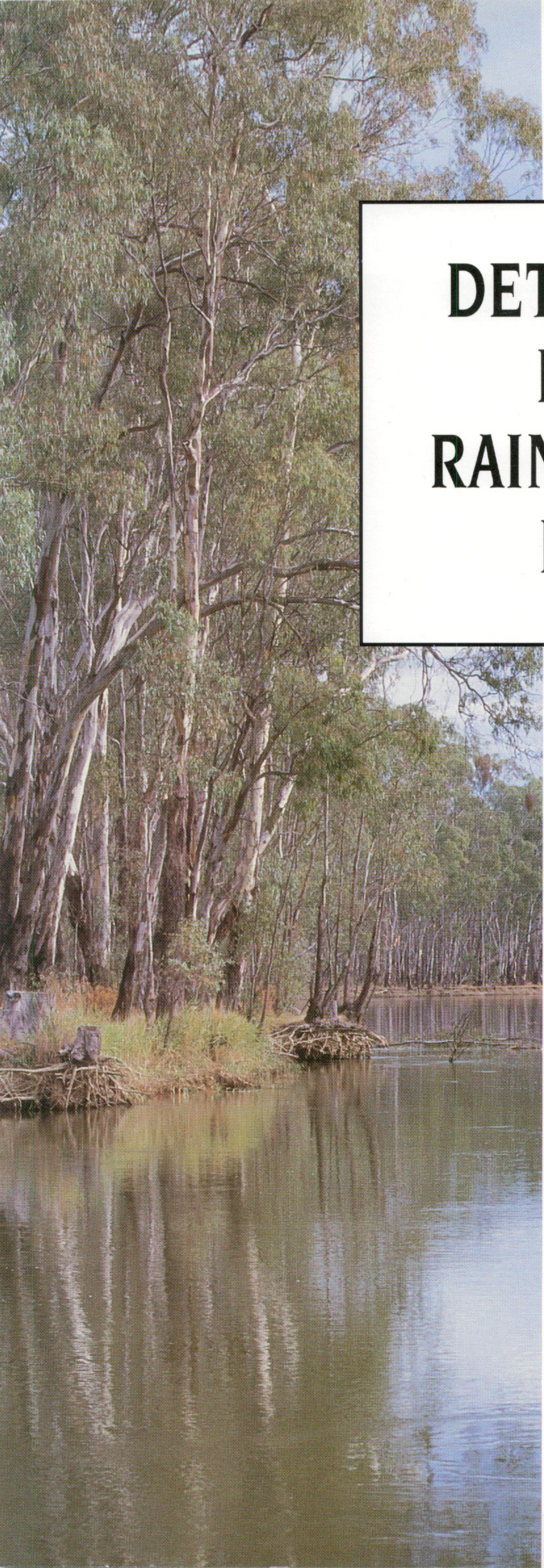

DETECTING DATA ERRORS IN RAINFALL-RUNOFF DATA SETS

Walter C. Boughton

Report 96/2
July 1996

COOPERATIVE RESEARCH CENTRE FOR
CATCHMENT HYDROLOGY

DETECTING DATA ERRORS IN RAINFALL-RUNOFF DATA SETS

Walter C. Boughton

Report 96/2

July 1996

COOPERATIVE RESEARCH CENTRE FOR
CATCHMENT HYDROLOGY

Boughton, W. C. (Walter Charles), 1931 -

Detecting data errors in rainfall-runoff data sets.

Bibliography.
ISBN 1 876006 09 9.

1. Runoff - Mathematical models. 2. Rain and rainfall - Mathematical models.
3. Runoff - Measurement.
I. Cooperative Research Centre for Catchment Hydrology. II. Title (Series: Report (Cooperative Research Centre for Catchment Hydrology); 96/2)

551.4880151

Keywords

Rainfall/Runoff Relationship
Modelling (Hydrological)
Data Handling
Storage
Calibration
AWBM Model

ISSN 1039-7361

PREFACE

One of the core projects in the CRC's Flood Hydrology Program deals with the issue of losses from storm rainfall. The uncertainty in current estimation methods for determining how much rainfall becomes runoff is a significant challenge in both design flood estimation and in flood forecasting.

Since losses from actual storms depend heavily on antecedent conditions, an approach which accounts for the wetting and drying of a catchment between major storms has some attraction. To this end, a modelling approach (using AWBM) has been included in CRC Project D1 "Improved Loss Modelling for Design Flood Estimation and Flood Forecasting". Professor Walter Boughton (currently: Honary Senior Fellow, Griffith University, Queensland) has been involved with the project team on this aspect, and has run two AWBM workshops in Melbourne as part of the CRC program.

In this report, Professor Boughton shows how AWBM can assist in checking the consistency of rainfall/evaporation/runoff data sets, a task which is sometimes overlooked. His advice should be of great assistance to would-be catchment modellers, particularly when it is accompanied by a full set of the AWBM software from the World Wide Web.

I am delighted to have Walter Boughton involved in the CRC in this way, and thank him for this contribution.

Russell Mein
Program Leader, Flood Hydrology

ABSTRACT

There are a number of ways to identify errors and inconsistencies in measurements of hydrologic variables. This report shows how the AWBM catchment water balance model can be used to identify data errors in rainfall-runoff data sets. In this model, average surface storage capacity is the main measure of runoff generating capacity. It is demonstrated that errors in the estimation of catchment rainfall produce substantial effects on the calibrated value of average surface storage capacity in the AWBM. Changes of ±10% in rainfall produce changes of +49%/-32% in the calibrated value of average surface storage capacity, changes of ±10% in runoff data produce changes of -15%/+30% in the calibrated value, while ±10% in evapotranspiration produce changes of -5%/+10%.

When the rainfall was scaled from 0.6 to 1.4 of measured values, the calibrated average surface storage capacity ranged from 7 mm to 962 mm around the calibrated value of 140 mm with no scaling. Over the range from 7 to 962 mm, the coefficient of determination (r^2) between actual and calculated monthly totals of runoff ranged only between 0.892 and 0.986. The high values of r^2 do not reflect the wide range of calibrated storage capacity and are a serious warning to those wishing to relate calibrated parameter values to catchment characteristics for use on ungauged catchments.

Cases in which rainfall data is systematically very low in relation to runoff data are identified by comparing actual runoff to calculated runoff from zero surface storage capacity (simulating an impervious catchment), and by the result of very low calibrated value of average surface storage capacity. Cases in which rainfall data are systematically very high in relation to runoff data are identified by the high values of calibrated average surface storage capacity, and by the infrequency of calculated runoff from the partial area of the catchment with the largest storage capacity. Isolated non-systematic errors in single storm events, and the errors introduced by human modifications of the catchment water balance, are identified by combinations of the above methods, and by checking for large monthly contributions to the sum of squares of differences between actual and calculated runoff.

TABLE OF CONTENTS

1. INTRODUCTION

The General Australian Axiom :
"She'll be right, mate. No worries!"
Boughton's Observation on the Quality of Hydrological Data :
"She will not be right, mate. Start worrying!"

The problems of errors in rainfall and runoff data, even on well-instrumented research catchments, are well known. Rating curves are a major source of error in the measurement of runoff, particularly of large flows. A single raingauge samples the rain falling in 1/30,000,000 of a square kilometre. The sparse network of raingauges over most of Australia produces significant problems of estimating areal rainfall in practical catchment scale studies.

Complex rainfall-runoff models compound the problem by making it difficult to distinguish between errors in input data and errors introduced by the modelling process. The AWBM model (Boughton, 1993, 1996) is capable of relating runoff to rainfall with an accuracy that equals or is better than many of the other models in common use in Australia, but is simple enough to allow for detection of a range of errors in input data, as demonstrated in this paper. This paper describes some errors in some rainfall-runoff data sets and demonstrates methods by which the errors can be systematically detected in other data sets using the AWBM.

Table 1 lists the data sets used in the paper and the sources of the data. Most of the rainfall and runoff data sets were collated by others independent of the writer and for purposes (water yield, flood forecasting, research) other than the preparation of this paper. The catchments cover a geographical range from north Queensland to Tasmania to Western Australia, and a size range from 16 ha to over 3000 sq km. The samples used illustrate that data errors are not confined to any particular subset of catchments based on size or climate or any other factor.

Chapters 2 and 3 show the effect that data errors have on fitted model parameters, and the type of errors which can be encountered. A systematic procedure for detecting errors is outlined in Chapter 4. The Conclusion re-emphasises the need for hydrologists to conduct thorough data checks before venturing into any rainfall/runoff modelling.

Table 1 Data Sets

(a) Sources of Information

Station	Stream	Station	Source of Data
Queensland			
111105	Babinda Ck	The Boulders	Chiew & McMahon, 1993
113004	Cochable Ck.	Powerline	Chiew & McMahon, 1993
132004	Munduran Ck.	Rundle Hills	Boughton, 1993
138007	Mary River	Fishermans P.	Bureau of Meteorology
143019	Oxley Ck.	Beattie Road	Brisbane City Council
-	Cressbrook	Dam	Macintosh, 1994
-	-	Brigalow C1	Boughton, 1985
New South Wales			
210022	Allyn River	Halton	Chiew & McMahon, 1993
215009	Endrick River	Nowra Road	Sharifi, 1996
401554	Tooma River	above Reserv.	Chiew & McMahon, 1993
420003	Belar Ck	Warkton	Chiew & McMahon, 1993
Victoria			
222213	Suggan Buggan	Suggan Buggan	Chiew & McMahon, 1993
227219	Bass River	Loch	Chiew & McMahon, 1993
Tasmania			
307001	Davey River	D/S Xing R.	Chiew & McMahon, 1993
South Australia			
503502	Scott Creek	Scotts Bottom	Chiew & McMahon, 1993
Western Australia			
612005	Stones Brook	Mast View	Chiew & McMahon, 1993

(b) Characteristics of catchments

Station	Area sq km	Rain mm pa	Runoff mm pa	Years of Data
Queensland				
111105	39	5400	4650	1974 - 1987
113004	93	2400	2100	1974 - 1986
132004	55	890	130	1978 - 1985
138007	3120	1440	580	1988 - 1992
143019	156	1025	422	1990 - 1991
Cressb.	210	844	87	1988 - 1992
Brigal.	0.16	670	35	1965 - 1979
New South Wales				
210022	205	1200	450	1977 - 1984
401554	114	1700	1400	1971 - 1979
420003	133	1100	110	1973 - 1984
Victoria				
222213	357	800	150	1972 - 1985
227219	52	1100	330	1974 - 1985
Tasmania				
307001	686	2100	2000	1974 - 1990
South Australia				
503502	27	950	130	1970 - 1985
Western Australia				
612005	15	1000	120	1974 - 1984

2. EFFECTS OF DATA ERRORS ON CALIBRATED AWBM PARAMETER VALUES

The runoff generating capacity of the AWBM can be summarised in a single parameter, the average surface storage capacity, which is calculated as :

$$C_{ave} = C_1*A_1 + C_2*A_2 + C_3*A_3 \qquad (1)$$

where C_{ave} is the average surface storage capacity
C_1,C_2,C_3 are the capacities of the 3 surface stores, and
A_1,A_2,A_3 are the partial areas of the 3 surface stores

If the measurement of catchment rainfall is in error, then so will the calibrated value of the average surface storage capacity. If measured rainfall is less than the true catchment rainfall, then the calibrated value of average surface storage capacity will also be too low because the calibration procedure will fix a lesser storage capacity in order to generate the amount of measured runoff from the lesser rainfall. Similarly, if measured rainfall is more than the true catchment rainfall, the calibrated value of average surface storage capacity will be too high in order to maintain the measured amount of runoff from the higher rainfall.

The effects of rainfall errors on the calibrated value of average surface storage capacity is illustrated in Figure 1 using 4 years of data from the 3120 sq km Mary River catchment at Fishermans Pocket in south east Queensland (138007). The data were collated by the Bureau of Meteorology for flood forecasting purposes. Using the rainfall and runoff data prepared by the Bureau, the calibrated value of average surface storage capacity in the AWBM was 140 mm.

Each daily rainfall was then scaled by a constant factor and the model was again calibrated. Figure 1 summarises the changes in average surface storage capacity for changes in rainfall from 0.6 to 1.4 of measured values.

The most important result shown in Figure 1 is the very large change in calibrated value of average surface storage capacity for relatively small change in rainfall. Changes of ±10% in rainfall produce changes of +49%/-32% in the calibrated value of average surface storage capacity. The range of 0.6 to 1.4 in rainfall scaling produced changes of 0.05 (7 mm) to 3.41 (478 mm) in the calibrated value of storage.

The second important result is that "acceptable" reproduction of measured runoff can be modelled by the AWBM across a wide range of erroneous rainfall. The coefficient of

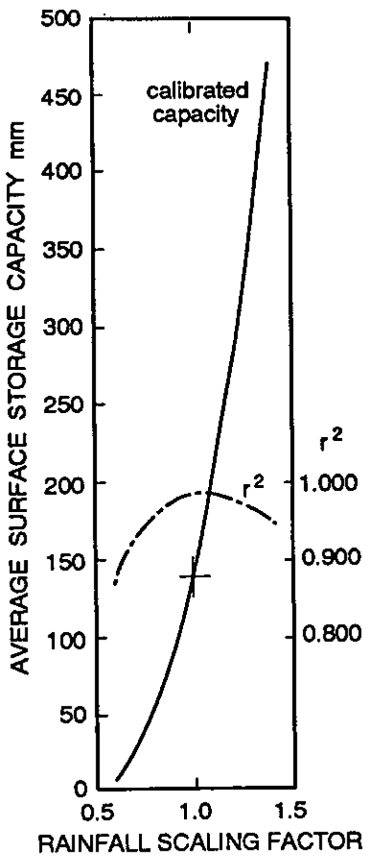

Figure 1: Effect of change in rainfall data on calibrated AWBM storage parameter

determination (r^2) values shown in Figure 1 are computed on measured and calculated monthly values of runoff. In the entire range of calibrated average surface storage capacity, from 7 to 478 mm, the coefficient only ranged from 0.892 to 0.986, i.e. the high values of r^2 give no indication of the wide range of calibrated storage values. *The results give warning of serious problems for those attempting to relate calibrated model parameter values to catchment characteristics for subsequent use on ungauged catchments. The variations in calibrated parameters values due to relatively small errors in estimation of catchment rainfall are so large that the calibrated values are virtually meaningless.*

In addition to errors in rainfall data, there can be errors in the measurement of runoff and in the estimation of evapotrans-piration. Using the same method as above, scaling factors were used to increase and decrease runoff and evapotranspiration, and the AWBM was calibrated to each change in the input data. Table 2 summarises the changes in the calibrated value of average surface storage capacity for changes in either runoff or evapotranspiration.

The results in Table 2 show that errors in runoff produce proportionally less error in the calibrated value of surface storage than do errors in rainfall; and error in the estimation of evapotranspiration produce even less. Changes of ±10% in runoff produce changes of -15%/+30% in the calibrated value of average surface storage capacity, while ±10% change in evapotranspiration produce changes of -5%/+10%. Note that these figures would be river dependent.

Table 2 Calibrated values of average surface storage capacity (mm) and coefficients of determination for changes in input data

Scaling Factor	Rain	Runoff	Evapotr.
0.6	7 (0.892)	310 (0.918)	243 (0.964
0.8	59 (0.960)	219 (0.982)	170 (0.984)
1.0	140 (0.986)	140 (0.986)	140 (0.986)
1.2	277 (0.981)	101 (0.971)	123 (0.985)
1.4	478 (0.949)	62 (0.948)	111 (0.984)

The values in brackets in Table 2 are the coefficients of determination (r^2) from the regression of actual monthly totals of runoff on calculated monthly totals. Again, the high values through-out the table give no indication of the wide range of calibrated storage value due to errors in input data. It is likely that similar results could be found on many of the models which are similar to the AWBM. The tolerance of systematic bias in input data means that the error will not be readily noticed by those involved in practical use of models, but the big changes in calibrated storage value for relatively small changes in input data will make it difficult to transfer calibrated parameter values to ungauged catchments.

3. DETECTING ERRORS IN INPUT DATA

3.1 Too much runoff, too little rainfall

By setting all surface storage capacities to zero, the AWBM can simulate a completely impervious catchment. The only loss from rainfall is the evapotranspiration on each day that rain falls. If measured runoff exceeds calculated runoff from such a model, there is obviously something wrong with the data.

This type of error is illustrated using 14 years of data from the 39 sq km catchment of Babinda Creek at the Boulders in north Queensland (111105). Two rainfall stations (31144, 31141) provide fair representation of the catchment rainfall.

In the 14-year study period, measured runoff totalled 65152 mm (4654 mm p.a.). With all surface stores in the AWBM set to zero capacity, calculated runoff was 64658 mm (4618 mm p.a.), i.e. less than the measured runoff. Figure 2 shows the calculated monthly totals compared with measured monthly totals. The 39 points below the 45° line of equality are months where measured runoff is greater than the calculated runoff using zero capacity.

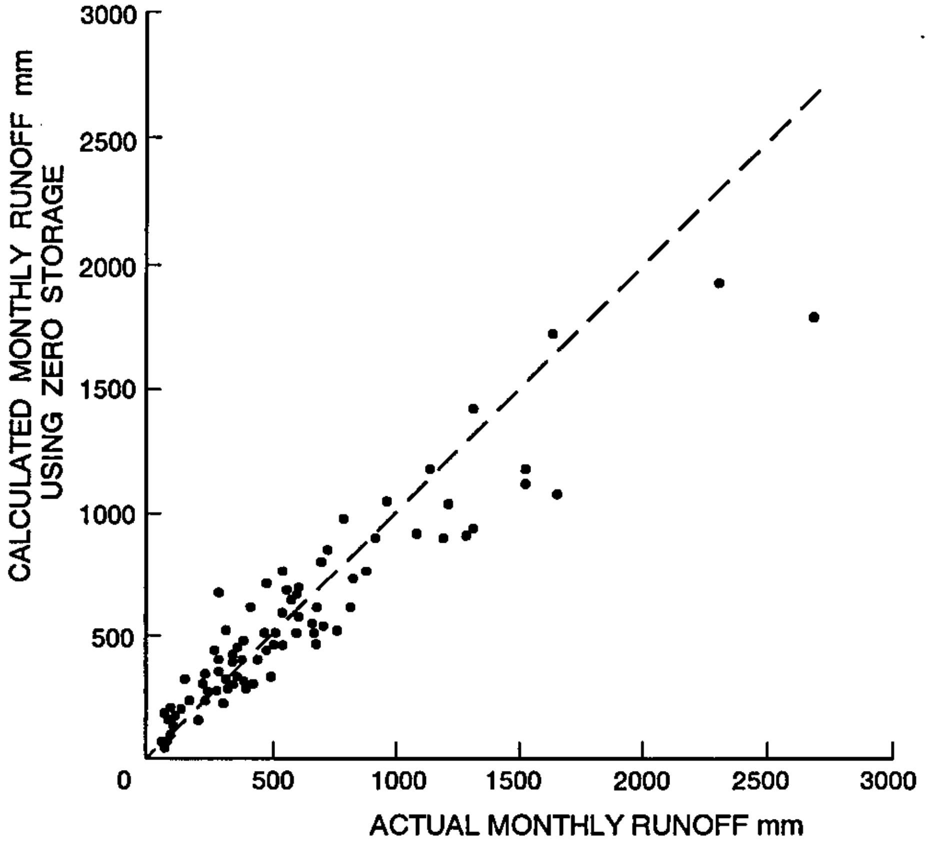

Figure 2: Babinda Creek - actual runoff compared with calculated runoff from zero storage capacity

It should be noted that runoff is generated in the AWBM only by the surface stores. The baseflow parameters serve only to delay some of that runoff by diverting it through the baseflow store, and can only influence the calculated runoff in a minor indirect way by affecting the amounts in baseflow storage at the start and end of the calculation. The baseflow parameters can affect the number of months in which measured runoff exceeds the calculated runoff from zero storage, and can affect the coefficient of determination between measured and calculated monthly totals. Figure 3 shows the effects of altering the baseflow parameters on the number of months in which measured runoff exceeds calculated runoff. The smallest number of 39 months was obtained with a baseflow index (BFI) of 0.65 and a daily recession constant (K) of 0.98. These values were used to calculate the results shown in Figure 2.

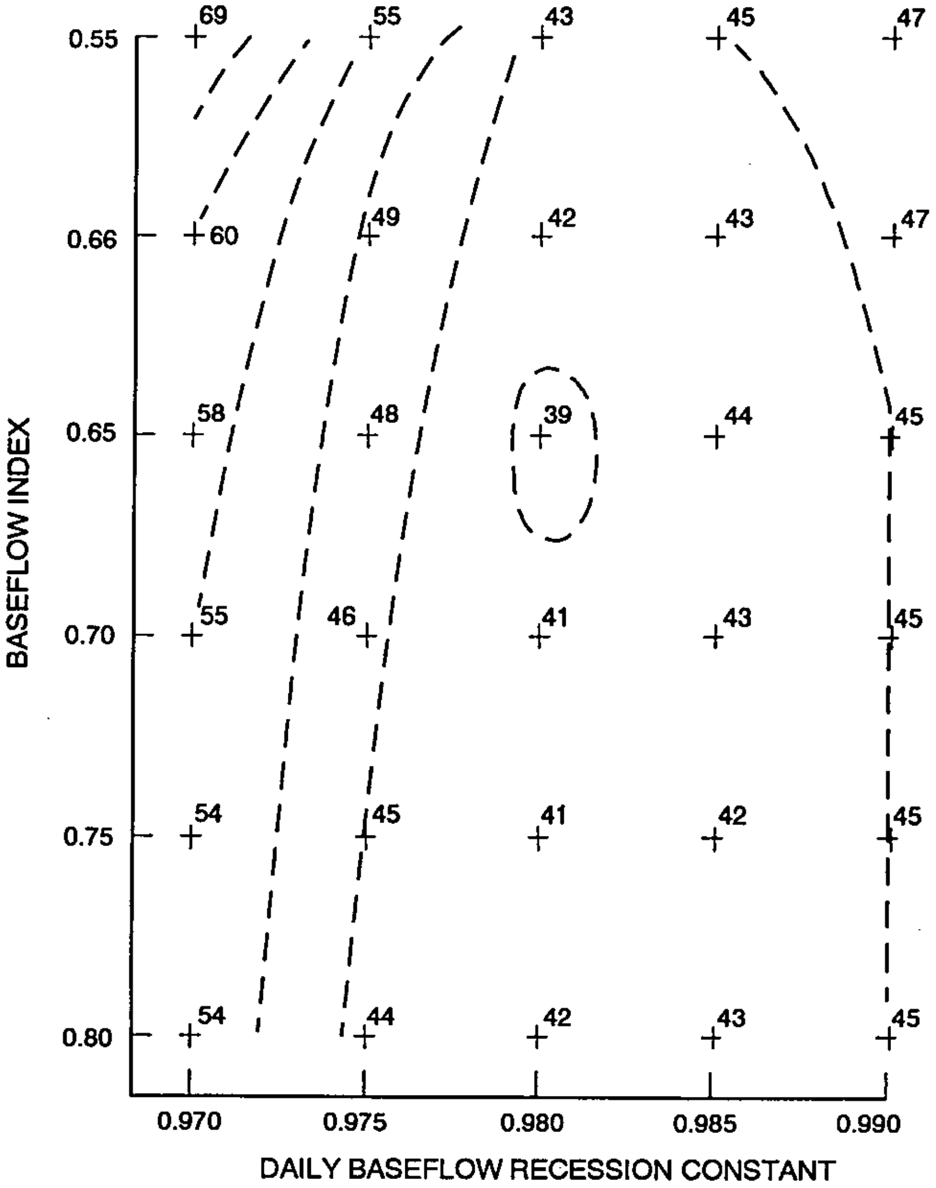

Figure 3: Babinda Creek - number of months when actual runoff exceeds calculated runoff from zero storage capacity

It would be pleasant if it could be reported that the Babinda Creek data set was an isolated and atypical example; however, this is not so. The records from the 686 sq km Davey River catchment in south west Tasmania (307001) show even clearer inconsistency between rainfall and runoff. Four rainfall stations were used to determine the input rainfall data. In the 13-year study period, total measured runoff was 26083 mm (2006 mm p.a.) while total runoff calculated by the AWBM with zero surface storage capacity was 22992 mm (1769 mm p.a.), i.e. less than the measured runoff. Annual totals of calculated runoff using zero storage were less than measured runoff in 12 of the 13 years in the study period.

Table 3 summarises the results from 5 catchments with apparent inconsistencies between rainfall and runoff.

Table 3 Comparison of measured runoff with calculated runoff from zero storage capacity (runoff values in mm p.a.)

Catchment	Area sq km	Years of Data	Actual Runoff	Calculated runoff from zero storage
Babinda Ck	39	14	4654	4618
Cochable Ck	93	13	2076	1835
Davey River	686	13	2006	1769
Endrick River	210	10	655	556
Tooma River	114	9	1388	1404

3.2 Too much rainfall, too little runoff

If data sets in which rainfall data are too low to be consistent with runoff can be readily found as illustrated above, then it is reasonable to assume that corresponding inconsistencies exist in which rainfall data are too high to be consistent with runoff.

Figure 4 shows the results of calibrating the AWBM on 12 years of data from the 133 sq km Belar Creek catchment at Warkton in New South Wales (420003). The coefficient of determination between monthly calculated and measured runoff $r^2 = 0.868$. At first sight, the calibration and reproduction of measured runoff by the AWBM seems to be satisfactory. Further study of the data and the results seem less convincing.

The calibrated values of surface storage capacities and partial areas are 70, 430 and 1550 mm over 0.22, 0.23 and 0.55 of the catchment area respectively. The average surface storage capacity of 962 mm is so high that the calibrated values are suspect. In addition, the largest

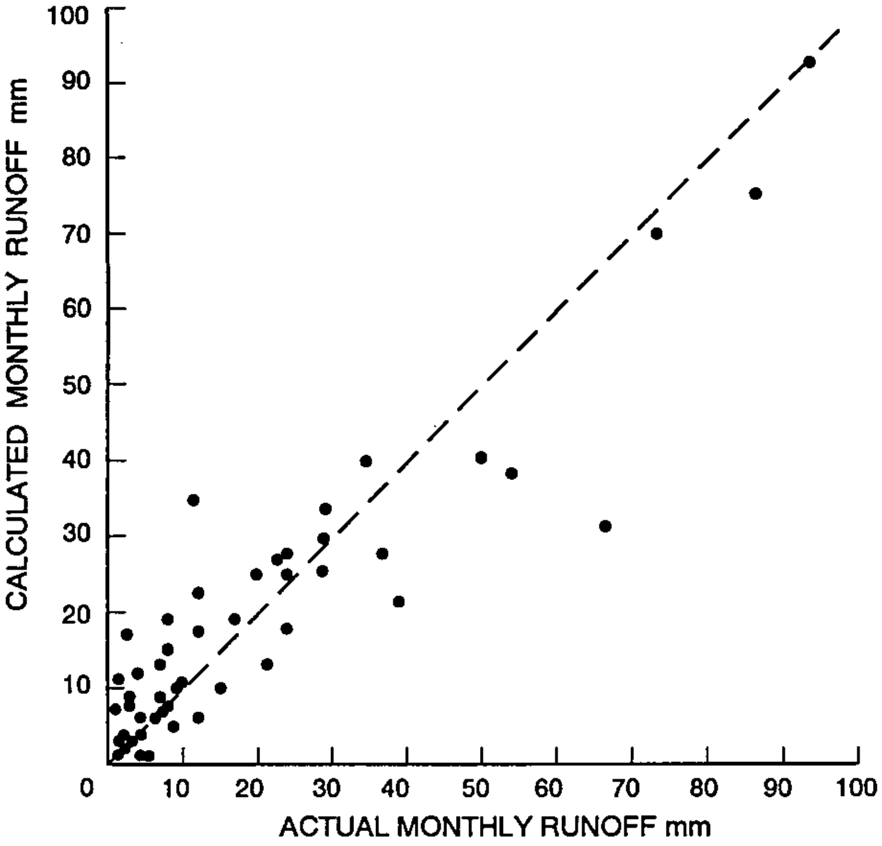

Figure 4: Belar Creek - comparison of actual runoff and calculated runoff from calibrated AWBM

store of 1550 mm capacity which represents 55% of the catchment area did not generate any runoff in the entire 12-year study period. All calculated runoff (including the baseflow runoff as well as the surface component of runoff) was generated by the two smaller storages which together represent only 45% of the area.

The two issues of concern here are whether:

(i) an average surface storage capacity of 962 mm is physically realistic ?
(ii) it is physically realistic that 55% of a catchment generates no runoff in a 12-year period?

Although the data from reliable calibrations of the AWBM are still limited, there are other sources of information about the normal range of surface storage capacity of catchments. For example, the Antecedent Precipitation Index (API) is a single storage rainfall-runoff model used by the Bureau of Meteorology for flood forecasting. In Queensland, the API model is used on about 75 catchments. The API capacities used on this large number of catchments have a relatively limited range from 100 to 150 mm (personal communication, unpublished information). The Queensland DPI Water Resources (1994) determined maximum initial loss values for the API model for 20 catchments in the Brisbane and Pine Rivers in south east Queensland. The 20 values ranged from 32 mm to 130 mm with an average of 81 mm.

An extensive set of information about the surface storage capacities of catchments was collated by Nathan and McMahon (1990) who calibrated the SFB model on 106 catchments in south east Australia. The calibrated values of the surface storage parameter S, as rearranged by Boughton (1991), are shown in Figure 5. These calibrated values of S are affected by errors in rainfall data in the same way as the corresponding average surface storage capacity in the AWBM (see Figure 1) so the spread of values in Figure 5 around the average values of about 120 mm can easily be explained by relatively small errors in the estimation of catchment rainfall. The centre of the spread of values agrees with the Bureau of Meteorology values of 100 to 150 mm in the API model. In addition, Boughton (1991) showed that the range of recommended curve numbers for the USDA Curve Number Method have a similar spread that corresponds to a similar range of surface storage capacities.

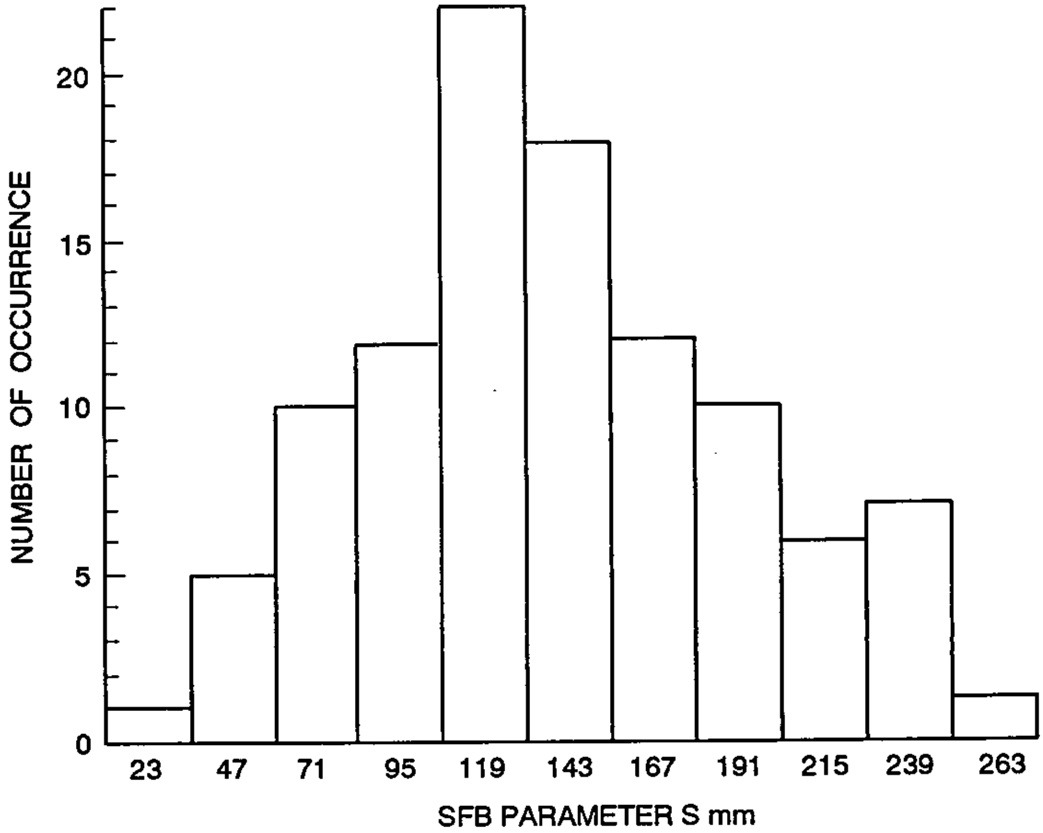

Figure 5: Calibrations of SFB storage parameter S (adapted from Nathan and McMahon, 1990)

There is a correspondence between the storage parameters of the SFB and AWBM models as shown in Table 4. Both models were calibrated on 5 sets of rainfall and runoff data. Although there is not a fixed relationship between corresponding values, it is obvious that the two storage parameters are approximately of the same magnitude.

Table 4 Comparison of calibrated storage parameters in the SFB and AWBM models (values in mm)

Catchment	Area sq km	Years of Data	SFB S par.	AWBM Ave. Cap
Allyn River	205	8	100	105
Brigalow Cl	0.16	15	100	115
Munduran Ck	55	8	110	135
Mary River	3100	4	150	140
Bass River	52	10	200	200

The available information suggests that the normal range of surface storage capacity is about 100 to 150 mm, with outer limits from zero to 100 mm, and from 150 to 250 mm (as likely to be due to data errors as to true variations in capacity). The calibrated value of 962 mm for the average surface storage capacity of Belar Creek is almost an order of magnitude higher than any other values of surface storage capacity and cannot be accepted as realistic. Soils of the Belar Creek catchment are classed as loam and the vegetation is a mixture of grassland and forest. None of the catchment characteristics is so unusual that it could account for the calibrated value.

When high runoff and/or low rainfall forces the calibrated storage capacity to zero, the lack of reality and the inconsistency between runoff and rainfall are immediately obvious. The alternative error of high rainfall and/or low runoff forces the calibrated storage capacity to a very high value, but how high is unrealistic? *In the opinion of the writer, any value above 200 mm should be treated with caution, and any value above 300 mm should be regarded as indicative of data errors.*

Is it physically realistic that 55% of a catchment contributes no runoff in a 12-year period? The phenomenon of partial area runoff is well known and can be readily identified and evaluated in rainfall-runoff data sets (Boughton, 1987b, 1990). The calibration procedures for the AWBM (Boughton, 1996) are a formalised way of routinely evaluating partial areas of runoff generation. However, there is no established body of knowledge about how frequently or infrequently different parts of a catchment generate runoff. Table 5 shows some examples of data sets where the combination of very high calibrated surface storage capacity and zero to very infrequent runoff generation from a major part of the catchment suggests that data errors are affecting the results.

Table 5 Calibrated average surface storage capacity (AWBM) and non-contributing area

Catchment	Area sq km	Years of Data	r^2	Average Cap.	Non-Contrib. Area
Belar Ck	133	12	0.868	962	55%
Stones Bk	15	11	0.926	608	58% *
Scott Ck.	27	16	0.915	502	58%
Suggan Bu.	357	14	0.678	380	44%
Cressbrook	210	5	0.913	295	60%

* The partial area of maximum capacity contributed a small amount of runoff in 2 runoff events in the 11-year period.

3.3 Errors in Individual Runoff Events

Minimising a sum of squares of differences between actual and calculated runoff, usually monthly totals of runoff, is a very common method of calibrating rainfall-runoff models. The method is popular because it simplifies the calibration criteria into a single number. The sum of squares can also indicate the possibility of data errors if the contribution of each month to the total is examined.

Table 6 shows 4 years of actual and calculated monthly runoff from the 3120 sq km Mary River catchment in south eastern Queensland. These data were used to produce Figure 1 and Table 2. Also shown are the percentages of the sum of squares of differences which are contributed by each of the 48 months. the values are rounded to the nearest percent for clarity so the 48 values do not add exactly to 100. The significant feature of the Table is that 69% of the sum of squares is contributed by the single month of April 1989. This is also the month with the largest runoff in the 4-year period.

The same month can also be identified as a potential point of data error by comparing actual runoff with calculated runoff from zero storage capacity (as used earlier in the paper). The 4-year total of actual runoff is less than the total of calculated runoff from zero storage capacity, but the actual runoff in the month of April 1989 is greater than the calculated total from zero storage capacity for that month.

Table 6 Mary River Data
Comparison of Actual and Calculated runoff - mm
Also shown is % contribution to the sum of squares

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
1988													
Act	-	-	-	-	-	-	96	13	8	2	1	102	222
Est	-	-	-	-	-	-	96	13	4	2	1	95	211
%							0	0	0	0	0	0	
1989													
Act	12	19	42	672	150	43	28	38	7	3	8	3	1025
Est	12	11	38	550	137	32	34	18	4	3	10	3	852
%	0	0	0	***69***	1	1	0	2	0	0	0	0	
1990													
Act	3	33	76	87	50	51	9	7	3	2	2	2	325
Est	6	44	83	113	77	64	11	6	2	1	4	1	412
%	0	1	0	3	3	1	0	0	0	0	0	0	
1991													
Act	2	16	3	1	1	1	1	0	0	1	1	46	73
Est	2	17	2	1	2	1	1	0	0	0	0	91	117
%	0	0	0	0	0	0	0	0	0	0	0	9	
1992													
Act	5	395	185	62	24	8	-	-	-	-	-	-	679
Est	8	355	193	83	30	9	-	-	-	-	-	-	678
%	0	7	0	2	0	0							

A more detailed examination of the catchment water balance in this period confirms the probability of data error. A period of 52 days from 9 April, when the flow on an established recession was 3.56 mm/day, to 1 June, when the flow on another recession was 3.45 mm/day, was selected to minimise errors in the estimation of baseflow storage. In this period, total rainfall was 484 mm and total runoff was 488 mm. Daily pan evaporation was 2.7 mm in April and 1.7 mm in May and June. It is highly unlikely that there was no evapotranspiration over a period of 52 days while 100% of measured rainfall became runoff.

Overall, the Mary River data set is of good quality, probably due to the need for careful selection of rainfall stations for flood forecasting purposes. The inconsistency in April 1989 is in the largest runoff event in the 4-year study period, and the circumstances of very large amounts of rainfall and runoff are most likely to produce errors in the data.

Similar inconsistencies in the water balance of individual storm events have been demonstrated by Boughton (1987a). The use of monthly contributions to the sum of squares of differences between actual and calculated runoff, and the comparison of actual runoff with runoff calculated from zero storage capacity, are two additional and quicker methods of identifying months in which measured runoff is too high to be consistent with measured rainfall.

There is a corresponding problem of single runoff event error when the rainfall is inconsistently high for the measured runoff; however, this type of problem is disguised by adjustment of model parameters during calibration, and the data error can not be readily identified by comparing with calculated runoff from zero storage capacity. Large errors of this type can be identified by using monthly percentages of the sum of squares of differences. These errors can also be seen as outliers on a graphical plot of calculated runoff against actual runoff.

3.4 Human Modifications of Catchment Water Balance

Oxley Creek is the largest (156 sq km) of the creeks that cross the urban area of the city of Brisbane. The catchment is mainly rural, but the lower reaches are in urbanised areas before the creek joins the Brisbane River. The streamgauging station at Beatty Road (143019) is used for flood forecasting by the Brisbane City Council because of flood problems in the lower urban reaches.

Table 7 shows 2 years of actual flow data compared with calculated runoff using zero surface storage capacity. Actual runoff exceeds the calculated runoff in a high flow period in April 1990, and in a low flow period in August and September 1991. Both of these inconsistencies are due to human influences on the catchment water balance.

Table 7 Oxley Creek Catchment
Comparison of actual runoff with calculated runoff
from zero storage capacity - mm

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1990												
Act	20	90	87	***213***	108	33	8	8	6	6	6	6
Cal	78	140	113	173	159	49	24	13	17	17	28	25
1991												
Act	7	38	7	6	8	8	8	***12***	***15***	13	11	122
Cal	60	114	22	8	34	29	28	7	7	20	31	268

The April 1990 flood event was the largest peak rate of runoff in the 2 years of data. The reason for the inconsistency was that a dam 4 metres high, constructed by sand and gravel miners on an anabranch not far upstream of the gauging station, breached during the rising stage of the flood (Boughton and Carroll, 1993). It was the additional water from the breached dam that makes it impossible to model the flood event by normal catchment water balance methods. Using the method of comparing actual runoff with calculated runoff from zero storage capacity, the inconsistency is identifiable as due to data error.

The inconsistency in the low flow period of August-September 1991 is also easy to identify by this method but is due to an entirely different reason. The period July-September 1991 was a severe drought in south-east Queensland, yet the baseflow in Oxley Creek increased during the drought. There are urban areas of the suburbs of Brisbane immediately upstream of the gauging station. Excess watering of home gardens from the piped city water supply during dry periods results in raised shallow groundwaters and artificially enhanced baseflows in the urban creeks of Brisbane (Boughton, 1986), and results in this inconsistency in the Oxley Creek data.

Although these inconsistencies can be identified by comparing actual runoff with calculated runoff from zero storage capacity, there is a need for local knowledge to determine the causes which can then confirm that it is data errors that cause the inconsistencies.

4. PROCEDURES FOR DETECTING ERRORS IN RAINFALL-RUNOFF DATA SETS

The methods used in Section 3 to identify various types of data errors can be codified into a systematic procedure for finding such errors in a new data set. The procedure is based on four checks, carried out in the following order :

(i) check actual monthly totals of runoff against calculated runoff from the AWBM using zero surface storage capacity;
(ii) calibrate the AWBM and check the magnitude of the calibrated average surface storage capacity;
(iii) check the amounts and frequency of runoff from the largest storage capacity;
(iv) calculate a sum of squares of differences between actual and calculated monthly totals of runoff, and check the percentage of the sum that is contributed by each month.

Details of each check follow in the sections below.

4.1 Compare actual runoff with calculated runoff from an impervious catchment

The purpose of this check is to see if rainfall data are unrealistically low when compared with runoff data in the whole period of record, or to see if there are inconsistencies in one or a few months.

If actual runoff exceeds the calculated runoff in more than one season of runoff, attention should be given to the possibility that the rainfall data are too low to be representative of catchment rainfall (likely) or that the runoff data are higher than true runoff (less likely) or both (a possibility) - see Section 3.1.

If actual runoff exceeds the calculated runoff in just a few months in a single season, attention should be given to the possibility that the rainfall data are too low or runoff data too high in just a single storm event. There can be carryover effects by baseflow into the following month or two when an error of this type occurs in a big storm. Even if the error occurs in a low flow period, e.g. see Table 7, further checks should be made to establish how the actual runoff can exceed the runoff that would have occurred from an impervious catchment.

The procedures for comparing actual runoff against calculated runoff from an impervious (zero storage capacity) catchment have been built into the CHECKCAL program (Boughton, 1996). This enables the check to be made in about one minute of time.

It is possible that the AWBM will not calibrate properly if the rainfall data are too low to be consistent with the runoff data. For this reason, the check of actual runoff against runoff from an impervious catchment should be made before attempting to calibrate the AWBM. The other 3 checks depend on the calibration of the model.

4.2 Check the magnitude of the calibrated AWBM average surface storage capacity

Table 8 summarises the rainfall-runoff data sets which have been used to illustrate various types of data errors in this paper. The average surface storage capacities of the 16 catchments range from zero to 962 mm. Suggestions are made in the comments in the Table as to the normal range of capacity and those values that are likely to be due to data error.

The commonest characteristic of a catchment that is likely to result in a low value of surface storage capacity is urbanisation, with impervious cover by roads, roofs, car parks, etc., over a significant part of the catchment. Such characteristics are obvious and difficult to overlook. The only natural, non-urban circumstances which can lead to a low calibrated storage value is a lack of vegetation cover combined with surface sealing of the soil profile, usually in arid and the drier semi-arid areas. For non-urbanised catchments that are not in very dry areas, a very low value of calibrated AWBM average surface storage capacity is most likely due to data error.

Table 8 Calibrated values of AWBM average surface storage capacity - mm

Station No.	Name	Capacity	Comment on value
420003	Belar	962	rainfall high/runoff low
612005	Stones	608	rainfall high/runoff low
503502	Scott	502	rainfall high/runoff low
222212	Suggan	380	rainfall high/runoff low
--	Cressbr.	295	rainfall high/runoff low
227219	Bass	200	? possible rainfall high
143019	Oxley	160	enhanced baseflow
138007	Mary	140	O.K.
132004	Munduran	135	O.K.
--	Brigalow	115	O.K.
210022	Allyn	105	O.K.
111105	Babinda	zero	rainfall low/runoff high
113004	Cochable	zero	rainfall low/runoff high
307001	Davey	zero	rainfall low/runoff high
215009	Endrick	zero	rainfall low/runoff high
401554	Tooma	zero	rainfall low/runoff high

Calibrated storage values above 200 mm should be treated as potential data errors. Any calibrated value above 300 mm should be treated as certainly due to data error.

4.3 Check runoff from the largest storage capacity

The three partial areas in the AWBM have different storage capacities. The smallest capacity always generates runoff before or at the same time as the other surface stores. The largest capacity always generates runoff after or at the same time as the other stores. When rainfall data are erroneously high compared with runoff data, the model adjusts by using the largest capacity to not produce runoff, effectively reducing the rainfall by the fraction of the catchment that is the partial area of the largest capacity.

When the largest capacity produces very little runoff in the period used for calibration, and particularly if this is in association with a high values of average surface storage capacity, it is an indication that rainfall data are too high to be consistent with the runoff data.

The AWBM calibration program BASE (or BASE5 or BASENL) has an option in the results menu for displaying the amounts of runoff generated from each of the 3 stores, and the number of days in the study period that each store generated runoff. This option should be used to check each new set of rainfall-runoff data in conjunction with the check on the magnitude of the average surface storage capacity.

4.4 Monthly contributions to sum of squares

The BASE program mentioned above has an option in the results menu for displaying the percentage that is contributed by each month to the sum of squares of differences between actual and calculated monthly totals of runoff. The option also includes a list of highest dozen monthly percentages, ranked in order of magnitude. This option quickly identifies outliers that are having a large effect on the calibration. The program also allows for the change of a value of actual runoff to a new value. By changing the value of actual runoff to the calculated value for that month, the effect of an outlier on the calibration can be eliminated. This allows for the detection of an isolated error in which rainfall is erroneously high for a short period, i.e. the calculated value of runoff is higher than the actual value. This type of error cannot be detected by the use of the impervious catchment approach as in Section 4(i) above.

5. CONCLUSIONS

The procedures set out in this paper form a systematic approach to detecting many of the errors which currently are common in rainfall-runoff data sets. There are other errors which the procedures will not detect, but the identification of those errors illustrated in the paper will make it easier to develop further procedures for eliminating more of the errors.

It was pointed out by reviewers of this report that an impression could be given to a reader that if the model does not produce a good fit between rainfall and runoff then the data must be in error. This is explicitly refuted in order to avoid misunder-standing. There are some catchments for which the AWBM will not simulate the dominant runoff processes. The purposes of the report are to demonstrate that major inconsistencies between rainfall and runoff are common in data sets which are being used for research as well as practical applications, and that it is possible to identify many of the errors using the methods described herein.

The purpose of rainfall-runoff modelling should be kept in mind during the analysis of data errors. If the modelling is for the practical purpose of using long rainfall records to extend a short record of streamflow, then it might be sensible to use scaling factors on rainfall and/or runoff to compensate for the effects of inconsistencies in the data. On the other hand, scaling of data would be worse than useless in a study where the calibrated parameter values are to be transfered for use on another, perhaps ungauged, catchment with different rainfall data.

The procedures described in this paper have been based on use of the AWBM model. All of the procedures have already been coded into the suite of programs used for calibrating and operating the AWBM (Boughton, 1996), so it is easy to make use of those facilities. However, it is emphasised that similar approaches can be used with other rainfall-runoff models, particularly those of simple structure that avoid the problems of interactions inherent in complex multi-parameter models.

6. ACKNOWLEDGMENTS

I am grateful to those who made data available for use with the AWBM which has contributed to the preparation of this paper. In particular, I thank Dr Chiew and Professor McMahon at the University of Melbourne who made available the 28 data sets that they collated from all States of Australia. Through them, I thank the State water authorities who prepared the data sets in the first instance. I also thank Brisbane City Council and the Bureau of Meteorology, Queensland Department of Primary Industries, Dr John Macintosh and Forood Sharifi who all helped by making data available.

7. REFERENCES

Boughton, W. C., 1985 : "Brigalow catchment study, calibration phase, 1965-1979", Qld. Dept. of Primary Industries, Div. Land Utilisation, Project Rept. 085012, 57 pp.

Boughton, W. C., 1986 : "Enhanced baseflows in an urban area", Working Paper 2/86, School of Aust. Envir. Studies, Griffith University, Brisbane. 14 pp.

Boughton, W. C., 1987a : "Hydrograph analysis as a basis for rainfall-runoff modelling", Inst. Engs. Aust., Civil Engg. Trans, v.CE29, no.1, pp.28-33.

Boughton, W. C., 1987b : "Evaluating the partial areas of watershed runoff", Amer. Soc. Civil Engs., J. Irr. & Drain., v.113, no.3, pp.356-366.

Boughton, W. C., 1990 : "Systematic procedure for evaluating partial areas of watershed runoff", Amer. Soc. Civil Engs., J. Irr. & Drain., v.116, no.1, pp.83-89.

Boughton, W. C., 1991 : Discussion of paper "The SFB Model Part II - Operational Considerations" by Nathan & McMahon, Inst. Engs. Aust., Civil Engg. Trans., v.CE23, no.2, pp.213-215.

Boughton, W. C., 1993 : "A hydrograph based model for estimating the water yield of ungauged catchments", Inst. Engs. Aust., Nat. Conf. Publ. 93/14, pp.317-324.

Boughton, W. C., 1996 : "AWBM water balance model calibration and operation manual, version 2.0", unpublished rept., 36 pp.

Boughton, W. C. and Carroll, D. G., 1993 : "A simple combined water balance/flood hydrograph model", Inst. Engs. Aust., Nat. Conf. Publ. 93/14, pp.299-304.

Chiew, F. H. and McMahon, T. A., 1993 : "Complete set of daily rainfall, potential evapotranspiration and streamflow data for twenty eight unregulated Australian catchments", Centre for Envir. Appl. Hydrology, Univ. of Melbourne, 53 pp.

Macintosh, J. C., 1994 : "Application of the AWBM, a case study", Inst. Engs. Aust., Qld. Div., Session Notes for AWBM Workshop, April 1994.

Nathan, R. J. and McMahon, T. A., 1990 : "The SFB model Part II - operational considerations", Inst. Engs. Aust., Civil Engg. Trans., v.CE32, no.3, pp.162-166.

Queensland Dept. of Primary Industries, Water Resources, 1994 : "Brisbane River and Pine River flood studies. Final draft report on regional loss model relationships", unpubl. report.

Sharifi, F., 1996 : "Catchment rainfall-runoff computer modelling", PhD thesis, Univ. of Wollongong.

APPENDIX A Obtaining copies of the AWBM.

The AWBM and accompanying programs for calibration and error detection and a copy of the operating manual are available without charge from the World Wide Web home page of the CRC for Catchment Hydrology. The address of the CRC home page is :

http://civil-www.eng.monash.edu.au/crcch/home.htm

The AWBM set of programs, sample data files and manual are available as pkzipped archived files. To download the pkzipped file, click the mouse on one of the filenames and then choose the option to save the file to your hard disk.

AWBM.ZIP	(368,009 bytes) -	all AWBM files
AWBMCODE.ZIP	(217,923 bytes) -	AWBM programs only
AWBMSAMP.ZIP	(89,918 bytes) -	sample data files only
AWBM_MAN.ZIP	(53,187 bytes) -	AWBM user manual

After saving the file to your hard disk, you will need a copy of pkunzip to unarchive the files by typing :

```
pkunzip  filename.zip
```

The Cooperative Research Centre for Catchment Hydrology is a cooperative venture formed under the Commonwealth CRC Program between:

Bureau of Meteorology

CSIRO Division of Water Resources

Department of Conservation and Natural Resources

Melbourne Water

Monash University

Murray-Darling Basin Commission

Rural Water Corporation

The University of Melbourne

COOPERATIVE RESEARCH CENTRE FOR
CATCHMENT HYDROLOGY

Centre Office
Department of Civil Engineering, Monash University, Clayton, Victoria 3168, Australia
Telephone: (03) 9905 2704 Facsimile: (03) 9905 5033 International: +61 3 9905 2704