NOTES AND CORRESPONDENCE

Digitizing Chart Recorder Data: Coordinate System Conversion for Rain Gauges and Similar Recording Instruments

M. TODD WALTER

Department of Agricultural and Biological Engineering, Cornell University, Ithaca, New York

PAUL MUTCH

USDA/Agricultural Research Service, Biosystems Engineering Department, Washington State University, Pullman, Washington

CHRISTA D. SALMON

Department of Ecology and Systematics, Cornell University, Ithaca, New York

DONALD K. MCCOOL

USDA/Agricultural Research Service, Biosystems Engineering Department, Washington State University, Pullman, Washington

LARS O. HEDIN

Department of Ecology and Systematics, Cornell University, Ithaca, New York

18 May 1998 and 8 September 1998

ABSTRACT

Though chart recorder data are widely available for many environmental parameters over long periods of record, the nonorthogonal coordinate systems of many types of recorder charts complicates data extraction such that it is almost always manual; a tedious chore involving hand-counting tiny red boxes. Perhaps the most common data available on these types of charts are precipitation from weighing-type recording rain gauges. Unfortunately, because most digitizer software assumes the data being digitized have an orthogonal coordinate system, the nonorthogonality inherent in these charts prevents simple digitization of the data. If methodologies exist for electronically extracting rain gauge data, they are not widely known; recent texts still outline manual break-point tabulation in detail with no reference to any software or published methodologies for electronic data extraction. A simple procedure is presented here to transform digitized chart recorder data into meaningful, accurate data using readily available software. A comparison of manually read precipitation data and precipitation data extracted with a digitizer showed average differences of less than half the smallest marked increment on the charts for both precipitation amount and time. Differences were almost exclusively due to human bias in locating break points rather than true error.

1. Introduction

While hourly and daily precipitation data are readily available for most of North America, research often requires more precise temporal distributions of precipitation. While tipping-bucket rain gauges, datalogging, and other related technological advances simplify the collection and downloading of continuous precipitation data, the most widely available form of continuous precipitation data is still chart recorder data from weighing-type gauges, especially for historical periods of record, that is, earlier than the 1980s.

Traditionally, recording rain gauge (RRG) charts have been manually read, counting tiny boxes to determine the time lapse and precipitation increase between break points in the slope of an inked line that traces the precipitation record. If less tedious methods for extracting rain gauge data exist, they are not widely known; recent texts like the manual on watershed instrumentation and measurement (Ffolliot 1990) still outline manual break-point tabulation in detail with no reference to methods of electronic data extraction. As a direct consequence

Corresponding author address: Dr. M. Todd Walter, Dept. of Agricultural and Biological Engineering, Cornell University, Riley-Robb Hall, Ithaca, NY 14853-5701.
E-mail: mtw5@cornell.edu

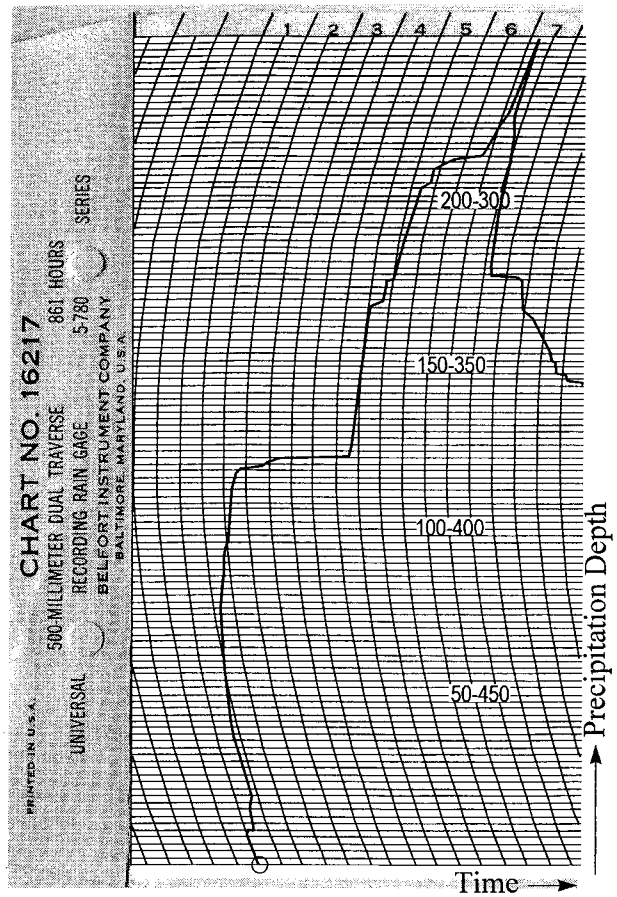

FIG. 1. An example of a portion of a recording, weighing-bucket rain gauge chart (data: Chiloe, Chile).

of the recording mechanism typically used in weighing-bucket RRGs, there is a curvilinear (nonorthogonal) component in rain gauge charts; namely, the lines depicting units of time are physically curved across the horizontal lines that depict units of precipitation (Fig. 1). This same mechanism is used in a variety of instrumentation to measure various properties such as stream stage, water table, temperature, humidity, and barometric pressure. Digitizing, using most readily available software, assumes an orthogonal coordinate system. Thus, direct digitizing does not accurately capture the time component from charts of this type.

A digitizer can be used to locate the physical location of points along a cumulative precipitation trace on a chart [chart dimensions are typically 15.2 cm $\times$ 29.2 cm (6 in. $\times$ 11.5 in.)]. The equations derived in the following section transform each digitized point, (x_i, y_i), into meaningful precipitation (P_i) and time (t_i) information. Of course this method could be extended to any similar chart data with minimal modifications. Though this report continually makes use of digitizer technology, the same approach could be used with any digital, image capturing devices such as scanners or digital cameras.

2. Derivation

Figure 2 shows an idealized nonorthogonal coordinate system analogous to a typical RRG chart. For convenience and computational ease, the *x* and *y* axis of the orthogonal coordinate system coincide with the horizontal line along the bottom edge of the chart and the vertical line along the left edge, respectively. Digitized data are associated with the orthogonal, or *x*–*y* system. A conversion is needed to transform the *x*–*y* data into *P* (precipitation)–*t* (time) data where the lines of constant *t* are curved. The variables x_i and y_i denote any point in the *x*–*y* system and t_i and P_i are the associated points in the *t*–*P* system. As can be seen in Fig. 2, *P* and *y* share an axis, hence, P_i and y_i are located at the some point on the vertical axis.

a. Precipitation conversion

It is apparent from Fig. 2 that the precipitation axis in the nonorthogonal system and the *y* in the orthogonal are superimposed:

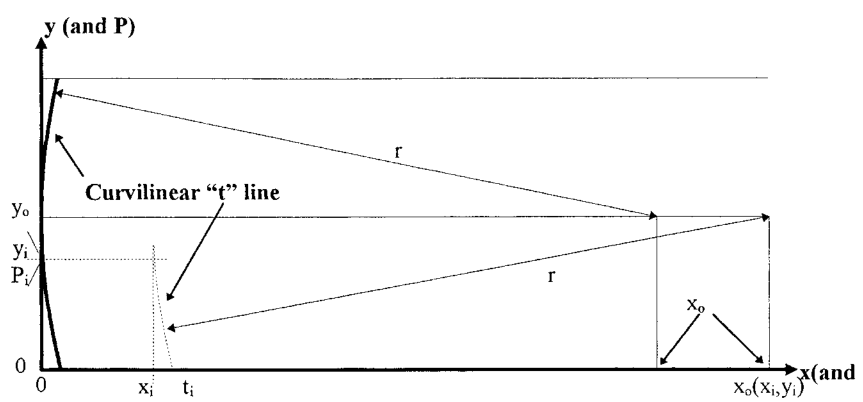

FIG. 2. The nonorthogonal *t*–*P* coordinate system used with RRG charts overlaid with an orthogonal *x*–*y* coordinate system.

$$P_i = C_P y_i, \tag{1}$$

where P_i is the cumulative precipitation at time t_i, y_i is the vertical location of point i, and C_p is the precipitation conversion constant (units of P per unit of y).

Dual-transverse gauges accumulate precipitation in both directions, up and down. As long as evaporation is negligible P_i can be determined with

$$P_i = P_{i-1} + C_p|y_i - y_{i-1}|. \tag{2}$$

When evaporation is significant, simple spreadsheet routines can be created to appropriately adjust tabulated data created with Eq. (1).

b. Time conversion

The curved lines representing units of time on the charts are arcs of a circle with radius r, where r is the distance from the tip of the recording gauge's pen to the pivot point of the pen arm. The function that converts digitized x–y data in the orthogonal system into time, t, in the nonorthogonal system is based on the equation of a circle of radius r centered at x_0, y_0:

$$r^2 = (x - x_0)^2 + (y - y_0)^2, \tag{3}$$

where r is the radius, (x_0, y_0) are the physical coordinates of the circle's center, and (x, y) define the physical location of a point on the arc. In Eq. (3), y_0 is a constant equal to half the total height of the chart and x_0 necessarily varies due to the motion imparted by the clock-driven motor during recording; as time progresses, x_0 moves to the right at a constant speed. Thus, x_0 is a function of x_i and y_i (Fig. 2). With the orthogonal and nonorthogonal systems superimposed, as in Fig. 2, we can conveniently set $t = 0$ at $x = 0$. If C_t is a factor that converts units of x into units of t, then t_i equals $C_t x_i$ for all points where y_i is zero. The general function for t_i for any digitized point x_i, y_i can be found by using Eq. (3) to find the point where the arc on which (x_i, y_i) lie intersects the x axis. From Fig. 2 and Eq. (3), x equals t/C_t when y is zero:

$$r^2 = \left(\frac{t_i}{C_t} - x_0\right)^2 + (0 - y_0)^2, \tag{4}$$

where t_i is the time associated with point (x_i, y_i) on a chart and C_t is the ratio of the units of t to units of x. Rearranging Eq. (4) and raising both sides of the equation to 0.5 gives

$$\pm\sqrt{\left(\frac{t_i}{C_t} - x_0\right)^2} = \pm\sqrt{r^2 - y_0^2}. \tag{5}$$

Because t/C_t is always less than x_0, the sign on the left-hand side radical is negative. Similarly, the sign on the right-hand side radical is positive. Solving Eq. (5) for t gives

$$t_i = C_t\left(x_0 - \sqrt{r^2 - y_0^2}\right). \tag{6}$$

As mentioned earlier, x_0 is not a constant but unique for each t_i. Setting x and y in Eq. (3) equal to x_i and y_i, respectively, Eq. (3) can be solved for x_0. Rearranging Eq. (3) and raising both sides of the equation to the power of 0.5 gives

$$\pm\sqrt{(x_i - x_0)^2} = \pm\sqrt{r^2 - (y_i - y_0)^2}. \tag{7}$$

As in Eq. (5), the physical description of the chart and recording mechanism are such that the left-hand side radical is negative and the right positive. Solving for x_0 gives

$$x_0 = x_i + \sqrt{r^2 - (y_i - y_0)^2}. \tag{8}$$

Replacing x_0 in Eq. (6) with the right-hand side of Eq. (8), the transformation for time is

$$t_i = C_t\left[x_i + \sqrt{r^2 - (y_i - y_0)^2} - \sqrt{r^2 - y_0^2}\right]. \tag{9}$$

As discussed earlier, r and y_0 are constants determined by the width of the chart and pen-arm length for the recording instrument, respectively, and C_t is the ratio of measured time per physical length on the chart (e.g., 1 h per centimeter of chart). The values, x_i and y_i, are any orthogonal coordinate pair obtained from digitizing.

3. Method

Six charts of recording gauge precipitation data were both manually and electronically tabulated for comparison. The data represented widely differing precipitation regimes; two charts were from the Palouse Conservation Field Station (PCFS) located near Pullman, Washington, which is in a semiarid region with approximately 0.25 m of average annual precipitation and four charts were from a research site in coastal Chile, which is a very wet area with approximately 6 m of average annual precipitation. Because the charts had different precipitation and time units, all comparisons are in units of the smallest marked increment (smi) on the charts; for example, for standard 6 in.–24 h charts, 1 smi in the horizontal direction is equivalent to 15 min of time (3.0 mm of physical chart length) and 1 smi in the vertical direction is approximately equivalent to 0.05 in. of precipitation (1.3 mm of physical chart length).

Rain gauge data were digitized using CalComp digitizers (DrawingBoard III and model 9500) and GIS software (ARC/INFO and GRASS4.1). The rain gauge charts were Belfort Instrument Company Dual Transverse Universal Rain Gauge charts; the PCFS data were on 12 in.–24 h charts and the Chile data were on 250 mm–31 day charts. The two PCFS charts each had seven traces, one for each day of a week. The dimensions of the recording area were 15.2 cm × 29.2 cm (11.5 in. × 6 in.).

The data were digitized into GIS files using the charts' physical dimensions for the x and y axis such that (0 cm, 0 cm) was the lower left-hand corner of each chart as shown in Fig. 2. The digitized data were transferred to ASCII format and loaded into Microsoft Excel. Equa-

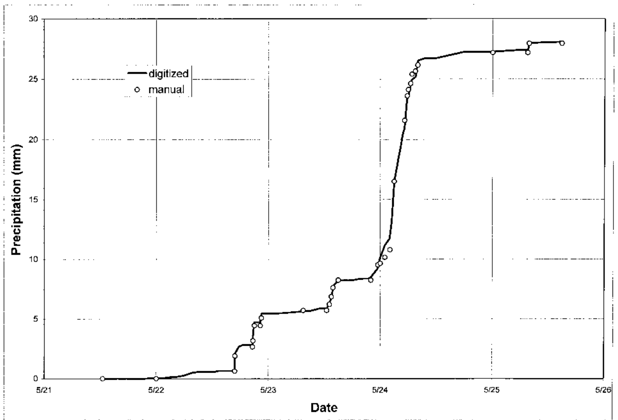

FIG. 3. Example comparison of digitized and manually tabulated precipitation data for PCFS.

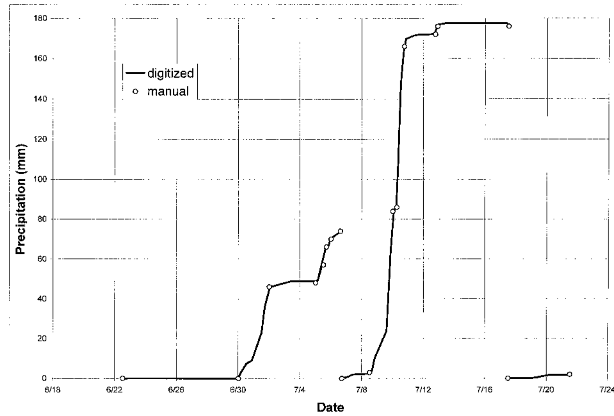

FIG. 4. Example comparison of digitized and manually tabulated precipitation data for Chile. Breaks in the line are dates in which the gauge was emptied.

tions (1) and (9) were used to convert the digitized x–y data into t–P data. The units of r had to be consistent with the digitized data; r was 20.3 cm (8 in.). To convert to meaningful units for comparison, C_p was 7.9 smi cm^{-1} and 8.2 smi cm^{-1} for the PCFS and Chile charts, respectively, and C_t was 3.3 smi cm^{-1} and 4.9 smi cm^{-1} for the PCFS and Chile charts, respectively. The value of y_0 was 7.6 cm for both the PCFS and Chile charts.

The transforming equations for the PCFS data were

$$P_i = 7.9y_i, \tag{10}$$

$$t_i = 3.3\left(x_i + \sqrt{20.3^2 - (y_i - 7.6)^2} - 18.82\right). \tag{11}$$

The transforming equations for the Chile data were

$$P_i = 8.2y_i, \tag{12}$$

$$t_i = 4.9\left(x_i + \sqrt{20.3^2 - (y_i - 7.6)^2} - 18.82\right). \tag{13}$$

Because of the relative ease of digitizing compared to manual tabulation, the digitized datasets were typically at least four times larger than the manually read sets. Comparisons were made by extracting digitized data points that most closely matched points in the manually read data; a total of 213 points were analyzed (91 for PCFS and 122 for Chile). To reduce human bias, the charts were manually read and digitized by different people. Though this precaution was taken, only three or four people were involved in data distillation, all hydrologists and agricultural researchers; that is, there is still significant human bias in this study.

4. Results

Figures 3 and 4 show examples of manually tabulated data (open symbols) superimposed on transformed digital data (solid lines) for both PCFS and Chile charts. Unlike the rest of this paper, Figs. 3 and 4 are presented in units of millimeters for precipitation and days for time.

Table 1 shows some comparisons between the manually tabulated data and the data extracted via digitization and transformed with Eqs. (10)–(13). Excluded from the comparisons in Table 1 are 14 points for which obvious human errors had been made; these accounted for all the gross discrepancies between the two datasets. A gross discrepancy is defined as an absolute difference between the two datasets of greater than one smi. Of these, 13 were errors in manual tabulations and 1 was an apparent digitizing error (an obviously stray or errant point). Twelve of the thirteen manual-reading errors were from simple miscounting and one from a failure to account for a period of excessive evaporation.

For the PCFS data there was only one gross discrepancy and it was with respect to the time units. There were 13 gross Chile discrepancies, 12 with respect to precipitation and 1 related to time. There were four Chile discrepancies greater than 2 smi.

5. Discussion

As can be seen in Figs. 3 and 4 and Table 1, discrepancies were relatively small on average, approaching the precision of the charts, which is typically about

TABLE 1. Comparison between manual and digitized data for PCFS and Chile charts.

	Standard error* (smi)	R^2	$\lvert\text{Difference}\rvert_{\text{avg}}$ (smi)	$\lvert\text{Difference}\rvert_{\text{max}}$ (smi)
Chile				
Time	0.29	1.00	0.22	1.36
Precipitation	0.82	1.00	0.41	4.88
PCFS				
Time	0.28	1.00	0.16	1.66
Precipitation	0.42	1.00	0.40	0.53

* The manual data were assumed to be the independent data and the digitized data the dependent.

0.2 smi. Detailed evaluation of the discrepancies showed that there were generally no identifiable true errors but that differences arose primarily due to human bias in locating break points on the charts. For example, 1) manual tabulators tended to round to the nearest marked unit when the break point was close to a marked ''grid line''; 2) when the slopes on either side of a break point were similar, the location of the break point was relatively subjective and it was difficult to identify any consistent trends or true ''errors'' in either the manually tabulated or digitized data; and 3) bias arose when slopes experienced a gradual transition. Typically the digitized data accurately depicted these transitions with several points whereas the manually tabulated data represented transitions with a single, ''average'' point and often the person digitizing did not extract that same point. Due to steep slopes in the inked trace (high precipitation intensities), small differences in locating the time of the break point resulted in large differences in precipitation. This was especially true for the Chile charts that had typical slopes in the trace of 35 to 40 vertical smi to horizontal smi. Therefore, a difference of 0.1 smi in locating the time of a break point resulted in a difference of 4 smi in precipitation. As an additional note, the width of clean inked lines on the charts used was approximately 0.1–0.2 smi.

Another source of discrepancy was evaporation. Digitized data generally accounted for evaporation more precisely than manual data. The manual tabulators occasionally overlooked or adjusted for small changes in evaporation, which created cumulative errors over long periods. This was especially problematic for the PCFS charts, which often experience long dry period or very tiny rainfall intensities that when combined with evapotranspiration appear as gentle rises and dips in the inked traces on charts. These dips were most often read as a ''no change'' in precipitation by the manual tabulators but recorded, correctly, as evaporation and precipitation by the digitizers. The effects of the evaporation errors were most significant at the end of the traces due to the accumulation of errors.

Much of the error observed in this report may be better identified and discussed by researchers focusing specifically on digitizing technology, human bias in reading charts, and science and technology of measurement.

As a final note, while it was not explicitly quantified, the time to digitize and transform data was only a small fraction of the time required to manually read the same information. A year's worth of data for the Chile site, 12 charts, were digitized and transformed in less than two hours. Based on the approximate time to manually tabulate chart recorder data, digitizing was at least three or four times faster than manual. This estimate did not include the additional time required for electronic data entry of the manually tabulated data.

6. Conclusions

The transformation methodology presented here allows for quick, easy, and accurate distillation of chart recorder data. No complicated or obscure software or hardware are required. In a comparison of six charts representing a variety of precipitation regimes and chart scales, the proposed method was at least as adequate, and in many ways much better, than the traditional, manual tabulation approach. While advances in the technology of data collection may someday make chart recorders obsolete, they are currently used throughout the world for a wide range of measurements. Also, because chart recorder data are usually the best source of continuous data for historical records, methods like the one presented here for rapidly and accurately extracting such information may continue to be valuable for many years.

Acknowledgments. This paper is a contribution to the Cordillera Piuchue Ecosystem Study (CPES) and has been supported in part by a grant from the A. W. Mellon Foundation.

REFERENCES

Ffolliot, P. F., 1990: *Manual on Watershed Instrumentation and Measurements.* ASEAN-US Watershed Project, Laguna, Philippines, 194 pp.