Daily Precipitation Grids for South America

BY BRANT LIEBMANN AND DAVE ALLURED

A gridded dataset of historical daily precipitation for South America is now available to the public. We believe this dataset is a substantial improvement over what heretofore has been easily accessible because it contains data from numerous sources. These data have been combined in a simple manner into daily 1° and 2.5° gridded fields for the period 1940–2003.

The data should help to improve our understanding of precipitation variability, a fundamental and difficult problem of meteorology and climatology. Rapid spatial and temporal variability of precipitation, even in the absence of topography, makes diagnosis of the regional- to large-scale component extremely challenging. An accurate depiction of precipitation is a first-order requirement for climate studies and model validation.

Research into the causes of precipitation variability is seriously impeded by a frequent lack of adequate observational data. A few scattered observations, some of which may be missing at any given time, are unlikely to reflect actual precipitation behavior. These and other problems, including timeliness of station reports and a nearly complete lack of coverage over the oceans, have prompted a large research effort into estimating precipitation via satellite retrievals.

Estimates derived from satellite measurements have proved immensely valuable in filling gaps in direct observations, and their accuracy improves as research continues. Nonetheless, gauge-based measurements of good quality and sufficient density provide the most accurate estimate of precipitation over a given area. Further, satellite research has introduced the additional need for gauge-based observations to validate and calibrate the retrievals.

AFFILIATIONS: LIEBMANN AND ALLURED—NOAA–CIRES Climate Diagnostics Center, Boulder, Colorado
CORRESPONDING AUTHOR: Brant Liebmann, NOAA–CIRES Climate Diagnostics Center, R/PSD1, 325 Broadway, Boulder, CO 80305-3328
E-mail: Brant.Liebmann@noaa.gov
DOI:10.1175/BAMS-86-11-1567

DATA. The gridded fields were constructed from about 7900 stations within 10,168 station data files. Most station records (see Fig. 1) are shorter than the full 65-yr period, with some missing observations within the available record. A given grid incorporates all station observations available for that day. The idea is that with a sufficient density of stations, an occasional missing value will not substantially affect the gridpoint average. One must be careful, however, when using gridded fields like these for studies such as trend analyses, which may sensitively depend on the number and temporal homogeneity of stations averaged into a given grid point.

The stations are almost entirely east of the Andes Mountains. There is substantial variation in station density. Brazil dominates the station count in both quantity and density, although station density varies substantially within Brazil as well, with a higher concentration in heavily settled areas to the east.

The yearly count of stations with at least one observation in a given year gradually increases from

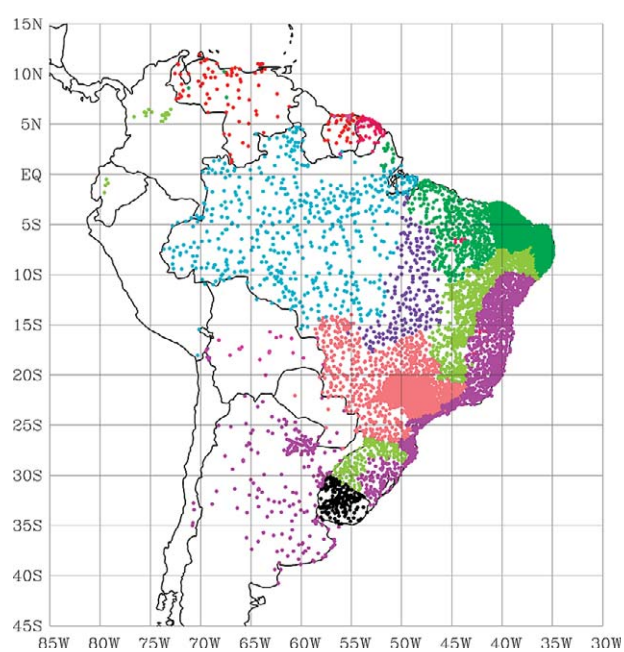
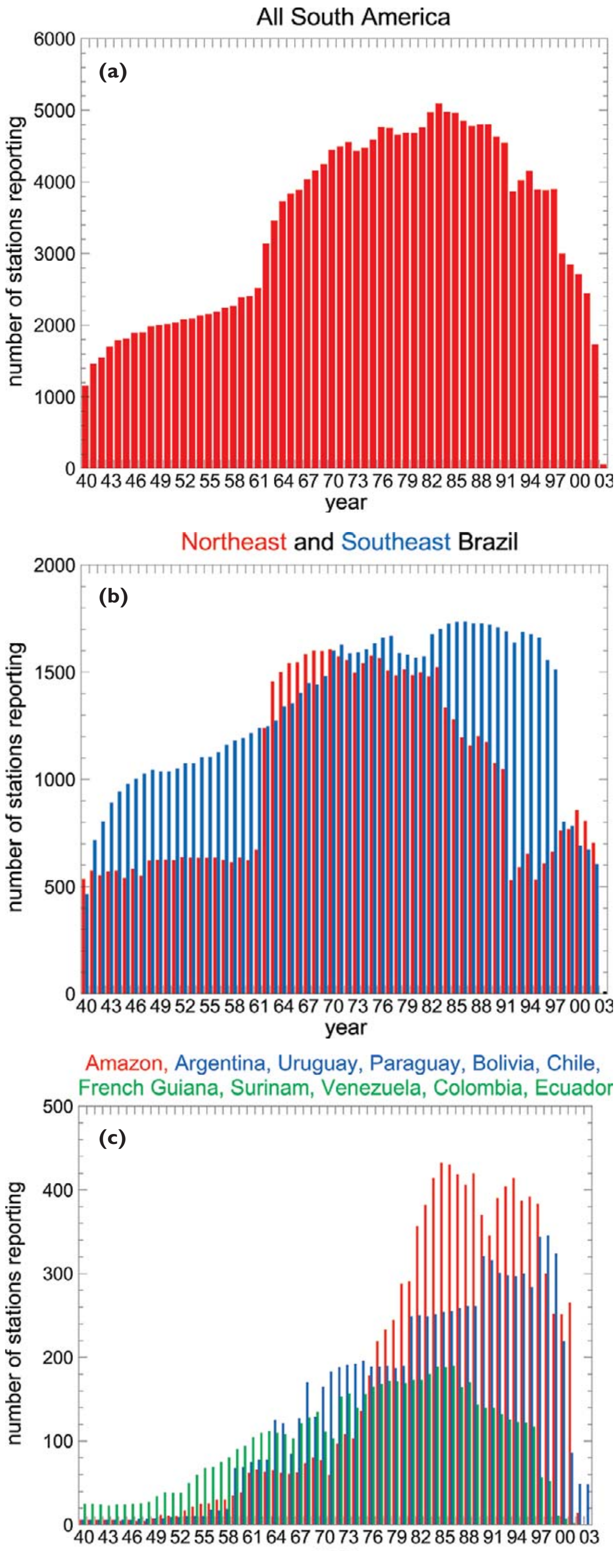


FIG. 1. Stations used to produce gridded precipitation fields described in text. For the most part, in Brazil colors correspond to basins. Elsewhere, colors correspond to different countries or datasets.



1940 until 1961 (Fig. 2). Several hundred stations were added in 1962, and the upward trend continues until 1983, after which there is a nearly continuous decline. Coverage in the last few years is expected to improve as institutional records are updated.

There are more than 500 stations during each year (except 2003) in northeast Brazil (Fig. 2b) (loosely defined—see figure caption for exact boundaries), although from 1962 to 1991 the count is doubled and for some years even tripled. Southeast Brazil shows a slow, nearly steady increase until the late 1990s. Density is poor in the Amazon Basin (Fig. 2c) except between about 1975 and 2001. Except for the last few years, density outside of Brazil has remained reasonably constant for several years. Prior to the 1960s, coverage was generally sparse and tended to follow settlement patterns.

While there is little doubt that coverage in the most recent years will eventually improve, it is not known whether data from stations that existed during the first years of the analysis have never been made available or whether there were few stations in place. It is quite possible that archives, likely in paper form, exist. Presently, we know that data from many stations could be added, but at a high purchase price. One hopes that the practice of charging for data will diminish.

QUALITY CONTROL. Some quality control issues have been addressed in a rudimentary way, some will be addressed before subsequent versions are released as time and resources allow, while some will forever add uncertainty to the precipitation estimates.

The most serious and difficult to resolve problems involve missing values in original station data. In some cases, missing values are recorded as zero. In other cases, blanks are recorded on days with zero precipitation. In either case, missing records and days with zero precipitation are impossible to distinguish. This leads to assumptions and biases in the final interpretation. In some cases, when the problem is obvious (e.g., a year or more with no recorded precipitation), we have removed a subset of that station from the record.

Hidden accumulated values are another problem. When a record includes one or more missing values followed by a large precipitation value, the pattern is

Fig. 2. Count of stations included in grids as a function of year. Station is counted if there was at least one day with data during that year. (a) All South America; (b) northeast Brazil (red; 0°–12°S, 45°W–coast), southeast Brazil (blue; 15°–30°S, 50°W–coast); (c) countries of Argentina, Uruguay, Paraguay, Bolivia, and Chile (blue), French Guiana, Surinam, Venezuela, Colombia, and Ecuador (green), and the Amazon Basin (red).

suspicious unless a confirming notation is included. Because most rain gauges accumulate the precipitation, one suspects that the recorded value for that day may actually include rain from the previous day or days reported as missing. In these cases, we discard any suspected accumulated value at or above an arbitrary limit of 20 mm. This practice almost certainly eliminates some valid observations. Further doubt arises when two problems might be combined (i.e., a long series of zeros followed by a large value). Additional study of original records would be of considerable value in mitigating the problems of ambiguously recorded zeros, missing observations, and accumulated observations.

Occasional, improbably large values are also a problem. Thresholds of 200–450 mm, based on geographic and historical considerations, are applied to constituent datasets. Observations above the threshold are discarded unless confirmed to be valid. Additional work to construct a database of confirmations of extreme events would be valuable to mitigate this problem.

The number of stations with erroneous coordinates is unknown, but is thought to be small. In some cases, we used redundant location information to check for inconsistencies. A few station coordinates were corrected from supplemental catalog or map references. A few stations were discarded due to unresolved inconsistencies in coordinates.

Most stations are measured at 1200 UTC and record precipitation as having occurred on the day on which the rain gauge reading is taken. Some agencies, however, have different observation times, and some times are not currently known. Further, a fraction of the stations record the 1200 UTC observations as precipitation on the day before. The logic of this method is that the majority of the 24 hours measured occurred on the day before the measurement was recorded.

With multiple data sources, the problem of duplicate data arises. This is easily dealt with when stations with identical coordinates have the same identifying names or numbers and identical records. But frequently there are two or more records at identical coordinates, with differing observed values on some dates. In these cases we average the differing values and merge the records into a single time series. This method has the added advantage of eliminating excessive weighting at a single location.

There are numerous other potential kinds of error in the station data, but they have a minor impact on the gridded fields, because of averaging of many nearby stations. For example, one automatic station, upon inspection, showed evidence of wine having been poured into the collection chamber. In spite of the caveats, however, we believe that the data will provide a useful tool with which to investigate South American precipitation variability.

METHODOLOGY. Once quality control is "complete," gridded fields are made by simply averaging all available stations within a specified radius of each grid point. More precisely, the sampling function is the mean of all stations within a circle of specified radius in degree space, with equal station weighting.

The radius was chosen to be 0.75 times the grid spacing, so as to ensure that every station was included in at least one grid point. For example, each point on a 2.5° grid contains data within a radius of 1.875° of the point. This results in a slight overlap, most pronounced in the longitudinal and latitudinal directions, with some stations included in the averages of two to four grid points. This method introduces a slight spatial smoother. Missing value codes are inserted at all grid points where there are no station observations within the sampling radius.

Figure 3 shows precipitation on the 2.5° grid for two consecutive days. The comparison demonstrates the large day-to-day variation of South American precipitation, especially during summer.

QUALITY OF GRIDDED DATA. On any given date, grid points with high station density benefit from spatial smoothing by blending numerous individual observations. In regions of low station density, however, there are many gridpoint values based on a single station report or a very small number of stations. This results in frequent "data noise" in these sparsely populated grid points, which is especially common at the outer edges of the regions with data.

The most extreme types of noise are zeros and intense individual storm events. These can manifest themselves as mathematical singularities and abnormal high spikes far in excess of surrounding gridpoint averages. Both of these singularities will play havoc with analysis software if not accommodated. We recommend that users of this data check for zeros and low station counts and take appropriate precautions in their applications.

Spatial smoothing causes another effect to be considered. Extreme events (heavy storms) that are localized to areas smaller than the grid spacing will be considerably muted by averaging with other stations. Therefore, we

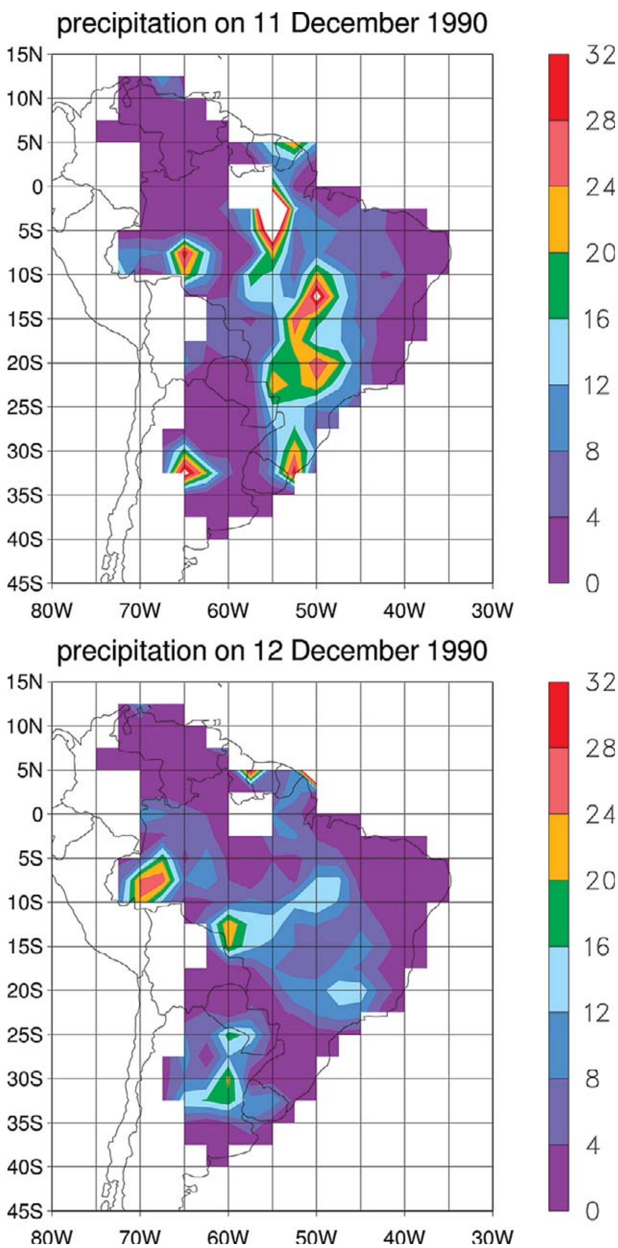


Fig. 3. Rainfall (mm) recorded on 11 and 12 Dec 1990, shown on a 2.5° grid.

recommend that these data not be exclusively relied on for studying small-area extreme events.

Extreme events are also subject to another muting effect in this dataset. The original suppression of suspiciously large values in the original station data is undoubtedly biasing the associated gridpoint averages downward.

Beyond these three considerations, we believe that this dataset is a very good representation of actual historical precipitation.

DISTRIBUTION. The first version of this dataset includes data averaged onto both 2.5° and 1° grids. The fields provided are daily precipitation totals and station counts. The counts give the number of stations that are included in each grid point for each day. These may be used to estimate level of confidence for gridpoint values.

To access this data set, visit the Web site **www.cdc.noaa.gov/people/brant.liebmann/south_america_precip.html** and follow instructions there. The file format is NetCDF, selected for cross-platform compatibility and incorporation of grid coordinates and other useful metadata.

We intend to remake this dataset at least once per year as new data are received and quality control is improved. Potentially, the largest single improvement to the dataset will come from increased station density, reducing the influence of errors. Therefore, we would greatly appreciate any contribution of station data to the database in any format. Electronic formats are preferred, but paper records will also be helpful.

ACKNOWLEDGEMENTS. We wish to thank the Inter-American Institute for Global Change Research (IAI-CRN-055). The following agencies made possible these gridded fields: Agência Nacional de Águas (Brazil); Agência Nacional de Energia Elétrica (Brazil); Superintendência do Desenvolvimento do Nordeste (Brazil); U.T.E. Uruguay, C.T.M. Salto Grande, Servicio Meteorologico Nacional de Argentina, Dirección Nacional de Aeronautica Civil (Paraguay); Administración de Aeropuertos y Servicios Auxiliares de la Navegación Aérea (Bolivia); Servicio Meteorologico Nacional de Uruguay, FUNCEME (Ceará, Brazil); Sistema Meteorologico do Paraná and Instituto Agronomico do Parana (Paraná, Brazil); DAEE (São Paulo, Brazil); Administración Provincial de Agua (Chaco Province, Argentina); Ministerio del Ambiente y los Recursos Naturales (Venezuela); Meteorologische Dienst Suriname; METEO-France; and the National Climatic Data Center (USA).