RC 20033 (04/11/95)
Physical Science

IBM Research Report

Visualization of Scattered Meteorological Data: Study of Severe Rainfall Events in Northwestern Peru

Lloyd A. Treinish

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Visualization of Scattered Meteorological Data: Study of Severe Rainfall Events in Northwestern Peru

Lloyd A. Treinish
IBM Thomas J. Watson Research Center
Yorktown Heights, NY
lloydt@watson.ibm.com

Abstract

The ordinarily arid climate of coastal Peru is disturbed every few years by a phenomenon called El Niño, characterized by a warming in the Pacific Ocean. Severe rainstorms are one of the consequences of El Niño, which cause great damage. An examination of daily data from 66 rainfall stations in the Chiura-Piura region of northwestern Peru from late 1982 through mid-1983 (associated with an El Niño episode) yields information on the mesoscale structure of these storms. These observational data are typical of a class that are scattered at irregular locations in two dimensions. The use of continuous realization techniques for qualitative visualization (e.g., surface deformation or contouring) requires an intermediate step to define a topological relationship between the locations of data to form a mesh structure. Several common methods are considered, and the results of their application to the study of the rainfall events are analyzed.

Keywords: scattered data interpolation, gridding, meteorology, El Niño, visualization

Visualization of Scattered Meteorological Data: Study of Severe Rainfall Events in Northwestern Peru

Lloyd A. Treinish
IBM Thomas J. Watson Research Center

Introduction

The climate of coastal Peru and southwestern Ecuador is mainly controlled by the Humboldt current, a cold ocean current which travels northward along the coastline of Chile and Peru before dispersing near the equator. The current helps cause dry conditions to persist continuously along the Peruvian littoral, making the land strip between the Andes and the Pacific Ocean one of the most arid deserts in the world. Every three to seven years this condition is disturbed by a phenomenon called El Niño, characterized by an ocean warming which appears off the coastlines of northwestern Peru and southwestern Ecuador. This warming modifies the Humboldt current destroying the persistent high pressure zone normally induced by the Humboldt on the west side of the Andes, which in turn generates major changes in the local meteorology and climate [1]. Excessive and severe rainstorms are the most disastrous consequences of El Niño, and such storms can cause great damage to human life, property, crops, and animal life. The rainfall from such episodes causes the flooding of existing rivers, huaycos (mud slides), and the sudden creation of new rivers and lakes.

The heating of the ocean off the Peruvian coast during El Niño periods is part of a larger scale warming of the eastern equatorial Pacific Ocean by several ˚C that creates large anomalies in oceanic and atmospheric circulation. These have, for example, led to the loss of much marine life. The El Niño of 1972 virtually destroyed the Peruvian anchovy fishing industry, which at that time represented a significant percentage of the world's protein supply with a catch of about 12 million tons per year [2].

The 1982-1983 El Niño has received wide attention for its severity [3]. In Peru alone, it was responsible for much loss of life, damage affecting over 80% of the highway system, railroad washouts, and material loss estimated in the billions of dollars. Such destruction emphasizes the need to better understand the meteorological forces unleashed by this powerful ocean-air interaction.

Goldberg et al [4] have investigated the mesoscale structure of severe rainfall events during the 1982-1983 period by examining daily data from 66 rainfall stations in the Chiura-Piura region of northwestern Peru. Figure 1 shows the location of this region, which was selected because it was most severely affected by the 1982-1983 El Niño and because the data were highly reliable and complete.

Figure 1. Location of Peruvian Rainfall Stations

These data support the study of rainfall characteristics over this localized region during El Niño and non-El Niño periods, as a function of elevation, geographic location, and time of year. The stations are enumerated in the Appendix.

Treatment of Scattered Data

Proper care in the treatment of these rain gauge data is required for effective analysis. Such care must be extended to the process of visualization, which creates pictorial representations of data that are a critical aid to interpreting these measurements. The data from the rainfall stations are typical of observational data that are scattered at irregular locations in two or three dimensions (i.e., data with no notion of connectivity or topology). Figure 2 is representative of a straightforward discrete realization of such data as a scatter plot to show the spatial distribution. Figure 3 illustrates the temporal distribution for a single station.

Although such techniques preserve the fidelity of the data, they fail to impart qualitative information about the spatial characteristics of the measurements or the phenomena of which they represent discrete samples. Thus, the application of continuous realization techniques (e.g., surface deformation or contouring for two-dimensional data, volume rendering or surface extraction for three-dimensional data) is necessary. An intermediate step of defining a topological relationship between the locations of data to form a mesh structure is required. Conventional continuous realization techniques can then be applied to such a mesh.

There is a long history of mathematical methods to create meshes from scattered data points. Each method does change the data and their artifacts must be understood because they will carry through to the actual visualization. This discussion is only meant as a very brief introduction to the topic. Nielson [5] summarizes many of the methods in use today and their relative advantages and disadvantages.

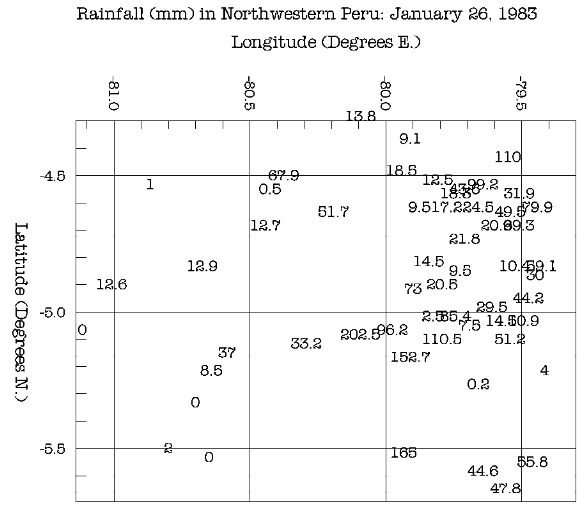

Figure 2. Spatial Distribution of Peruvian Rainfall Measurements on January 26, 1983.

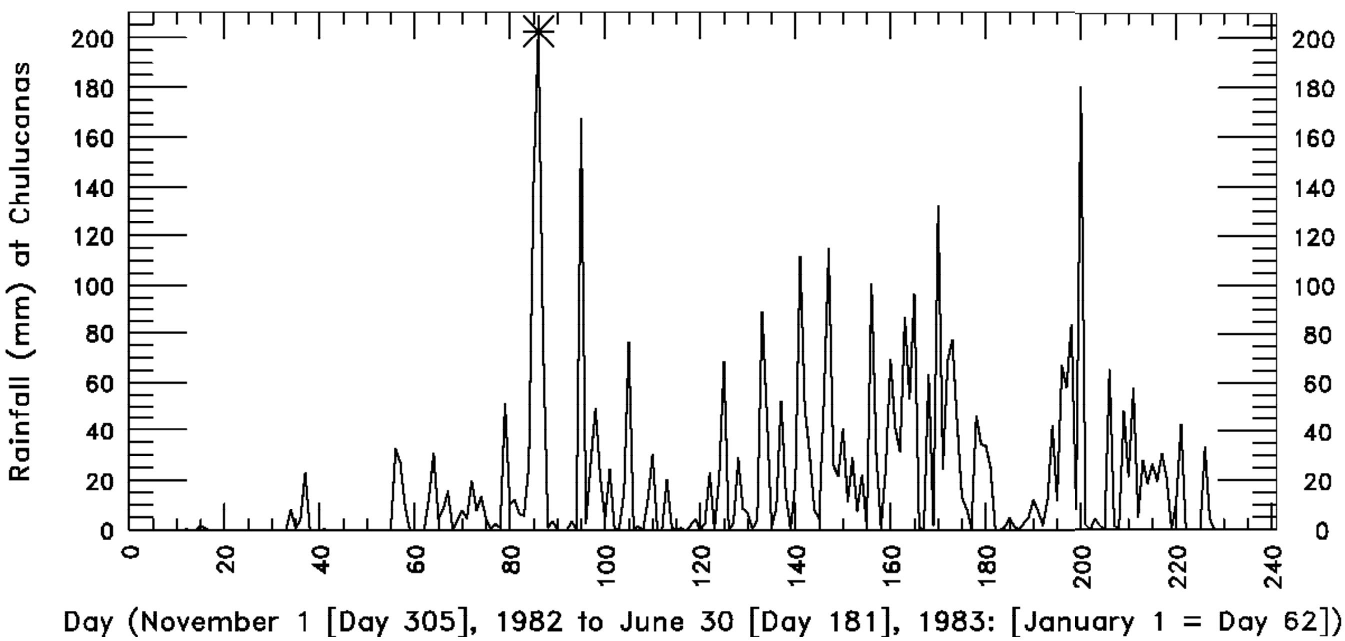

Figure 3. Daily Distribution of Rainfall Measurements at Chulucanas, Peru with January 26, 1983 Highlighted

The simplest and quickest approach is to create a regular grid from the point data by nearest neighbor meshing -- find the nearest point to each cell in the resultant grid and assign that cell the point's value as illustrated in figure 4. Such a technique is valuable because it preserves the original data values and distribution of a grid after a coordinate transformation may have taken place on a collection of points. Although computationally inexpensive, the results may not be very suitable for qualitative display because of the preservation of the discrete spatial structure.

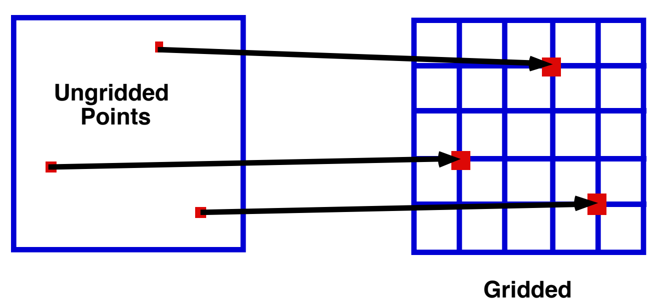

Figure 4. Nearest Neighbor Gridding

An alternate approach that preserves the original data values involves imposing an unstructured grid dependent on the distribution of the scattered points. In two dimensions, this would be a method for triangulating a set of scattered points in a plane [6]. This technique first requires the Voronoi tesselation of the plane with a polygonal tile surrounding each of the scattered points. These tiles are such that the locus of all points within a particular tile are closer to the scattered point associated with that tile than they are to any other points in the set. A triangulation can then be constructed which is the dual of the Voronoi tesselation (i.e., connecting a line between every pair of points whose tiles share edges). This is known as Delauney triangulation and is illustrated in figure 5 as applied to the rainfall stations.

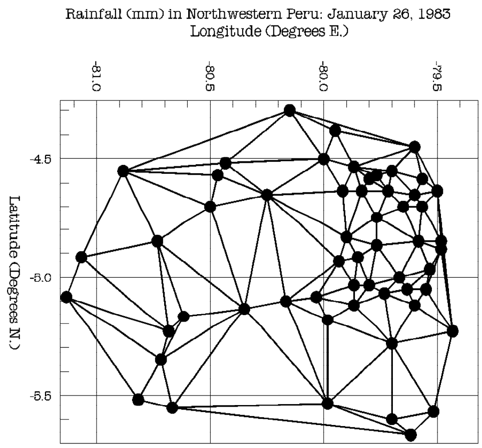

Figure 5. Delauney Triangulation of Rainfall Stations.

For a relatively random distribution of a small number of points such as these rainfall data, the application of continuous realization techniques to the triangulated mesh does not yield useful qualitative results. Consider figure 6, in which the mesh from figure 5 is pseudo-colored by amount of rainfall. Over the pseudo-colored grid are two magenta lines, which are the coastline of the Pacific Ocean and the Peru-Ecuador border, respectively, registered with the rainfall and altitude data. The rendering process applies bilinear interpolation to the value at each node to determine the color of each pixel in the image. Although the original data are preserved, the sparseness of the points results in a pseudo-colored distribution that is difficult to interpret. Artifacts of interpolation in color space from rendering dominate the appearance of the image instead of the spatial distribution of the data.

A potentially more appropriate method, and certainly one that is more accurate than nearest neighbor meshing, uses weighted averaging as illustrated in figure 7. For any given cell in a grid, the weighted average of the n nearest values in the original data distribution spatially nearest to that cell is chosen. A weighting factor, $w_i = f(d_i)$, where d_i is the distance between the cell and the ith (i = 1 , …, m) point in the original distribution, is applied to each of the n values. Figure 7 illustrates the case where n = 3. A common weight is $w = d^{-2}$. These are variants of Shepard's method [7]. For example, Renka [8] modified this approach with local adaptive surface fitting. Collectively, these methods are typically O[nlog(n)] in cost. Intermediate in quality and computational expense would be using linear instead of weighted averaging.

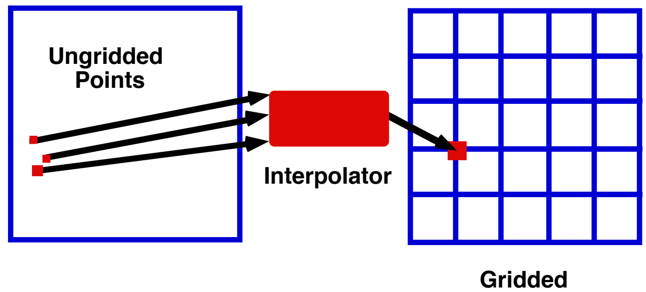

Figure 7. Weighted Average Gridding

All methods in the aforementioned class do introduce aliasing or smoothing of the data to achieve a gridded structure. The form of the interpolation may also impart artifacts on the results depending on the relative spatial variability of the original data vs. how close the interpolation function may be able to model that structure. Given a goal of qualitative visualization, such artifacts may be acceptable. Figure 8 shows a regular mesh with spacing of 0.04 degree (of latitude and longitude), onto which the rainfall data for January 26, 1983 has been gridded using d^{-2} weighting for n = 5. A quadratic was chosen to have local extrema at data sites. The number of neighbors was chosen empirically by experimentation with the application described in the Implementation section below. Five nearest neighbors provides a reasonably smooth distribution that is not significantly enhanced by the use of additional points.

The mesh and the data locations have been similarly pseudo-colored. There is good, but NOT perfect correspondence between the original data values and that of the interpolated grid, which is sufficient for qualitative realization. From this grid, isocontour lines of constant rainfall every 25 mm and a pseudo-colored image are created as shown in figures 9 and 10, respectively. Unlike in figure 6, the pseudo-color image in figure 10 is dominated by the distribution of the data, rather than rendering artifacts.

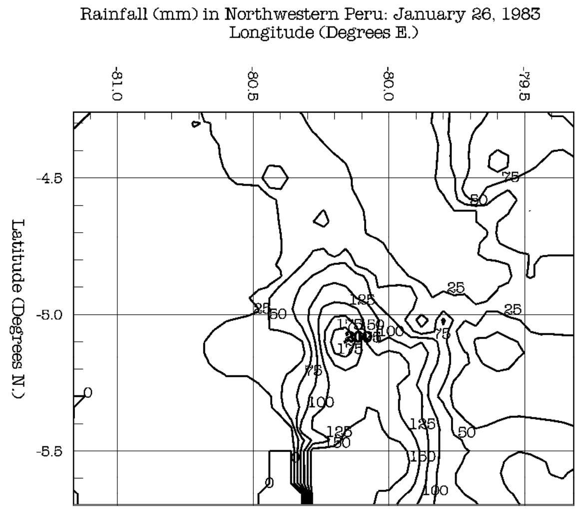

Figure 9. Isocontour Lines of Rainfall from Weighted Average Gridding of Stations.

Results

The detailed results of the analysis of the rainfall data are presented by Goldberg et al [4]. The data have shown that enhanced rainfall during El Niño can be divided into two categories; first, a general increase in background levels over non-El Niño periods and second, sporadic bursts of intense rainfall superimposed over the enhanced background levels (cf., figure 3). The two categories of enhanced rainfall in comparison to the normally dry regions led to the choice of the hue-based pseudo-color map (e.g., figure 10) in which variations in each are illustrated. Hence, this increased background level is shown in yellow and green instead of brown (dry regions) while the bursts are shown with increasingly darker shades of blue. It is these sporadic bursts which were most responsible for the great damage during the 1982-1983 El Niño episode.

Effective gridding of the observations is critical for analysis. The weighted average method described earlier is used to create a pseudo-colored mesh independently for each day over the eight months of daily data being examined (November 1, 1982 through June 30, 1983). As a further aid to the study of spatial and temporal variations, the mesh is deformed by the altitude at each node, which is determined from the same gridding process applied to the altitude of each station as shown in the Appendix. The result is a simple elevation model, which gives a reasonable approximation of the topography in northwestern Peru, especially given the paucity of high-resolution elevation data for this region. This surface with pseudo-colored rainfall is used in figures 12 and 13.

The data further show that the severe storms (or bursts) often originate near the Andean foothills and may be induced by the interaction of rainbands moving inland from the coast with mountain downslope winds. This is best shown with a time sequence during such an event with local topography pseudo-colored by rainfall amount. Figures 12 and 13 illustrate the distribution of rainfall for January 24-27, 1983, and May 19-22, 1983, respectively, when a severe series of storms took place. Deep blue regions corresponding to these storms can be seen starting near the coast and diverting to the northwest after reaching the foothills of the Andes.

Applications to Other Data

To illustrate the generality of the distance weighted average approach to gridding of scattered data, consider its application to other data sets. Figure 13 shows a visualization of a collection of yearly averages of weather data for 1960 from 1702 stations scattered around the earth. Temperature and precipitation data are independently gridded to an irregular mesh of quads, which is a digitized representation of the earth's land masses at one-degree resolution in latitude and longitude. The precipitation data are shown as a pseudo-color map in a Mollweide projection while the temperature data are shown as a pseudo-color contour overlay every five ˚C. The gridded representation illustrates correlation between lack of precipitation with very high temperatures and high precipitation with moderately-high temperatures, for example. These relationships would be difficult to see with scattered realization methods.

Figure 14 illustrates observations made by a nadir-looking instrument on board a polar-orbiting spacecraft, which examined backscattered ultraviolet radiation in the earth's upper atmosphere. These measurements are used to determine the vertical distribution of ozone at discrete geographic locations along its flight path. The set of such ozone profiles measured on October 1, 1987 are shown on the left as pseudo-colored tubes, whose color corresponds to the ozone density and the diameter is the size of the instrument's viewing footprint at the earth's surface. The radial extent of these tubes on the scale of the earth's diameter is exaggerated to reveal the profile structure. The locations of the observations are warped onto samplings on spherical shells in an earth-centered, three-dimensional coordinate system and registered with a gray-scale globe, which is derived from topography data. At the right is a direct volume rendering of the same data in the same coordinate system. The scattered data are gridded at each level to a mesh defined by the output of a complementary instrument on the same spacecraft. The data are only interpolated within the valid geographic coverage of the instrument, then stacked vertically, spherically warped, pseudo-colored and opacity mapped. The spatial structure of ozone can be seen including a heavy concentration in the upper troposphere and lower stratosphere (i.e., the red shell) and a signature of the so-called ozone hole at intermediate and upper levels in the southern hemisphere in cyan. Application of familiar volumetric techniques to reveal this structure would not be possible without introducing the gridding process.

Implementation

The techniques described herein have been developed with IBM Visualization Data Explorer, a general-purpose software package for scientific data visualization and analysis. It employs a client-server architecture with an extended data-flow execution model and is available on Unix workstations (e.g., Sun, Silicon Graphics, Hewlett-Packard, IBM, DEC and Data General) [9]. Data Explorer provides tools for operating on both scattered and gridded data. The Data Explorer *Connect* module performs the Delauney triangulation used in figures 5 and 6 while the *Regrid* module performs the weighted average interpolation used in figures 8, 9, 10, 11, 12, 13 and 14. The *Regrid* module provides independent control of the exponent of the weighting factor, the size of n and the radius of influence from each node of the grid within which to consider data points. In this case a radius of 0.36 degree (of latitude and longitude) was used. It should be noted that for each day of data, not all stations have rainfall measurements. This is NOT the same as a station reporting no rain. Data Explorer supports a notion of data invalidity. Hence, for any given day, only those stations having a measurement are considered by both the *Connect* and *Regrid* modules in creating gridded versions of the data. The choice of modules that support continuous realization is independent of the use of *Connect* or *Regrid* even though they result in different mesh structures because these Data Explorer operations are polymorphic and appear typeless to the user. This polymorphism is a consequence of Data Explorer being built on a foundation of an unified data model, which describes and provides consistent access services for any data that is to be studied independent of shape, rank, type, mesh structure or dependency or aggregation.

To illustrate how these tools may be applied to scattered data, consider figure 15, which contains a Data Explorer visual program. There are nine modules in this program:

Import - reads rainfall data from disk.

Sequencer - provides a frame counter for controlling daily animation.

Select - chooses which day from the time series to process.

Include - flags which stations failed to report a measurement.

Construct - defines the aforementioned 0.04-degree-resolution grid (36 x 46) for interpolation.

Regrid - interpolates the point data to a grid, which is highlighted. The configuration for this module is also shown, where the number of nearest neighbors, radius of influence and weighting factor are defined.

Colormap - provides interactive construction of a custom color map, which is shown.

Color - applies the color map to the gridded rainfall data.

Image - renders an image and provides interactive tools for its manipulation.

This program will generate an animation sequence of the daily rainfall as (gridded) pseudo-color images, one of which is shown. A similar visual program was used to create figures 8, 9, 10 and 13. If Delauney triangulation were preferred over weighted average interpolation, then the use of the *Construct* and *Regrid* modules would be replaced by the *Connect* module (e.g., for figures 5 and 6).

Such a visual program can be expanded into a complete application that supports the study of the rainfall data. A snapshot of such an application in operation is shown in figure 16. The data for one day are shown as a pseudo-colored deformed surface, using the aforementioned technique. A time history of the rainfall for a single station is shown as a plot. Continuous and discrete realization techniques are provided. For the continuous techniques (e.g., contouring, imaging), the data are gridded as discussed earlier. The discrete techniques (e.g., scatter plots) may be overlaid with the continuous ones to evaluate the quality of the interpolation. Optionally, the topographic surface may be shown. The daily distribution of the rainfall at an individual station may be plotted. In addition, the gridding process may be adjusted interactively. The number of stations used to compute a value at each grid point and the radius of influence from each data point may be independently selected. These features are selected through three control panels containing Motif widgets. A VCR-like widget may be used to create time-sequenced animations of the daily data. Of course, the geographic view of the data may be varied interactively.

Conclusions and Future Work

The characteristic topography near regions such as Chulucanas (roughly in the center, cf., figures 3, 11, 12 and 16 and the Appendix), where such storms were observed to occur on a frequent basis, is ideal for the aforementioned interaction between the rainbands and the Andean foothills. The origin of the east-west rainbands near the north Peruvian coast is less clear but may be caused by low altitude wind surges, which are driven northward along the coast of Peru by a large and quasi-permanent high in the south-eastern Pacific.

The potential influence of this phenomena may be determined by looking for changes in ocean conditions due to El Niño prior to the onset of rainbands. Figure 17 illustrates the ocean surface from an empirical model on January 24, 1983 before the major storm that is shown in figure 11. Sea level pressure is shown as a deformed surface, pseudo-colored by surface temperature. Streamlines of surface winds are also shown, pseudo-colored by wind speed. A region of ocean warming in the southeastern Pacific Ocean (pinkish-white) is associated with a decrease in sea level pressure and winds moving toward the South American coast.

Careful methods of gridding scattered (measured) data are critical for effective visualization, especially when used for continuous realization to yield qualitative information. The techniques described herein appear to be suitable for earth sciences applications other than meteorology (e.g., hydrological samplings, petroleum or mining well logs) as well as independent disciplines as diverse as medicine (e.g., measurements distributed on a patient's skin) or aerospace engineering (e.g., pressure along an airfoil or temperature inside a jet engine). Enhancement of the current study or extensions to other domains will require investigation of the applicability of other methods of gridding scattered data (e.g., [5], [10], [11], [12]).

Acknowledgments

The data are available courtesy of NASA/Goddard Space Flight Center, Greenbelt, Maryland.

The author thanks the anonymous reviewers for a number of valuable suggestions to improve the quality of this paper.

References

1. E. M. Rasmusson. "El Niño and Variations in Climate". **American Scientist**, **73**, 168, 1985.

2. W. H. Quinn, D. O. Zopf, K. S. Short and R. T. W. Kuo-Yang. "Historical Trends and Statistics of the Southern Oscillation, El Niño, and Indonesian Droughts". **Fishery Bulletin**, **76**, p. 663, 1978.

3. S. G. H Philander. "Anomalous El Niño of 1982-1983". **Nature**, **305**, p. 16, 1983.

4. R. A. Goldberg, G. Tisnado M. and R. A. Scofield. "Characteristics of Extreme Rainfall Events in Northwestern Peru during the 1982-1983 El Niño Period". **Journal of Geophysical Research**, **92**, C13, pp. 14225-14241, 1987.

5. G. M. Nielson. "Scattered Data Modeling". **IEEE Computer Graphics and Applications**, **13**, n. 1. pp. 60-70, January 1993.

6. Agishtein and Migdal. "Smooth surface reconstruction from scattered data points". **Computer Graphics (UK)**, **15**, n. 1, pp. 29-39, 1991.

7. D. Shepard. "A Two-dimensional interpolation function for irregularly-spaced data". **Proceedings 23rd National ACM Conference**, pp. 517-524, 1968.

8. R. J. Renka. "Multivariate interpolation of large sets of scattered data". **ACM Transactions on Mathematical Software**, **14**, 2, pp. 139-148, June 1988.

9. G. Abram and L. Treinish. "An Extended Data-Flow Architecture for Data Analysis and Visualization". **Computer Graphics**, **29**, n. 3, May 1995.

10. R. H. J. Gmelig-Meyling and P. R. Pfluger. "Smooth Interpolation to scattered data by bivariate piecewise polynomials of odd degree". **Computer Aided Geometric Design**, **7**, pp. 439-458, 1990.

11. W. F. Smith and P. Wessel. "Gridding with continuous curvature splines in tension". **Geophysics**, **3**, pp. 293-305, 1990.

12. L. Yue-sheng and G. Lü-tai. "Bivariate Polynomial Natural Spline Interpolation to Scattered Data". **Journal of Computational Mathematics**, **8**, n.2, pp. 135-146, 1990.

Appendix. Alphabetical Listing of Rainfall Stations and Coordinates.

Name	Number	Latitude (°S)	Longitude (°W)	Altitude (m)	Name	Number	Latitude (°S)	Longitude (°W)	Altitude (m)
ALTAMIZA	29	5.07	79.73	2600	LOS ALISOS	21	4.97	79.53	2150
ANIA	17	4.85	79.48	2450	MALLARES	6	4.85	80.73	45
ARANZA	16	4.85	79.58	1300	MIRAFLORES	11	5.17	80.62	30
ARDILLA	40	4.52	80.43	150	MONTEGRANDE	13	5.35	80.72	27
ARENALES	8	4.92	79.85	3010	MONTERO	48	4.63	79.83	1070
ARRENDAMIENTOS	57	4.83	79.90	3010	MORROPON	10	5.18	79.98	140
AUL	42	4.55	79.70	640	NANGAY DE MATALACAS	18	4.87	79.77	2100
AYABACA	2	4.63	79.72	2700	OLLEROS	53	4.70	79.65	1360
BARRIOS	33	5.28	79.70	310	PACAYPAMPA	23	5.00	79.67	1960
BERNAL	36	5.47	80.73	32	PAITA	67	5.08	81.13	6
BIGOTE	34	5.33	79.78	200	PALOBLANCO	28	5.05	79.63	2800
CANCHAQUE	35	5.37	79.60	1200	PALTASHACO	30	5.12	79.87	900
CHALACO	25	5.03	79.80	2250	PANANGA	43	4.55	80.88	450
CHIGNIA	60	5.60	79.70	360	PARAJE GRANDE	65	4.63	79.92	1500
CHILACO	3	4.70	80.50	90	PASAPAMPA	31	5.12	79.60	2410
CHULUCANAS	9	5.10	80.17	95	PICO DEL ORO	41	4.53	79.87	1325
CHUSIS	37	5.52	80.82	12	PIRGA	61	5.67	79.62	1510
CIRUELO	64	4.30	80.15	202	PUENTE INTERNACIONAL	63	4.38	79.95	408
CORPAC	15	5.20	80.62	49	SAN JOAQUIN	32	5.13	80.35	100
ESPINDOLA	49	4.63	79.50	2300	SAN MIGUEL	12	5.23	80.68	29
FRIAS	20	4.93	79.93	1700	SAN PEDRO	27	5.08	80.03	254
HUANCABAMBA	68	5.23	79.43	1052	SANTO DOMINGO	24	5.03	79.87	1475
HUAR HUAR	62	5.08	79.47	3150	SAPILLICA	56	4.78	79.98	1446
HUARA DE VERAS	47	4.58	79.57	1680	SAUSAL DE CULUCAN	4	4.75	79.77	980
HUARMACA	14	5.57	79.52	2100	SICCHEZ	44	4.57	79.77	1435
JILILI	46	4.58	79.80	1330	SUYO	39	4.50	80.00	250
LA ESPERANZA	7	4.92	81.07	12	TACALPO	50	4.65	79.60	2010
LA TINA	1	4.40	79.95	427	TALANEO	26	5.05	79.55	3430
LAGARTERA	54	4.73	80.07	307	TAPAL	55	4.77	79.55	1890
LAGUNA RAMON	59	5.55	80.67	9	TEJEDORES	5	4.75	80.25	230
LAGUNA SECA	19	4.88	79.48	2450	TIPULCO	52	4.70	79.57	2600
LANCONES	45	4.57	80.47	110	VADO GRANDE	38	4.45	79.60	900
LAS LOMAS	66	4.65	80.25	265	VIRREY	58	5.53	79.98	230

Color Figure Captions

6. Pseudo-Colored Rainfall Distribution from Delauney Triangulation of Stations.

8. Pseudo-Colored Mesh from Weighted Averaging and Rainfall Stations.

10. Pseudo-Colored Rainfall Distribution from Weighted Average Gridding of Stations.

11. Rainfall Distribution in Northwestern Peru for January 24-27, 1983.

12. Rainfall Distribution in Northwestern Peru for May 19-22, 1983.

13. Pseudo-Colored Precipitation and Pseudo-Color Temperature Contours from Gridded Global Station Data.

14. Weighted Average Gridding Applied to Atmospheric Profile Data.

15. Data Explorer Visual Program Illustrating Weighted Average Gridding of Stations.

16. User Interface of a Data Explorer-based Application for Studying Peruvian Rainfall Data.

17. Conditions in the Pacific Ocean on January 24, 1983.