Journal of Hydrology (N.Z.) Vol. 26, No. 2, 1987

DIGITISING PLUVIOGRAPHS

John Sansom

New Zealand Meteorological Service
P.O. Box 722, Wellington

ABSTRACT

A scheme for digitising daily pluviographs from Dines-type automatic raingauges is described. The data set derived from digitisation consists of a division of time into periods of variable length in which rain is assessed as having fallen at a steady rate. The results obtained from the digitised data contain fewer errors than data extracted manually from the pluviographs. It is estimated that 98% of the rainfalls in rain periods of any length lie within 0.1 mm of the true amount, and that the time resolution of the system is about 2.5 minutes.

INTRODUCTION

Most rainfall measurements for the New Zealand Meteorological Service's rainfall observation network are made using the standard 127 mm diameter gauge which is read once a day in units of 0.1 mm. There are several thousand of these gauges in New Zealand at present, but there are also about 150 automatic recording gauges which produce a continuous trace or pluviograph. Some automatic gauges have charts which need be changed only once a month or once a week, but the vast majority produce a chart every day. At present the pluviographs have the month's maximum rainfall in 10, 20 and 30 minutes and 1, 2, 6, 12, 24, 48 and 72 hours extracted from them. Rainfall amounts for each clock hour are also estimated from gauges at stations run by New Zealand Meteorological Service staff. This extraction of maxima and hourly estimates requires double handling of the charts; once for the maxima and again for the hourly estimates.

Any scheme to digitise the daily Dine's pluviographs requires that currently extracted information still be available with no greater input of labour. In addition, the scheme should extract all the available information from the charts with one handling (i.e. the digitisation) so that any future requirements will be met by this extracted data. Finally, the amount of data produced from the current network of stations is too large for manual checking and so automatic means of validating the data are required. No currently available scheme met all these conditions or was suitable for adapting to the equipment available to the New Zealand Meteorological Service. A scheme described by Thompson (1971) was not suitable for the available hardware, and the data, produced by the trace following technique rather than that of taking breakpoints (see the next section), would not define the timings of changes in rainfall rate, thus precluding the extraction of accurate rainfall intensity statistics. Finally, the scheme's checking procedures were insufficiently automated, so that manpower saved by digitising the charts might well be lost in manually checking the results.

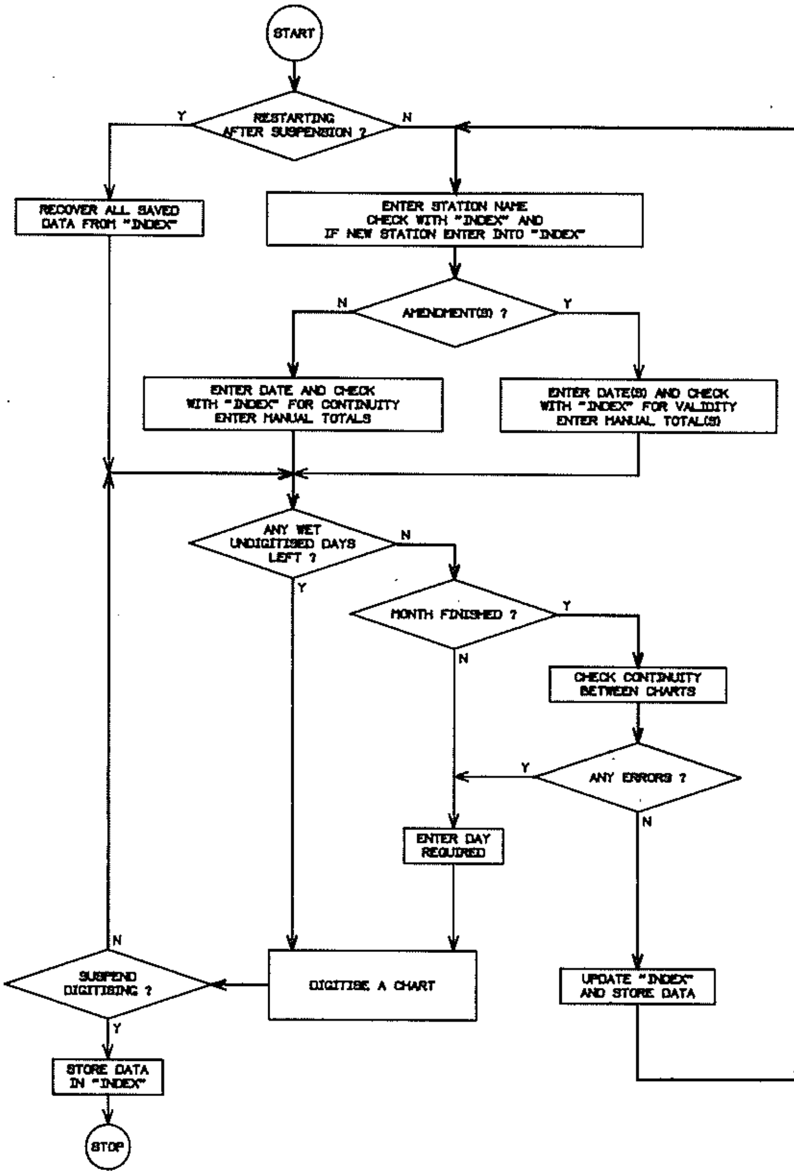

FIG. 1. Flow chart illustrating the digitisation of a monthly batch of pluviographs (the heavier line shows the most common route).

The digitisation scheme described herein started operation in January 1986. It will provide a data base flexible enough to determine both the maximum falls for the standard periods and the hourly rainfalls. In addition, by simply re-processing the data, information such as rainfall rate statistics, exceedence frequencies and maxima for periods other than the standard ones can be found.

To test the digitisation scheme, charts from a station where both hourly and maximum falls had been manually extracted needed to be digitised. Charts for Invercargill were suitable since the rainfall there is not excessive, being on average just over 1000 mm per year, and is well spread with very dry periods of more than a week. The New Zealand Meteorological Service changed to metric measurement in 1971, so 1972 was chosen as a starting year for this test data. The years 1972 to 1975 inclusive were digitised by the writer and the years 1976 to 1979 inclusive by potential future users of the scheme (i.e. four operators digitised one year each). These four operators and the writer each digitised December 1975 four times (20 digitisations of the same month) so that comparisons could be made among the various operators.

THE DIGITISATION SYSTEM

Pluviographs arrive in monthly batches from various stations. The archiving of a data set suitable for further processing into maximum falls requires digitising the charts, and transforming the data.

1) *Digitising charts*

To digitise charts, pluviographs are placed on a graphics tablet which electronically generates numbers in response to applied pressure from an attached pen; these numbers are equivalent to the position of the pen on the tablet. The graphics tablet is controlled by a personal computer, with the digital data being temporarily stored on floppy disc before transfer to the Meteorological Service's main computer. The trace on a pluviograph consists of a series of straight lines, either horizontal when no rain fell, or rising during periods of rain, with heavier rain causing a steeper rise. Only the break points are digitised i.e. those points where one steady rate of rainfall changes to another steady rate.

The pluviographs are digitised in monthly batches (Fig. 1). The station's name is checked in an index file from which the station's number and date of last digitisation are read. The number is the prime identifier of the station, and the date provided can be checked against the date of the current batch to ensure continuity of the archive. Before digitisation starts, the manual totals are entered; these come from a manual gauge positioned near the autogauge to provide accurate daily readings and are used at the data transformation stage to provide a scaling factor.

Charts for each wet day are digitised. After placing the chart on the graphics tablet a reference line must be entered which not only enables points taken from the randomly-placed chart to be corrected for horizontal display on a monitor screen but also provides a time scale. As the digitisation of a chart proceeds, a representation of it is drawn on the monitor screen. Facilities are available for editing the points taken, changing the date, the check gauge total, or the time of origin; and designating periods during which the trace is missing or invalid. At the end of the chart, the on and off times and the amount of rain represented by the pen trace are computed and displayed for checking.

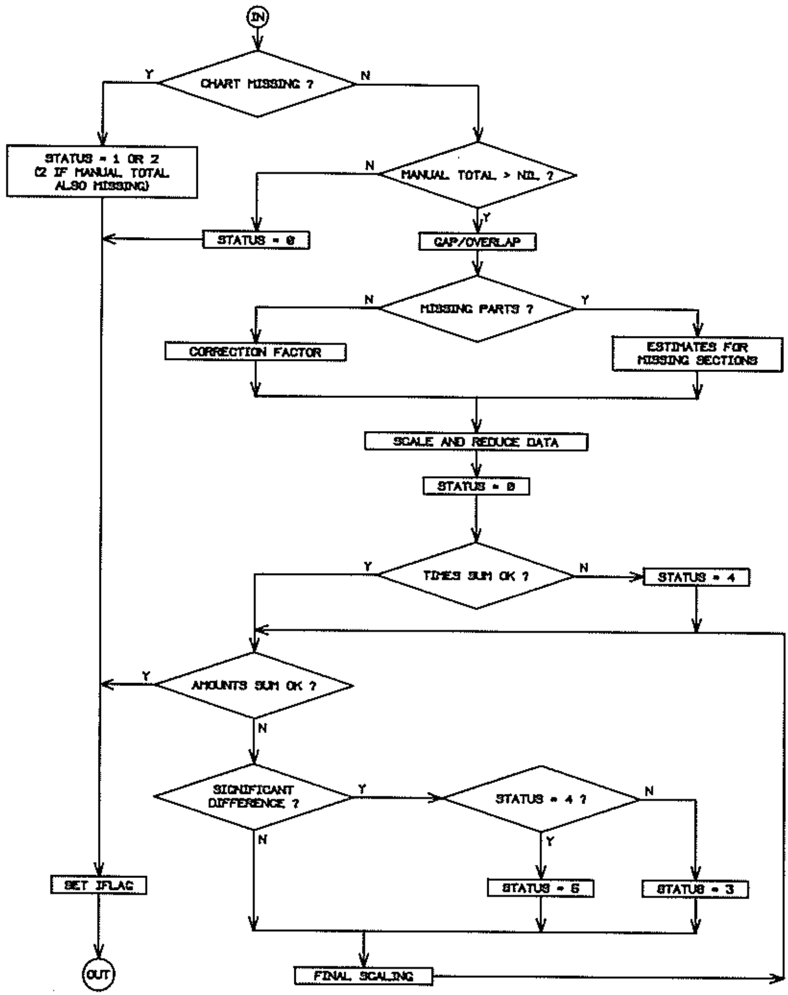

FIG. 2. Flow chart illustrating the transformation of the digitisation data into periods of time and amounts of rain.

At the end of the monthly batch, the continuity of the pen trace from one chart to the next is checked before the data taken is stored on floppy disc. At this point there have been sufficient checks on names, dates, manual totals and continuity of trace to ensure that the digitised data are labelled with the correct station number and date, and that correct scaling factors for both the time and rain axes have been provided. For any particular station, the data archive, from the date of the first chart that was digitised to the last, is continuous in time. When the temporary data store on the floppy disc becomes full this data must be transferred to the main computer.

2) *Data transformation*

At this stage the digitised data are just numerical representations of the pluviograph traces, but can be transformed to a true represention of how rainfall varied with time. The logical flow of this transformation, which is performed for one day at a time, is shown in Figure 2. "STATUS" returns a value specifying how good the transformation is for that day. Initially, those days which were dry or for which the chart was missing are recognised and require little treatment. The first requirement for a day with data is to deal with the junction of this day with the previous day. The time on of one chart is seldom the same as the time off of the previous one — there could be a gap or an overlap. The action necessary to fill the gap or share out the overlap depends on the state at the ends of the charts — was it dry or raining or was the trace missing? The state at the end of the previous chart is remembered by "IFLAG" (see Fig. 2) and the state at the beginning of the present one can be easily found and the appropriate action taken. Of the 18 different types of chart junction most can be dealt with within the present day, but a few require an adjustment to the end of the previous day.

The next steps result in the actual conversion of the data. If the trace is complete a scaling factor is calculated; alternatively estimates of how much rain to accredit to the missing sections are made. As exactly horizontal lines cannot be digitised a criterion for recognising when no rain is falling is required; the criterion adopted is that a minimum rainfall rate of 0.1 mm/hr is required. Another difficulty is recognising short dry periods within a basically wet time; the criterion adopted here was that, during any period between wet periods, at least 0.1 mm of rain must have fallen otherwise it was a dry period. Using these criteria and the scaling factor, the data is scaled and put into increments i.e. pairs of numbers which specify an amount of rain (in hundreths of a millimetre) in a certain length of time (in tenths of a minute). Adjacent periods may both be dry or missing or have nearly equal rain rates, in which case such periods are amalgamated to reduce the data to an essential set.

Finally, the status (or how well the transformation has been performed) is found by checking that the lengths of the periods sum to the actual day length and that the increments sum to the manual total. If the difference from the manual total is significant (i.e. over 10% and at least 5 mm of rain on the chart) then the status number is adjusted, but in any case a final scaling is performed. At the end of the transformation, charts with status values of 3 or more must be re-digitised. The final data set consists of a division of time into variable length periods such that, throughout the period, rain fell at a steady rate to give a certain total amount for that period.

Comparison of Digitised Data with Manually Extracted Data

a) *Hourly rainfalls*

For the years 1972 to 1975 the digitised data set was processed to give hourly rainfalls, and these were compared with the hourly rainfalls from the New Zealand Meteorological Service data archive. Since the archived data were derived manually from the same pluviographs that had now been digitised the comparison is a direct one between the methods (Table 1). In both methods the resulting hourly rainfalls are in tenths of a millimetre. An hour where rain can be seen on the pluviograph but the amount is insuffícent to accredit 0.1

TABLE 1—Frequency of differences between manual and digitisation estimates of hourly rain

Manual Estimate	Digitisation Estimate	Number of Differences
Trace	>.3mm	8
Trace	.3mm	5
Trace	.2mm	29
Trace	.1mm	227
Nil	Nil	29504
(Estimates agree and are non-zero)		2504
R+	R±. 1mm	2288
R	R±.2mm	503
R	R±.3mm	159
R	R±.4mm	47
R	R±.5mm	28
R	R±>.5mm	48
(At least one estimate is missing)		140

+R represents a non-zero amount of rainfall.

mm to that hour is said to have had a "Trace" in the manual method. For simplicity, amounts under 0.05 mm were just set to zero in the digitising method. As rain falls on average for only 7% of the available time and methods should agree for the dry hours, the 29504 cases of agreement under dry conditions are not unexpected. There are 2504 cases of exact agreement during rain and another 2288 within 0.1 mm (Table 1). These, together with the dry hours and those Trace hours given 0.1 mm when digitised, account for 97.2% of the time. The 56 cases (i.e. 48 + 8) making up the extremes of Table 1 represent only 0.16% of the time.

The causes of the disagreement between methods in the 56 extreme cases were as follows: 10 cases where, for no discernible reason, the manual estimate was incorrect, 13 cases at the beginning or end of a pluviograph, 30 cases of allocation disagreement, and 3 cases of poor digitisation. As pluviographs are changed daily and the manual gauge read at about, but not necessarily exactly on, 9 a.m. there can be an allocation error between one day and the next at the ends of the pluviographs. The error involved is illustrated in Figure 3 (a); the rain from "B" to "C" continues on from "A" and should be accredited to 8 a.m. to 9 a.m., but a common mistake made during manual estimation of hourly rainfalls is to accredit it to 9 a.m. to 10 a.m. Figure 3 (b) illustrates

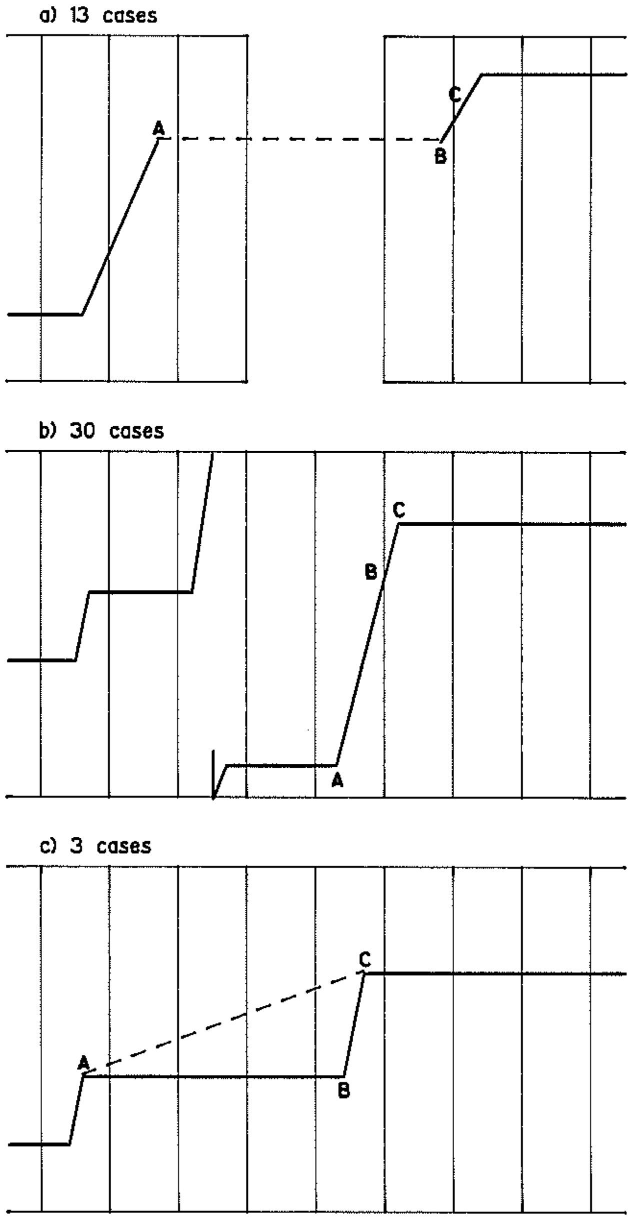

FIG. 3. Schematic partial pluviographs illustrating, a) a common allocation error made at the beginning of a chart b) how allocation errors of rainfall amounts into consecutive hours may occur and c) how the missed digitisation of one breakpoint can lead to wrong amount estimates for several prior hours.

TABLE 2: Mean annual maximum falls and standard deviations for 1972—1979 as estimated manually and from the digitised data set.

Period	Manual Mean	Manual St.Dv.	Digitisation Mean	Digitisation St.Dv.	Ratio of Means
10 mins	45	10	47	11	1.044
20 mins	67	11	67	12	1.000
30 mins	81	13	82	12	1.012
1 hr	109	27	114	34	1.046
2 hrs	158	46	167	47	1.057
6 hrs	275	67	280	66	1.018
12 hrs	360	94	351	83	0.975
24 hrs	414	108	440	111	1.063
48 hrs	507	109	544	134	1.073
72 hrs	628	132	650	129	1.035

(All means and standard deviations in tenths of mm's)

how allocation disagreement arises; if rain is falling at an hour crossing then how should the rain be shared out between these hours? Manually, all the levels "A", "B" and "C" must be estimated whereas, digitally, only the points "A" and "C" are taken and "B" calculated from them. In either case, estimates become less certain as the rainfall rate at the crossing increases (i.e. the pluviograph trace becomes steeper and intersects the hour division line at a more acute angle). In most of these 30 cases it was not possible to decide which method had given the better estimate. Finally, Figure 3 (c) illustrates how poor digitisation may lead to incorrect hourly amounts. If, after a period of nil rain, a breakpoint, (point "B" in this case) is omitted, then the effective trace that has been digitised is that shown by the dashed line between points "A" and "C". The amount of rain between "B" and "C" is thus spread out between "A" and "C". If, where disagreements exist, the reasons for disagreements are in the same ratio as for these 56 cases, then the digitisation method is as good as, or often better than, the manual method.

b) *Maximum Falls*

For each month of the years 1972 and 1979 the digitised data set was processed to give the standard maximum falls of 10, 20 and 30 minutes and 1, 2, 6, 12, 24, 48 and 72 hours, and these were compared to the maximum falls already archived (Table 2). Generally, the two sets of results agree to within a few percent, with the average percentage difference being 3.2%. Those months which had significant disagreements were reprocessed manually (and not by the writer) as an independent check. In all cases the reprocessed results agreed to within 5 percent of the digitised results. The differences between the manual and digitisation estimates (Table 2) are not significant, but, apart from one

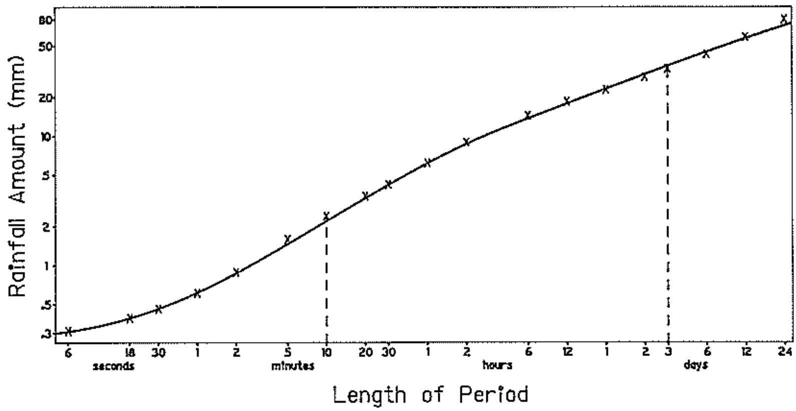

FIG. 4. The variation of the mean of the monthly maximum falls with period length.

overestimate, the manual values tend to be underestimates. Supposing over- and underestimates to be equally likely, the probability of only one overestimate out of 10 estimates is about 1%.

Apart from those for the standard time periods between 10 minutes and three days, maximum falls were found for other time periods (Fig. 4). Since the time periods in the digitised data set are rounded to the nearest 0.1 minute, the shortest time period for which a maximum rainfall could be found was 0.1 minute; the longest time period was 24 days. The values for the shorter time periods are limited by the physical properties of the recording raingauge and are related more to the mean intensities over longer periods than to the true maximum fall for the particular period. However, they do provide a lower limit to the true value (Fig. 4), this limit being considerably higher (about double for 0.1 minute) than might be expected by extrapolating the reasonably linear section between 2 hours and 2 minutes. Therefore, as linear extrapolation must be invalid for periods shorter than 2 minutes, the linearity down to 2 minutes cannot be accepted as proof of validity of the 2 minute mean maximum. However, Hershfield (1982) estimates the ratio of the 5 minute maximum to the 2 minute maximum for stations in the 40°N to 50°N latitude band of the USA as 0.55 and the writer's estimate for Invercargill is also 0.55. Another feature (Fig. 4) is the break, at periods of about 3 or 4 hours, between the linear sections of shorter periods and longer periods. This break is probably due to the maxima for the shorter periods arising from continuous rain whereas the longer periods, which have smaller maxima than might be expected by extrapolation, would generally contain dry periods which would suppress the maxima.

Estimation of Digitisation Systems Accuracy

a) *Time Resolution*

Breuer and Kreuels (1984) measured rainfall using equipment with a claimed time resolution of 5 seconds and fitted their data to the following relationship:

$$R_S = \alpha_{S,L} R_L^{\eta_{S,L}}$$

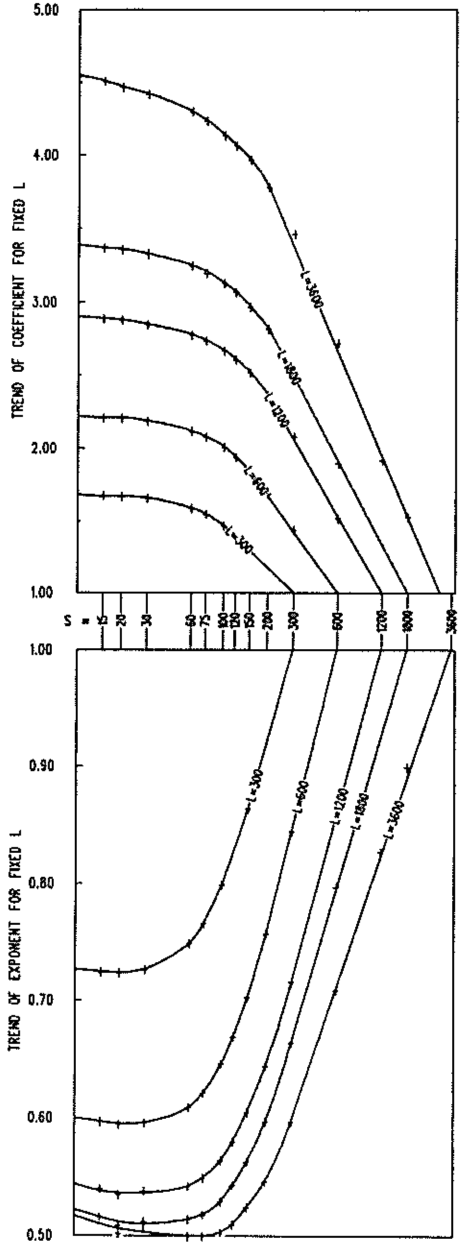

FIG. 5. The trends of $\alpha_{S,L}$ and $\eta_{S,L}$ for fixed long periods, L in seconds, with varying short period, S in seconds.

which relates the measured rain rate R_L during a particular period of length L to R_S, the maximum rain rate among the L/S periods of length S within that particular L. For example, if L were 1 hour and S were 10 minutes then, provided some rain fell in the hour, the 6 clock 10 minutes in that hour would be examined to find that period in which the rain were heaviest; R_L is the mean rate of rainfall over the hour and R_S the rate in the 10 minutes which had the highest rate. For each clock hour in which there was some rain this examination generates a pair of numbers (R_L, R_S) and the equation above was used to model this, with the parameters $\alpha_{S,L}$ and $\eta_{S,L}$ being estimated by a least squares method. The estimations made by Breuer and Kreuels (1984) of the coefficients $\alpha_{S,L}$ and the exponents $\eta_{S,L}$ for many different pairs of S and L show that, for a fixed value of L, there was a linear trend with the logarithm of S. Data derived from digitising rain charts has been similarly treated to provide estimates of $\alpha_{S,L}$ and $\eta_{S,L}$. In accord with Breuer and Kreuels only summertime data were used and only those hours in which the maximum 1 minute rainfall intensity was at least 1mm/hr. These estimates are shown in Figure 5. Since values only exist where L/S is integer, the lines are drawn to help the eye establish the trends that exist in these results. For S values of more than about 200 seconds the trends for fixed values of L are linear (Fig. 5). In the range of S values of between about 200 seconds and 60 seconds a transition takes place so that, for smaller values of S, both $\alpha_{S,L}$ and $\eta_{S,L}$ are nearly constant. This implies that no extra information can be extracted by dividing time into periods shorter than 1 minute than that extractable from 1 minute periods. As the transition to this state starts at about 3 minutes, it can be inferred that the raingauge/digitiser combination enables good time resolution down to 3 minutes, poor resolution for 1 to 3 minutes and no resolution for times of less than 1 minute.

b) *Amount Accuracy*

Twenty digitisations of December 1975 were made by five operators each digitising it four times. Since the choice of which points to digitise is not absolute, the comparison between the 20 samples was made after processing the data into fixed period rainfalls (i.e. "fixed" with respect to the clock). Mean values for each of these fixed periods can be found and hence individual anomalies from the mean. Initially, the whole month was processed to give rain amounts for each quarter of each day. Analysis of Variance showed no significant difference between either people or their individual runs, but in the tested model the interaction term between the operator and the particular quarter day did appear significant. This was interpreted as a systematic error i.e. for reasons unknown a particular operator's results for a particular quarter day tended to cluster about a value away from the overall mean. In the series of anomalies generated by a particular run much compensation occurred from one period to the next e.g. an anomaly of 0.3mm from the mean might well be followed in the next period by an anomaly of –0.3mm. This was interpreted as an allocation anomaly i.e. again for reasons unknown the digitisation has led to an allocation of rainfalls into consecutive quarter days which are noticeably different to the mean values. This type of error occurred when rain was falling at the junction

TABLE 3: Absolute and relative errors in amounts estimated for periods of various length.

Period	Days of Month	Total Error	Number within ±1 of "True" value	Mean Fall	Relative Error
6 hrs	1–29	.62	98.5%	23.7	3%
1 hr	20–27	.68	97.0%	11.1	6%
10 min	20–21	.64	98.0%	3.2	20%
10 min	23–24	.57	99.0%	3.4	17%
2 min	21	.68	97.0%	1.2	57%

(Rainfall amounts are in tenths of millimetres)

of consecutive periods (i.e. as was illustrated in Fig. 3 (b)). On top of these errors will be a random error so that overall:

$$\text{TOTAL ANOMALY} = \text{ALLOCATON ANOMALY} + \text{SYSTEMATIC ERROR} + \text{RANDOM ERROR}$$

The distributions of both the total and allocation anomalies had more extreme values than would be expected for normal distributions. However, both the systematic and random errors appear to be normally distributed, and the total error was taken to be their pooled standard deviation. This provides an estimate of the absolute error involved while, by comparing the total error to the mean rainfall amount, a relative error can be estimated. Errors were re-assessed for other period lengths using fewer days of the month's data as the period length decreased. The results (Table 3) show that, for all periods, the absolute error remains at about 0.06mm. About 98% of the periods — no matter the length of the period — will lie within 0.1mm of the "true" value where the mean of the 20 digitisations is the "true" value and is taken as an estimate of the true value. However, as the mean rainfall decreases with the period length the relative error increases to become about half of the mean value for 2 minute periods.

CONCLUSION

The digitisation of the eight years of pluviographs from Invercargill indicates that the resulting data base is an adequate representation of the actual variation of rainfall with time and no limitations were found due to the taking of just the breakpoints. From this data base the hourly rainfalls and monthly maxima agreed well with those already extracted by manual methods, and where disagreements were found in most cases the manual estimate was found to be in error. Also from this data base many other statistics could be determined. It is estimated that 98% of the rainfalls in rain periods of any length lie within 0.1 mm of the true amount and that the time resolution of the system is about 2.5 minutes.

REFERENCES

Breuer, L. J. and Kreuels, R. K. 1984: Dependence of rain intensity maxima on time base. *Beitr. Phys. Atmosph. 57:* 39-54.

Hershfield, D. M. 1982: 2-minute rainfall extremes. *International Symposium on Hydrometeorology (American Water Resources Association)*: 585-588.

Thompson, P. J. 1971: A rainfall data supply system. *New Zealand Engineering,* 26: 207-209.