive and high-quality source of Pacific island and atoll rainfall for the global climate community and all other interested users. To this end, the database is undergoing many improvements.

## 2. DATABASE DESCRIPTION

The PACRAIN database is made up of four core data sources, and additional datasets are being added as they are discovered. The core data sources are the National Climatic Data Center (United States), the National Institute of Water and Atmospheric Research (New Zealand), the French Polynesian Meteorological Service, and the Schools of the Pacific Rainfall Climate Experiment. New supplemental datasets include the *Atlas of Pacific Island Rainfall* and observations from Japanese possessions prior to World War II. The database currently contains over 1.3 million daily observations from 653 sites, extending from 1971 to the present. There are more than 40 thousand monthly observations from 201 sites, extending from 1874 through 1970. There are 54 sites with both daily and monthly values, giving them very long periods of record. Daily data are organized into files by observation site, and monthly data are organized into files by month. There is also an interactive query form, which allows the user to select data based on criteria like date and location. The data are publicly available via the internet at the PACRAIN web site: <http://pacrain.evac.ou.edu>.

## 3. DATA IMPROVEMENTS

The most challenging aspect of managing the PACRAIN database is the number of disparate datasets that are

used. Besides having different physical formats, each source uses a unique set of reporting conventions. A significant mismatch is the difference in which times are reported. Some sources use the observation time (the end of the accumulation period), while others use the beginning of the accumulation period. All times in the PACRAIN database now refer to the beginning of the accumulation period, except in cases where the source reporting convention is unknown. Even the definition of a day is problematic for rainfall data. Most rainfall observations are not taken at midnight, so a rainfall day spans more than one calendar day. For this reason both the date and time must be part of the record. The time of day is now part of each new PACRAIN record, and is being added retroactively to older records. Once this is complete all timestamps will be converted to UTC, which is essential for comparing data from different locations. Another planned improvement is enhanced quality control procedures. Statistical analysis will be used to determine site and individual records which are suspect. This information will be available as part of the site and observation metadata.

## 4. INFRASTRUCTURE IMPROVEMENTS

The most significant change to the PACRAIN database to date has been moving from a flat file scheme to a true database management system (DBMS). Flat file systems are usually faster, but at the expense of flexibility, data integrity, and security. PACRAIN now uses the open-source PostgreSQL DBMS. PostgreSQL supports both object-oriented and relational database schemes. It uses the nearly universal Structured Query Language (SQL) for data access and manipulation. A variety of database interfaces are available. For PACRAIN, a C++ interface has been built around the native PostgreSQL C interface. This C++ interface is used to create in-house data applications, and will be used for the new web interface as well.

## 5. DATA ACCESS IMPROVEMENTS

The ongoing and planned database improvements will all serve to make data access more flexible and complete. The current public data format available does not allow for things such as time of day and observation metadata, so a new format is being created to include these important parameters. A new interactive query form is being developed to take advantage of this new format and all of the infrastructure improvements. The ability to use SQL to access the data will allow for much more functionality for the user. The new query form will also include different data layout and format options; the goal is to make the data available in a form that will require the least amount of manipulation by the user.

---

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

## 6. REFERENCES

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