WATER RESOURCES OF SOUTH AFRICA, 2012 STUDY (WR2012)

Volume 4: A review of the accuracy of calibrations undertaken within the WRSM/Pitman Model

WV Pitman

WATER RESEARCH COMMISSION

TT 686/16

WATER RESOURCES OF SOUTH AFRICA, 2012 STUDY (WR2012)

A Review of the Accuracy of Calibrations undertaken with the WRSM/Pitman Model

Report to the
Water Research Commission

by

WV Pitman
Royal HaskoningDHV (Pty) Ltd

WRC Report No. TT 686/16

August 2016

Obtainable from

Water Research Commission
Private Bag X03
GEZINA, 0031

orders@wrc.org.za or download from www.wrc.org.za

The publication of this report emanates from a project entitled Water Resources of South Africa, 2012 (WR2012) (WRC Project No. K5/2143/1).

This report forms part of a series of nine reports. The reports are:

1. WR2012 Executive Summary (WRC Report No. TT 683/16)
2. WR2012 User Guide (WRC Report No. TT 684/16)
3. WR2012 Book of Maps (WRC Report No. TT 685/16)
4. **WR2012 Calibration Accuracy (WRC Report No TT 686/16 – this report)**
5. WR2012 SAMI Groundwater module: Verification Studies, Default Parameters and Calibration Guide (WRC Report No. TT 687/16)
6. WR2012 SALMOD: Salinity Modelling of the Upper Vaal, Middle Vaal and Lower Vaal sub-Water Management Areas (new Vaal Water Management Area) (WRC Report No. TT 688/16)
7. WRSM/Pitman User Manual (WRC Report No. TT 689/16)
8. WRSM/Pitman Theory Manual (WRC Report No. TT 690/16)
9. WRSM/Pitman Programmer's Code Manual WRC Report No. TT 691/16)

ISBN 978-1-4312-0844-9
Printed in the Republic of South Africa

ACKNOWLEDGEMENTS

The authors would like to acknowledge:

The Water Research Commission for their commissioning and funding of this entire project.

The Department of Water and Sanitation for their rainfall, streamflow, Reservoir Record and water quality data, some GIS maps and their participation on the Reference Group.

The South African Weather Services (SAWS) for their rainfall data.

The following firms and their staff who provided major input:

- Royal HaskoningDHV (Pty) Ltd: Mr Allan Bailey, Dr Marieke de Groen, Miss Kerry Grimmer (now WSP Group), Mr Sipho Dingiso, Miss Saieshni Thantony, Miss Sarah Collinge, Mr Niell du Plooy and consultant Dr Bill Pitman (all aspects of the study);
- SRK Consulting (SA) (Pty) Ltd: Ms Ansu Louw, Miss Joyce Mathole and Ms Janet Fowler (Land use and GIS maps);
- Umfula Wempilo Consulting cc: Dr Chris Herold (water quality);
- Alborak: Mr Grant Nyland (model development);
- GTIS: Mr Töbias Goebel (website) and
- WSM:; Mr Karim Sami (groundwater).

The following persons who provided input into the coding of the WRSM/Pitman model:

- Dr Bill Pitman;
- Mr Allan Bailey;
- Mr Grant Nyland;
- Mrs Riana Steyn and
- Mr Pieter Van Rooyen.

Other involvement as follows:

Many other organizations and individuals provided information and assistance and the contributions were of tremendous value.

REFERENCE GROUP

Reference Group Members	
Mr Wandile Nomquphu (Chairman)	Water Research Commission
Mrs Isa Thompson	Department of Water and Sanitation
Mr Elias Nel	Department of Water and Sanitation (now retired)
Mr Fanus Fourie	Department of Water and Sanitation
Mr Herman Keuris	Department of Water and Sanitation
Dr Nadene Slabbert	Department of Water and Sanitation
Miss Nana Mthethwa	Department of Water and Sanitation
Mr Kwazi Majola	Department of Water and Sanitation
Dr Chris Moseki	Department of Water and Sanitation
Professor Denis Hughes	Rhodes University
Professor Andre Görgens	Aurecon
Mr Anton Sparks	Aurecon
Mr Bennie Haasbroek	Hydrosol
Mr Anton Sparks	Aurecon
Mr Gerald de Jager	AECOM
Mr Stephen Mallory	Water for Africa
Mr Pieter van Rooyen	WRP
Mr Brian Jackson	Inkomati CMA
Dr Evison Kapangaziwiri	CSIR
Dr Jean-Marc Mwenge Kahinda	CSIR
Research Team members	
Mr Allan Bailey (Project Leader)	Royal HaskoningDHV
Dr Marieke de Groen	Royal HaskoningDHV
Dr Bill Pitman	Consultant to Royal HaskoningDHV
Dr Chris Herold	Umfula Wempilo
Mr Karim Sami	WSMLeshika
Ms Ans Louw	SRK
Ms Janet Fowler	SRK
Mr Niell du Plooy	Royal HaskoningDHV

Mr Töbias Gobel	GTIS
Miss Saieshni Thantony	Royal HaskoningDHV
Mr Grant Nyland	Alborak
Miss Sarah Collinge	Royal HaskoningDHV
Miss Kerry Grimmer	WSP
Ms Riana Steyn	Consultant
Miss Joyce Mathole	SRK
Mr Sipho Dingiso	Ex Royal HaskoningDHV

EXECUTIVE SUMMARY

This WRSM/Pitman model produces simulated streamflow which is compared to observed streamflow by means of graphs and statistics. The model is then calibrated to compare as closely as possible to observed streamflow. This report deals with the process of calibrating the approximately 600 streamflow gauging stations in South Africa, Lesotho and Swaziland. The approximately 600 gauges were categorized into six categories based on experience gained in calibrating at these stations and their reliability. Typical problems encountered during calibration were identified. Criteria for the acceptance of statistical parameters were dealt with. The calibration accuracy was analysed graphically. The statistical parameters were described and their relationship to the six categories were analysed and discussed. Modelling inaccuracy versus gauging errors was discussed.

Table of Contents

List of Tables

List of Figures

LIST OF ABBREVIATIONS

DWS	Department of Water and Sanitation
MAP	Mean Annual Precipitation
MAR	Mean Annual Runoff
WMA	Water Management Area
WRSM/Pitman	Water Resources Simulation Model 2000, also referred to as the Pitman model
WR2005	Water Resources of South Africa 2005 study
WR2012	Water Resources of South Africa 2012 study

1 INTRODUCTION

The previous WR2005 project incorporated the calibration of WRSM/Pitman calibrations on almost 600 streamflow records. For the WR2012 project it was necessary to review and possibly update the WR2005 calibrations for the following reasons:

- an additional 5 years of data has been added to the WR2005 data set;
- a number of new gauges have been added;
- in many catchments the land-use data has been improved;
- the Sami groundwater parameters have been revised for the whole study area; and
- in some catchments the default Sami parameters (i.e. those adjusted as part of the calibration process) were left unaltered in WR2005.

The revised calibrations were used to assess model performance for a wide variety of catchments, ranging in size from 15 km^2 to over 400 000 km^2, with unit runoff ranging from about 1 mm to over 1 000 mm. The catchments also reflect a wide variance in land use, from almost pristine mountain catchments to those embracing major dams, large irrigation schemes, afforestation or urbanization.

2 THE PROCESS OF CALIBRATION

Model calibration can be defined as the adjustment of model parameters until a satisfactory fit has been obtained between observed and simulated flows. Calibration should not be viewed as an exercise in isolation: it is also an important part of the process of verifying hydrological data, including that of land and water use. Before calibration is initiated the data should be thoroughly screened: however, some of the data problems may only become evident during the process of calibration. To this end the various graphs should be studied before comparing the statistics of the observed and simulated flows. For example, the yearly hydrograph can be used to identify outliers, reflected by large differences between observed and simulated flows. Some of these outliers could be due to unpatched flow data where the simulated flow is much higher than observed: however, in this study a concerted effort was made to patch all such flows.

All the graphs are used in the calibration process, in conjunction with the statistics, in order to arrive at the “best” calibration. In this analysis of calibration accuracy, however, it was necessary to focus on the strictly numerical statistics. These statistics are displayed in the calibration screen and are as follows:

- MAR – mean annual runoff;
- Mean (log) – mean of logs of annual flows;
- Std. dev. – standard deviation of annual flows;
- Std. dev. (log) – standard deviation of logs of annual flows; and
- Seasonal index – mean variability of flow over a year.

Statistical measures can be a bit misleading looked at in isolation as positive and negative differences in observed and simulated flows can tend to cancel each other out. This pattern is easily noticed in the graphs, however, which is a very good reason to analyze both graphs and statistics. It is, however, very unlikely that poor graphs would reflect good statistics for all the statistical measures though.

A discussion of the role the various statistics have in the assessment of model calibration follows.

2.1 Mean Annual Runoff (MAR)

This is probably the most important statistic as it is the MAR that plays the dominant role in the determination of yield from a reservoir. Problems with matching MAR will obviously be reflected by an unsatisfactorily high error in MAR. Outliers will also have an impact (depending on the length of record). If the outlier (or outliers) results from the simulated flow being much greater than observed, one will usually end up with a simulated MAR that is higher than observed, although the difference will normally not be too significant, unless the record is short; the opposite applies when the observed flow is much greater than simulated.

2.2 Mean Annual Log Runoff

This statistic has the effect of minimizing the impact of outliers, but has the disadvantage of being very sensitive to annual flows at the low end of the spectrum, especially where they are very small relative to the mean. Furthermore, as the log of zero is $-\infty$, annual flows are assigned a nominal (very small) value in such cases. As it is possible in small catchments to have a log mean of zero, or very close to zero, small differences can result in very large percentage errors. In an attempt to circumvent this problem log means were increased by 2 (equivalent to multiplying all values by 100) to ensure that all log means were positive.

2.3 Standard Deviation

This statistic is the classic indicator of the variation about the mean. If the simulated MAR is too high it is more than likely that the standard deviation will also be overestimated, and vice versa. Outliers can have a significant impact (depending on the length of record). If the outlier (or outliers) results from the simulated flow being much greater than observed, one will almost definitely end up with a simulated standard deviation that is higher than observed, and the difference can be quite significant; the opposite applies when the observed flow is much greater than simulated. The standard deviation will show up a poor correlation where the simulated flow oscillates above and below the observed flow which may balance out in the MAR and look quite good.

2.4 Log Standard Deviation

This statistic reduces the impact of outliers but is very sensitive to annual flows at the low end of the spectrum. As mentioned above, there is a problem when zero annual flows are encountered, when the assignment of an arbitrary small value has a large impact on the log standard deviation. It goes without saying that there can be a large discrepancy between observed and simulated log standard deviation when one or more of the observed or simulated flows are equal to zero.

In the original verification of the WRSM/Pitman model the log-normal distribution was selected for statistical comparisons between observed and simulated annual flows as the logs of such flows tend to be normally distributed. These statistics, namely mean and standard deviation have been retained in the WRSM/Pitman table of statistics but are probably of lesser importance than the un-transformed statistics.

2.5 Seasonal Index

It is more productive to study the graph of mean monthly flows than to merely compare this statistic: however, unlike the previous four statistics, which are based on annual flows, it does attempt to take into account the ability to model the variation in monthly flow over the year. If the simulated MAR is too high at least some – if not all – of the mean monthly flows will exceed the observed flows: however, this may not reflect in the seasonal index. Outliers will have a moderate impact (depending on the length of record). If the outlier (or outliers) results from the simulated flow being much greater than observed, one will usually end up with a simulated seasonal index that is higher than observed, although the difference will normally not be too significant; the opposite applies when the observed flow is much greater than simulated. It was noted that large percentage errors were sometimes encountered in the year-round rainfall zone where the seasonal index is generally small. Accordingly it was considered more reasonable to express the error as the straight difference between the seasonal indices, as they are already expressed as percentages.

3 TYPICAL PROBLEMS ENCOUNTERED DURING CALIBRATION

Although most calibrations are straightforward, there are many instances where it is difficult to arrive at a reasonable calibration. It is taken for granted in this analysis that all rainfall data has been checked and all observed streamflow records have been patched, hence there is no reference to problems caused by missing data. An analysis of the approximately 600 calibrations performed for WR2012 revealed that calibrations could be classified into 6 categories, as follows:

- (1) No apparent problems;
- (2) Problems caused by outliers;
- (3) Problems caused by imbalance among records on same river (or in same catchment);
- (4) Problems with log statistics caused by zero or near- zero annual flows;
- (5) Problems due to very short records; and
- (6) Data problems rendering calibration impossible.

A discussion of these categories follows.

3.1 No apparent problems

This category is more or less self-explanatory. In the context of this analysis it implies there were no problems related to outliers, imbalance among gauges, zero annual flows or any obvious data errors. However, there was still a fairly wide variation in calibration accuracy within this category.

3.2 Problems caused by outliers

As mentioned above, the presence of outliers can best be detected by plotting the annual hydrograph. Outliers in this context are large differences between observed and simulated flows. The treatment of outliers is a complex subject in which common sense can be more important than statistical theory. Outliers can have a significant impact on the MAR and standard deviation. For example, where the simulated flow is very much higher than the observed flow, the simulated MAR will be greater and, the shorter the record, the greater the impact. The same applies to the standard deviation, which is even more sensitive to outliers.

Outliers are usually associated with extreme events, when measurement of both rainfall and streamflow can be problematical. As far as is possible, outliers due to such data problems have been eliminated as part of the calibration process. A review of the calibrations revealed about 100 outliers, split fairly evenly between those with simulated flows greater than observed and those with observed flows greater than simulated. This result indicates no bias with regard to outliers, but a trend of increasing frequency in the occurrence of outliers emerged as shown in Table 3.1. The drop in the 1980s is probably due to the fact that this decade was considerably drier than the 1970s.

Table 3.1: Frequency of outliers

Period	< 1960	1960s	1970s	1980s	1990s	2000s
No. outliers	3	6	20	14	25	30

The high number of outliers in the 1990s is mainly attributed to the extreme wet years in 1995 and 1999, but the higher number since 2000 could be due to the decline in rain gauge coverage and (as yet) undetected errors in the more recently processed streamflow data.

3.3 Problems caused by imbalances among records on same river

Even very good streamflow records are not 100% accurate – some may overestimate by a few percent and others may underestimate. When a number of gauges are situated on the same river (or in the same catchment), it is unlikely that it will be possible to “home in” on the observed MAR at each gauge, without resorting to highly unrealistic model parameters. In such cases it is necessary to over- or under-estimate flows at one or more of the gauges. Note that this category does not include records that fall into category 6, where the error in MAR can be considerable.

Also placed in this category are records split into two (or more) segments. Some records may have large gaps running for several years. Such gaps may be the result of damage caused by a major flood and a consequent delay in reinstating the gauge. It is not advisable to patch such long gaps and it is preferable to split the record into segments. A problem often arises with split records where one cannot closely match the MAR for both segments. In this case the focus has been on the longer segment or, if both are of similar length, the more recent segment.

3.4 Problems with log statistics

As mentioned under the descriptions of the statistics mean (log) and standard deviation (log), the presence of zero (or near zero) annual flows can distort these statistics to a great degree. Calibrations in this category, however, often tend to be satisfactory in terms of the other 3 statistics and the graphs generally do not reflect this problem.

3.5 Problems due to short records

It is generally recognized that a record should be at least 10 years in length for a reliable calibration to be undertaken. However, some such records were included in WR2012 where flow data were sparse: the thinking here was to have them in the system so that, if and when future studies are done, they will eventually become useful for model calibration.

3.6 Data problems rendering calibration impossible

There can be several reasons why a gauge is placed in this category, as discussed below.

3.6.1 Catchment MAP over- or under-estimated

This can be the situation in small mountain catchments where rainfall gauges are sparse and interpolation techniques are inadequate. This is most likely to occur in mountainous area where rainfall gradients are steep, especially in the Western Cape. If MAP has been underestimated to a significant degree it will not be possible for the model to generate sufficient runoff whereas, if MAP has been overestimated, the model will require parameters outside the normal range to suppress runoff sufficiently.

3.6.2 Land-use impacts over or under-estimated

Data on land-use is not always reliable or even available necessitating extension of data from previous years: if it is way out it will usually not be possible to obtain a reasonable calibration.

3.6.3 Data errors in the streamflow record

Plotting of the data can highlight obvious errors, where flows are unreasonably high or low. In some cases it is only a part of the record that is affected; hence this part has been excluded in the calibration process. In some cases gauges in this category also had large gaps but, as the records were too unreliable for calibration, no attempt was made to patch them.

4 CRITERIA FOR ASSESSMENT OF CALIBRATIONS

As mentioned above, the process of calibration relies to a large extent on the evaluation of certain graphs that depict the relationship between observed and simulated flows. However, in this assessment it was necessary to confine the analysis to numerical indices of goodness of fit. Five indices were selected as follows:

- percentage error in mean annual runoff (MAR);
- percentage error in mean annual log runoff;
- percentage error in standard deviation;
- percentage error in standard deviation of logs; and
- percentage error in seasonal index.

In order to detect bias in the calibrations, the actual errors (i.e. with + or – sign) were determined in addition to the absolute errors. As a measure of the overall accuracy of calibration the mean absolute error of all 5 statistics was calculated.

A spreadsheet was compiled with the observed and simulated statistics for each gauge so that the percentage errors could be calculated. After studying the results of the calibrations – and this included inspection of the various graphs – they were classified according to the six categories mentioned above. In cases where more than one category could apply, the “most severe” category was selected. For example, if a gauge could be classified as categories 3 and 6, the latter applied. Table 4.1 lists the number of gauges in each category and the percentage of the full data set. It should be noted that the list includes record segments where they have been split due to long gaps of missing data, thus the actual number of gauges (593) is less than the total given in the table.

Table 4.1: Number of gauges per category

Category	No. of gauges	Percent total
1	359	58
2	94	15
3	64	11
4	51	8
5	10	2
6	34	6
Total	612	100

5 ANALYSIS OF CALIBRATION ACCURACY

The analysis concentrated on the determination of the 1st, 2nd (i.e. median) and 3rd quartiles of the percentage errors derived from the calibrations. The median was preferred to the mean to avoid undue influence by extreme errors. The two quartiles were selected to indicate the range for half of the calibrations closest to the median: this means that a quarter of the results are better and a quarter of them are worse than shown by this range.

5.1 Bias in calibrations

The first test undertaken was for the detection of any bias in the calibrations, i.e. to check if any of the simulated statistics showed a definite difference to the observed statistics. For this analysis all 6 categories were combined into a single data set. For each statistic the 1st quartile, median and 3rd quartile was determined, as shown in Figure 5.1.

This graph shows that the median error is very close to zero for MAR, Mean (log) and Seasonal Index, with a range of a few percent between the 1st and 3rd quartiles. The median error in the Standard Deviation is only about 2 to 3 percent below the observed, although there is quite a wide range between the 1st and 3rd quartiles. This wide range is due to the problem of outliers, which can have a significant impact on Standard Deviation. The only statistic with a significant bias is the Log Standard Deviation, with a median error of nearly 10% below observed. This result is largely due to the problems with zero (or near-zero) annual flows, which are the reason for category 4.

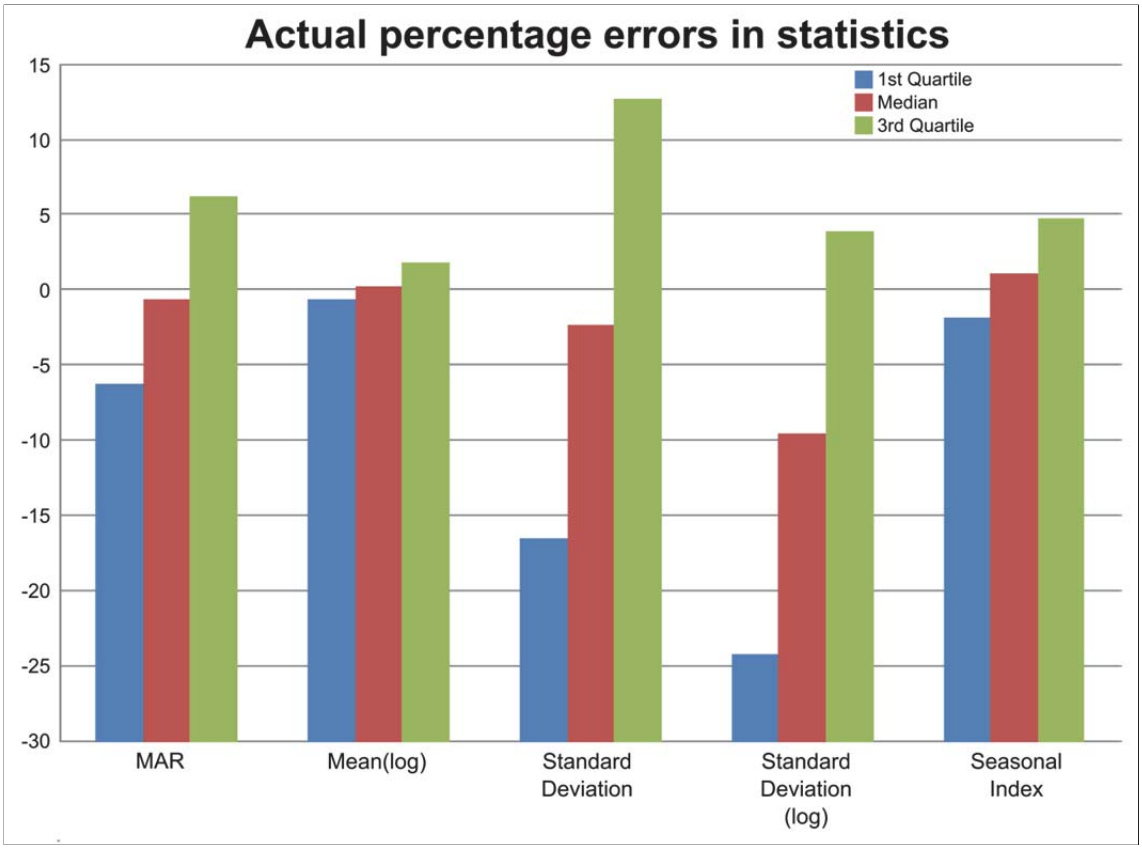

Figure 5.1: Analysis of actual percentage errors in each statistic

5.2 Accuracy of calibrations

In analyzing the calibration results (in terms of the absolute errors), the following groups were selected to illustrate the deterioration with the inclusion of each category.

- category 1;
- categories 1 and 2;
- categories 1, 2 and 3;
- categories 1, 2, 3 and 4 ;and
- all categories (i.e. 1 to 6)

Categories 5 and 6 were included together as category 5 represents only about 1% of the full data set. Each statistic is dealt with in turn, including the mean error of all 5 statistics.

5.2.1 MAR (mean annual runoff)

The 1st quartile, median and 3rd quartile errors are plotted in Figure 5.2 for each of the 5 category groups shown above.

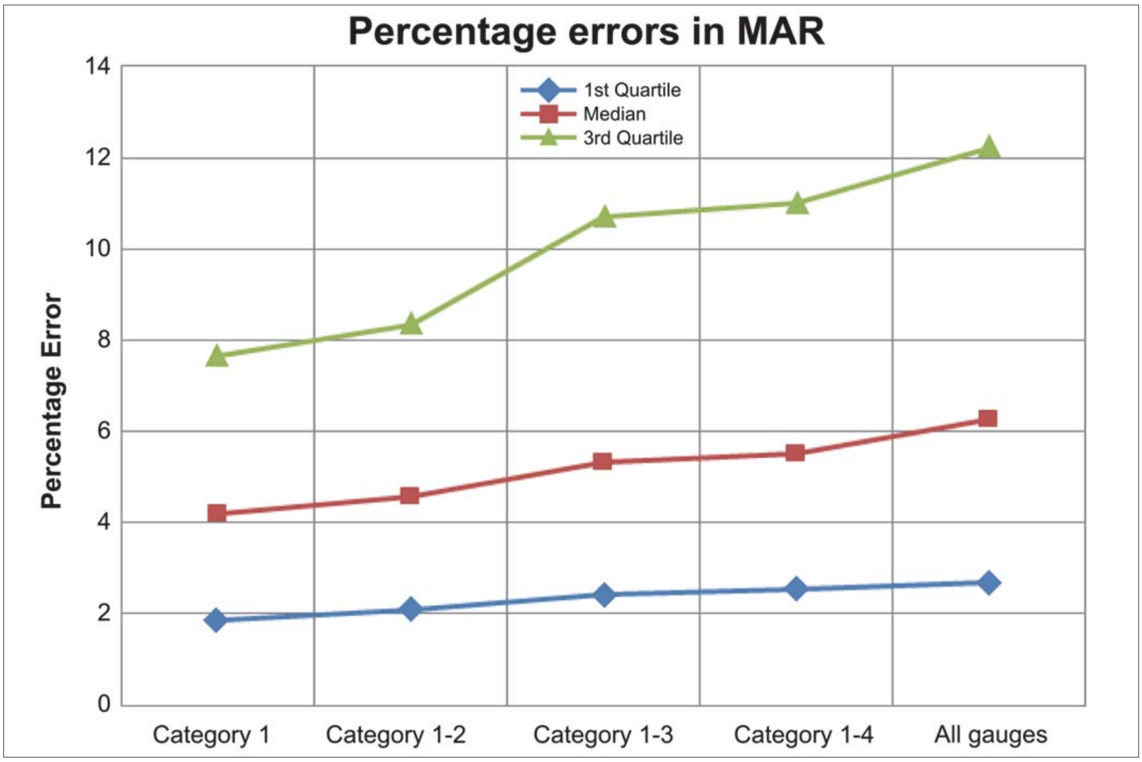

Figure 5.2: Percentage absolute error in MAR

Figure 5.2 shows a fairly gradual increase in the median error from about 4% (category 1) to 6% for all categories. The largest increase occurs when category 3 is included, as this category reflects difficulties with matching observed and simulated MAR: the increase is particularly evident in the 3rd quartile. Another large increase occurs when categories 5 and 6 are included.

5.2.2 Log mean annual runoff

The 1st quartile, median and 3rd quartile errors are plotted in Figure 5.3 for each of the 5 category groups shown above.

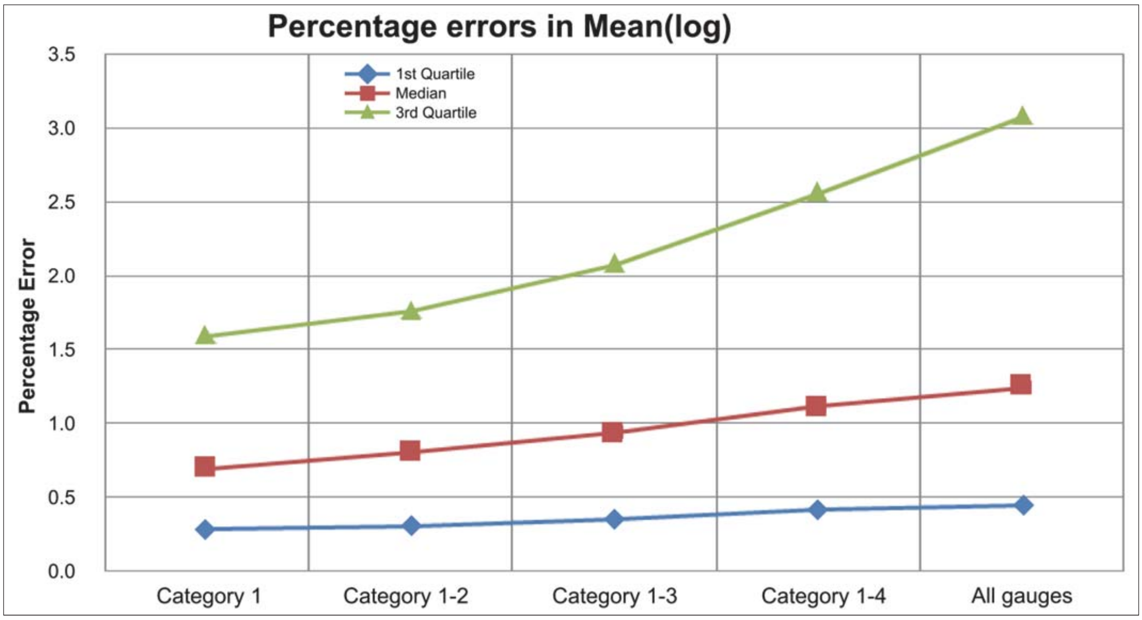

Figure 5.3: Percentage error in log mean annual runoff

The median errors are low at around 1%, but the error for all categories is about double that of category 1. The 3rd quartile also doubles – from about 1.5% to 3%.

5.2.3 Standard Deviation

The 1st quartile, median and 3rd quartile errors are plotted in Figure 5.4 for each of the 5 category groups shown above.

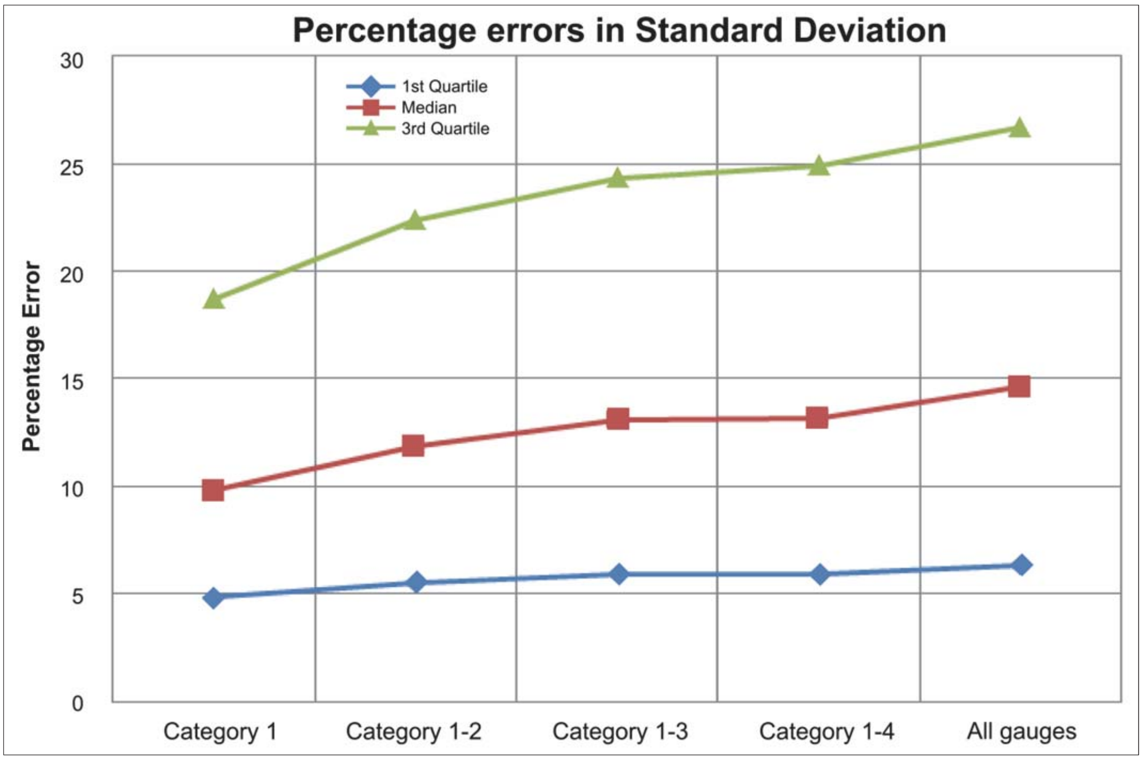

Figure 5.4: Percent error in Standard Deviation

The significant increase with the inclusion of category 2 is due to the impact of outliers on the standard deviation. Thereafter there is a steady decline with the median error for all categories about 50% higher than for category 1.

5.2.4 Log Standard Deviation

The 1st quartile, median and 3rd quartile errors are plotted in Figure 5.5 for each of the 5 category groups shown above.

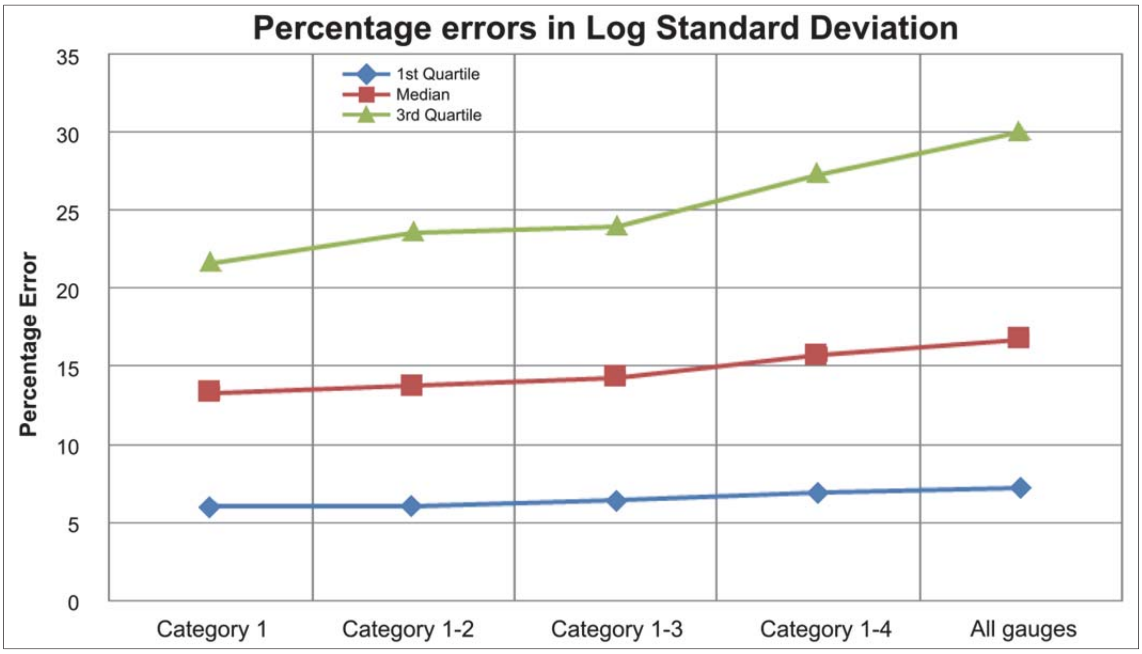

Figure 5.5: Percent error in Log Standard Deviation

The most noticeable increase is with the inclusion of category 4, which is to be expected as this category is related to problems with log statistics. However, even category 1 is associated with relatively high errors for this statistic.

5.2.5 Seasonal Index

The 1st quartile, median and 3rd quartile errors are plotted in Figure 5.6 for each of the 5 category groups shown above.

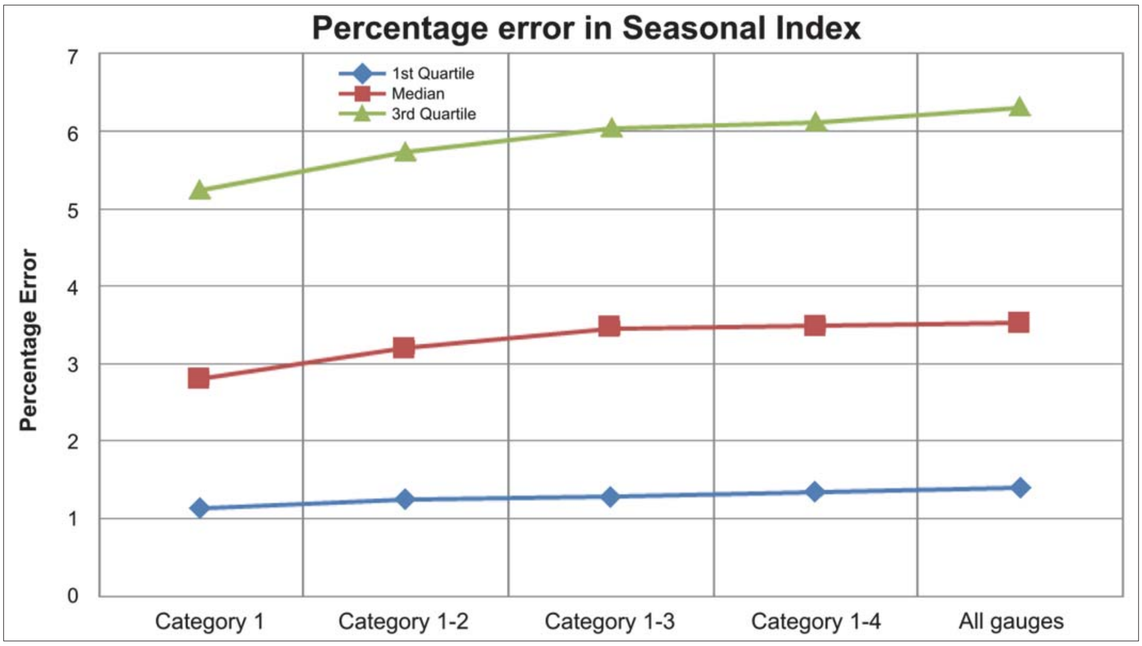

Figure 5.6: Percent error in Seasonal Index

Although category 1 is, as expected, the one with the lowest error, there is relatively little deterioration with the inclusion of the other categories.

5.2.6 Mean error for all statistics

This error serves as an approximate indication of the overall calibration.

The 1st quartile, median and 3rd quartile errors are plotted in Figure 5.7 for each of the 5 category groups shown above.

Figure 5.7: Mean percentage error for all 5 statistics

The graph shows a fairly steady deterioration as each higher category is included: however the errors (1st, 2nd and 3rd quartiles) for the entire data set are only about 50% higher than those for category 1.

5.3 Other factors affecting accuracy of calibration

WRSM/Pitman, like any other rainfall-runoff model, attempts to model the hydrological cycle, whereby the residual streamflow is determined after the rainfall-evaporation driven hydrological processes have taken place. The streamflow is then further modified by the various land-use activities. As rainfall is the primary input to the model, it stands to reason that the closer the residual streamflow is to the causative rainfall, the more accurate will be the model (provided the rainfall data is reasonably accurate). As pointed out in the Introduction, unit runoff for the catchments calibrated in this study varied from about 1mm to over 1000 mm. When expressed as a percentage of MAP, this represents a range of about 1% to well over 50%.

In the following Figures 5.8, 5.9 and 5.10, only the first four categories are plotted since nearly all of category 5 and 6 gauges could not be properly calibrated.

A plot of mean calibration error against unit runoff did show a slight improvement with increasing MAR, but the trend was not significant. It was only with MAR > 100 mm that a clear trend emerged (see Figure 5.8).

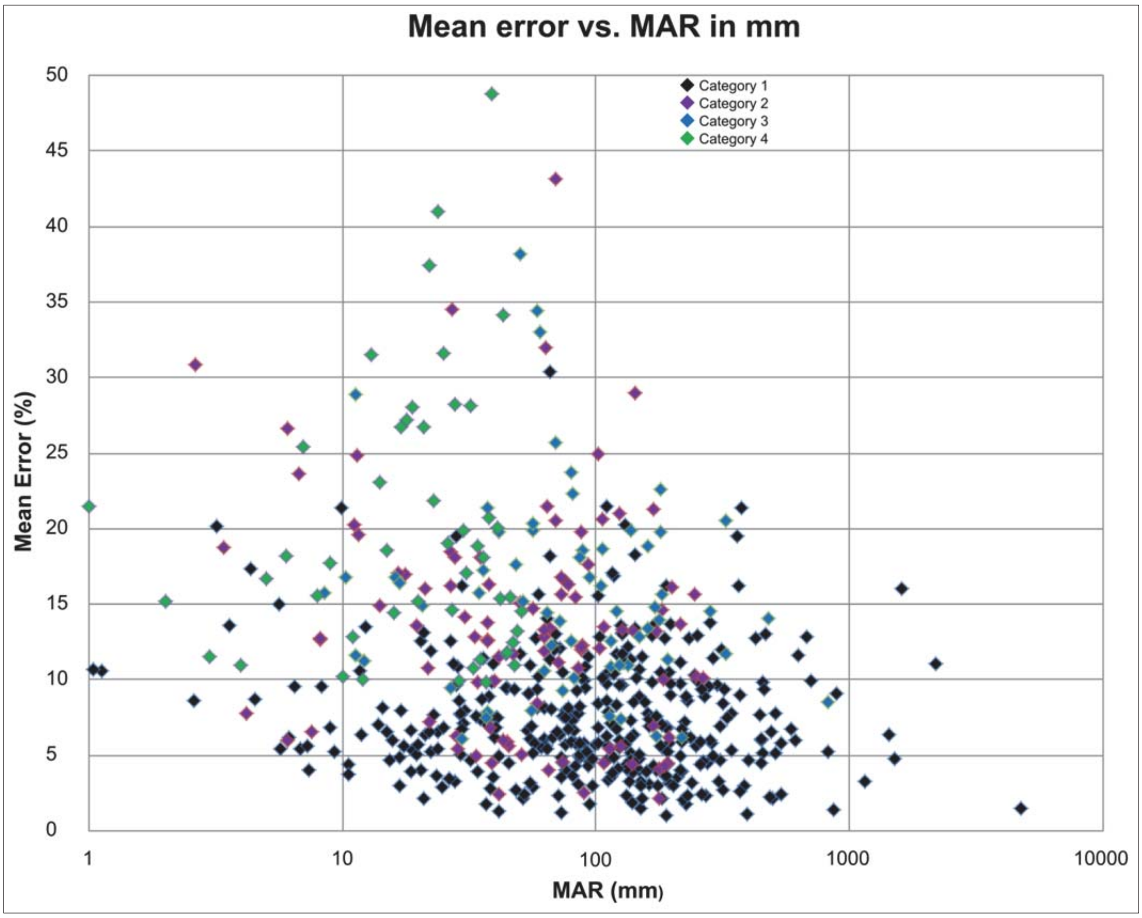

Figure 5.8: Mean calibration error vs. MAR

Another factor that can affect calibration is length of record: the longer the record, the lower impact an outlier will have on the record statistics. Nevertheless, a plot of mean error against record length did not yield a significant trend (see Figure 5.9).

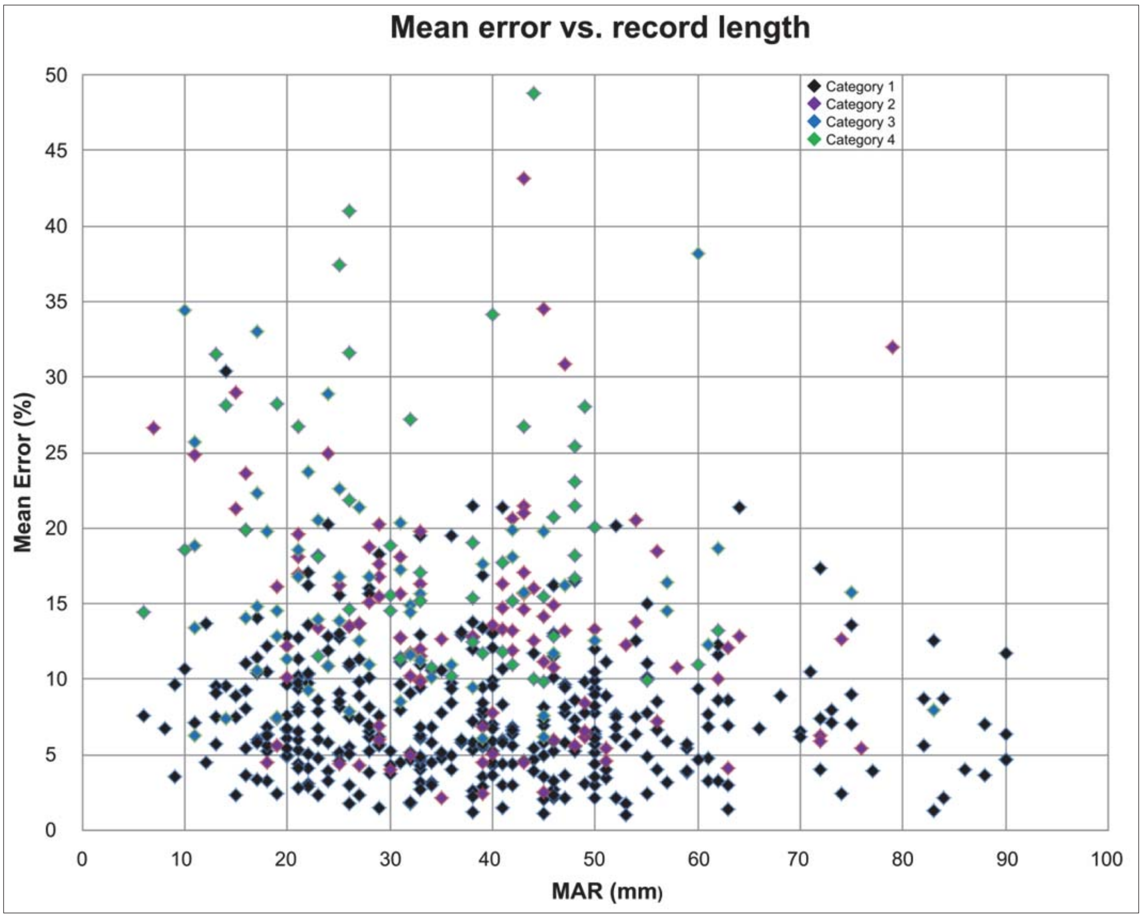

Figure 5.9: Mean calibration error vs. length of record

By the same token, catchment area should also be a factor in evening out outliers. Apart from outliers, larger catchments benefit from the fact that rainfall zones (used to determine catchment rainfall) comprise a number of quaternaries: catchments much smaller than the zone in which it lies will be at a disadvantage in this respect. However, a plot of mean area against catchment area did not yield a significant trend. In fact, some gauges with small catchments produced very good calibrations (see Figure 5.10).

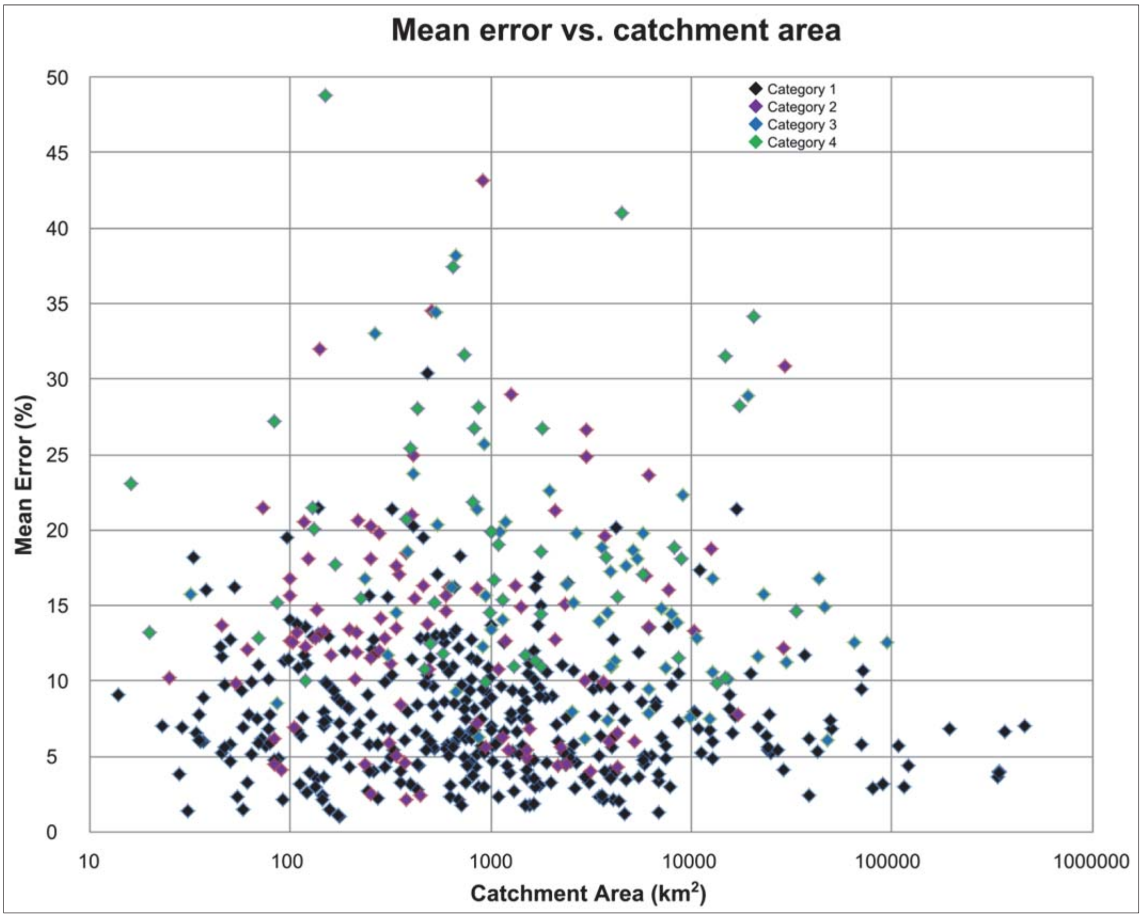

Figure 5.10: Mean calibration error vs. catchment area

Although they display somewhat weak trends, Figures 5.8, 5.9 and 5.10 also show a considerable overlap between the various categories. In other words, category 1 calibrations are not necessarily better than all those belonging to the other categories. What this means is that the results for the other categories would have been better if it were not for the problems associated with these categories, namely outliers, imbalance among gauges on same river and problems with log statistics where zero annual flows are encountered.

Table 5.1 illustrates this aspect with a simple rating of the calibrations based on the mean percentage error of all five statistics.

Table 5.1: Rating of calibrations for each category

Mean % error	Percentage of gauges in category per accuracy rating						
	Cat. 1	**Cat. 2**	**Cat. 3**	**Cat. 4**	**Cat. 5**	**Cat. 6**	**All**
< 5	28	12	0	0	0	0	18
5-10	49	18	19	4	10	3	34
10-15	17	33	36	29	10	6	22
> 15	6	37	45	67	80	91	26

The data in Table 5.1 clearly shows the decline in accuracy as the category number increases, but that even category 1 has a number of gauges (6%) with a mean error > 15%. This percentage increases with each higher category, reaching 91% for gauges in category 6.

A factor that should always be borne in mind when calibrating against observed flow data is the impact of missing/incomplete data that has been patched. Where possible, patching was done by comparisons with other gauges on the same river or in the same catchment. However, in many cases this approach was not possible and one had to use simulated flows for patching purposes. It stands to reason that calibrations on such patched records will appear to be better than they should be due to the use of simulated flows.

6 MODELLING INACCURACY VERSUS GAUGING ERRORS

When one encounters problems during calibration it is not always obvious what the source of the problem is. As discussed in Section 3.6 (dealing with category 6), there is more than one reason for a poor calibration – one of which could be due to errors in the observed streamflow record. However, category 2 (outliers) and category 3 (imbalance among gauges) could also arise from gauging errors. However, one can usually assume that streamflow records associated with category 1 have no significant errors. In summing up, one can infer the following with regard to gauging errors and calibration categories.

Category 1 (no apparent problems): it is unlikely that the record contains any significant errors, so it can be accepted as being reliable.

Category 2 (outliers): this may possibly be due to gauging errors associated with extreme high flows: if the outlier(s) show simulated flows much greater than observed the gauge may be underestimating such high flows. (The opposite applies if the observed flows are higher.) However, it is more than likely that outliers are a result of errors in the causative rainfall.

Category 3 (imbalance among gauges): calibrations in this category result from having to over- or underestimate flows to achieve a balance among gauges on same river (or in same catchment). Such imbalances are of the order of 10 to 20 percent and may warrant a check on the discharge rating of the gauge (gauges with a considerable imbalance belong to category 6.)

Category 4 (problem with log statistics due to zero flows): this is not related to any gauging errors but is most likely due to model's inability to simulate zero flows in certain instances.

Category 5 (short records): as the record is too short for reliable calibration it is not usually possible to infer any gauging errors, unless they are very obvious.

Category 6 (calibration not possible): as explained in Section 3.6, there are other reasons for this category, apart from gauging errors. However, gauging errors in this category can sometimes be detected by studying the yearly hydrograph to detect a period where the observed flows do not appear to conform to the remainder of the record.

7 SUMMARY AND CONCLUSIONS

The question is often asked by users of the WRSM/Pitman model – *"what constitutes a good (or acceptable) calibration?"* There is no straightforward answer to this question; nevertheless it is necessary to provide the user with some guidelines concerning this aspect of calibration. Accordingly, the calibrations undertaken for WR2012 have been analyzed for the purpose of showing the user what to expect when calibrating the model against observed streamflow records.

The analysis showed that not all calibrated streamflow records could be simply lumped together for the purpose of establishing the accuracy of the calibrations. Not only does the accuracy of the streamflow records play a role: other factors such as confidence in land use and rainfall input also have an impact on the calibrations.

A detailed analysis of the calibrations indicated that they could be placed into 6 different categories as follows:

- (1) No apparent problems;
- (2) Problems caused by outliers;
- (3) Problems caused by imbalance among records on same river (or in same catchment);
- (4) Problems with log statistics caused by zero or near- zero annual flows;
- (5) Problems due to very short records and
- (6) Data problems rendering calibration impossible.

Although the various graphs provided by WRSM/Pitman were used to help classify the calibrations in addition to the statistics of the streamflow records, it was the statistics that were used to assess calibration accuracy as these are strictly numerical. It is important to stress, however, that the graphs are an extremely useful aid to calibration and that it would be unwise to base any calibration solely on statistics.

The statistics used to rate the calibrations were those given in the model's calibration screen, namely:

MAR – mean annual runoff

- Mean (log) – mean of logs of annual flows;
- Std. dev. – standard deviation of annual flows;
- Std. dev. (log) – standard deviation of logs of annual flows and
- Seasonal index – mean variability of flow over a year.

The percentage error in each statistic was determined for each gauge where a calibration was undertaken. Also, as a measure of the overall calibration, the mean (absolute) percentage error was obtained.

The first tests undertaken were those for bias in any of the statistics by examining the actual errors, i.e. those with a + or – sign. This test showed a tendency to underestimate the log standard deviation by about 10%, but this result was not unexpected, owing to the problem with zero annual flows. The remaining four statistics showed no bias, although the range in standard deviation was relatively high due to the influence of outliers.

The tests on the absolute errors revealed an expected deterioration as one moves from category 1 to category 6, although median errors for the entire data set are only about 50% greater than for category 1.

The graphs relating mean percentage error (for all 5 statistics) against MAR, length of record and catchment area revealed a slight tendency for improvement with an increase of these variables. What is perhaps more significant is the overlap between the various categories. In other words, a category 1 calibration is not necessarily better than any other category. What it means is that the calibrations in the higher categories would have been better if the problems associated with these categories did not present themselves.

In summing up, it is perhaps appropriate to mention the existence of automatic calibration algorithms, which rely on optimization of an objective function related to the minimization of differences between characteristics of observed and simulated streamflow records. Whilst automatic calibration can be useful, especially where there are no apparent data problems, they can go astray where outliers have a large impact (category 2) or where it is necessary to settle for a less than optimum calibration (category 3). As pointed out in the course notes for WRSM/Pitman, it is advisable to have an understanding of how the model functions and, in particular, the role of the various parameters before attempting to calibrate the model. Calibration can be thought of as a heuristic process, whereby users learn by experience to improve their skills at calibrating the model.

Appendix A shows calibrations at a few select gauges, spread across the study area and with a range of hydrological characteristics, to indicate what is meant by a “good calibration”.

APPENDIX A – RESULTS OF SELECTED CALIBRATIONS

The following table summarizes the characteristics of 5 gauges spread across the study area in addition to the calibration results in terms of errors in the 5 main statistics. The following pages contain the graphs of annual and mean monthly flows, as they are the most important graphs used for calibration refer to Figures A.1 to A.10.

Gauge	B6H003	E2H002	K3H003	U2H011	V1H038
River	Treur	Doorn	Maalgate	Msunduze	Klip
WMA	4	17	16	11	7
Rainfall season	Summer (dolomite)	Winter	All-year	Summer (coastal)	Summer (inland)
Catchment area (km^2)	92	6903	145	176	1644
MAP (mm)	1352	413	753	943	826
Record	1959-2009	1922-2009	1960-2009	1957-2009	1978-2009
% error in:-					
MAR	1.50	1.41	1.70	0.48	0.10
Mean(log)	0	0.23	0.30	0.29	0.24
Std. dev.	6.40	0.10	1.75	4.14	2.38
SD (log)	0	3.85	6.25	0	5.88
SI	2.79	1.00	0.64	0.21	0.68
Mean	2.14	1.32	2.13	1.02	1.86

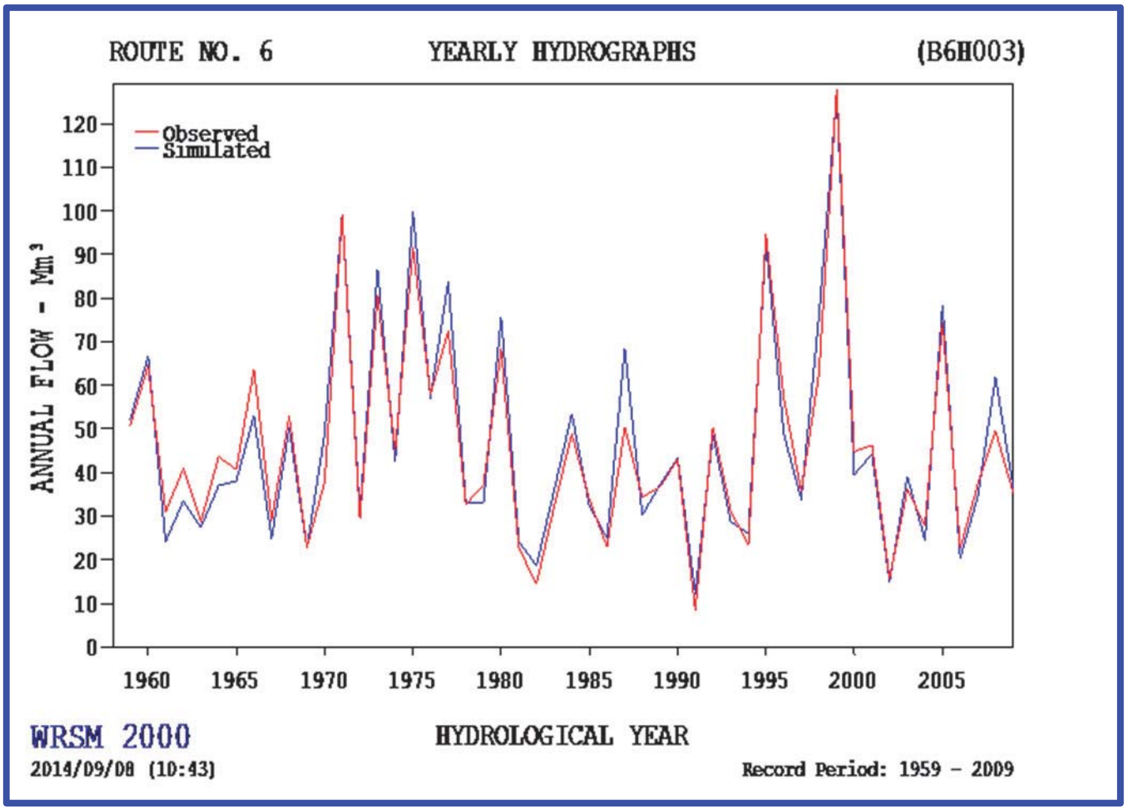

Figure A.1: Yearly Hydrograph for B6H003

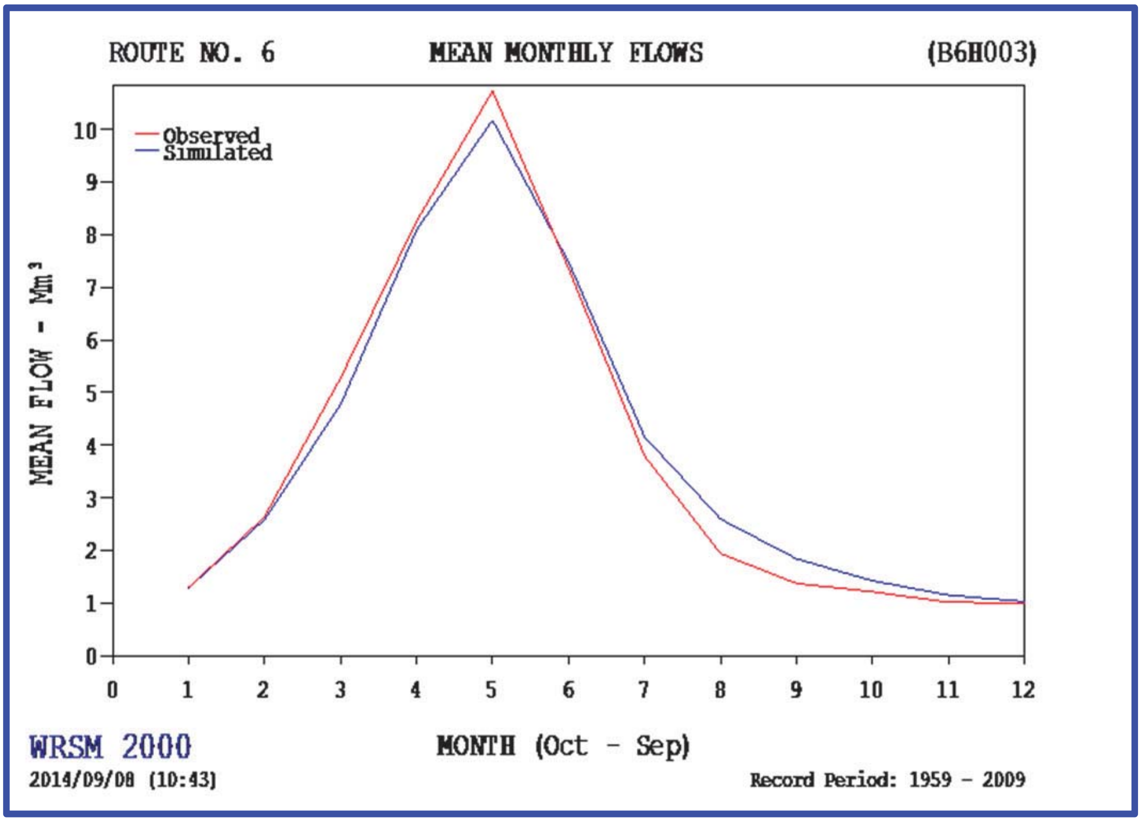

Figure A.2: Mean Monthly Flows for B6H003

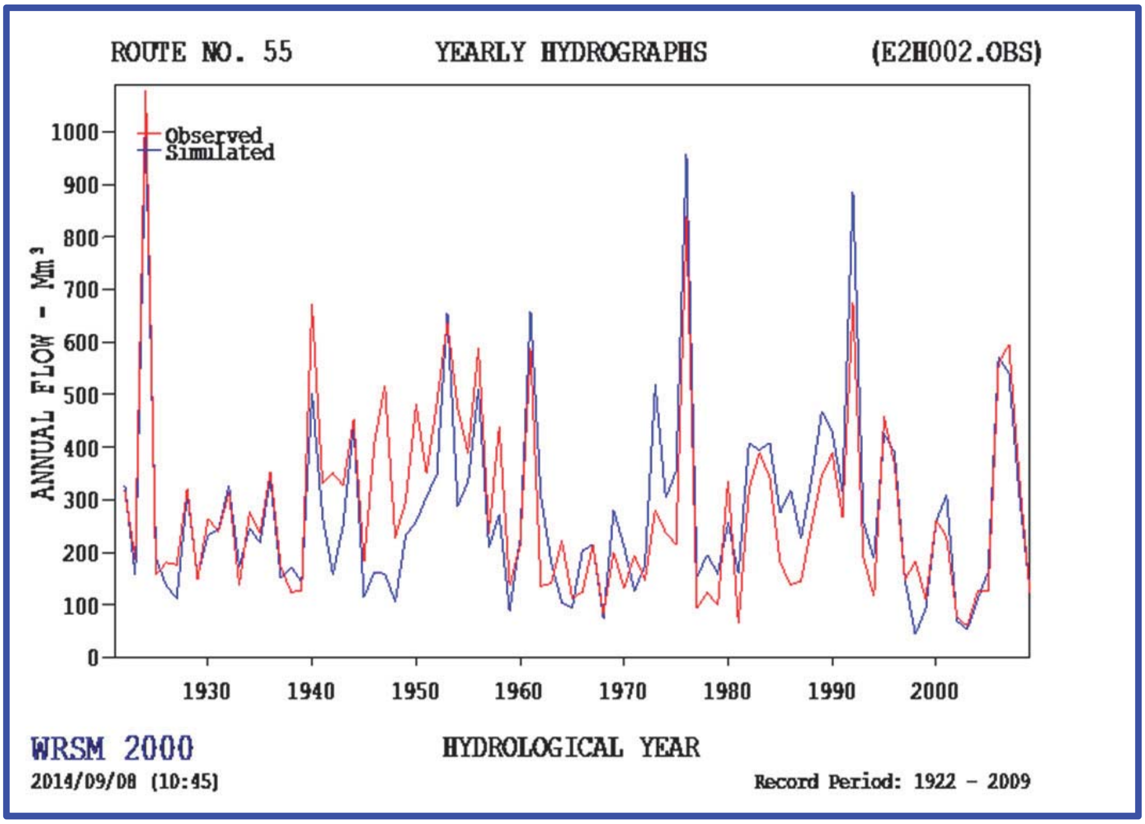

Figure A.3: Yearly Hydrograph for E2H002

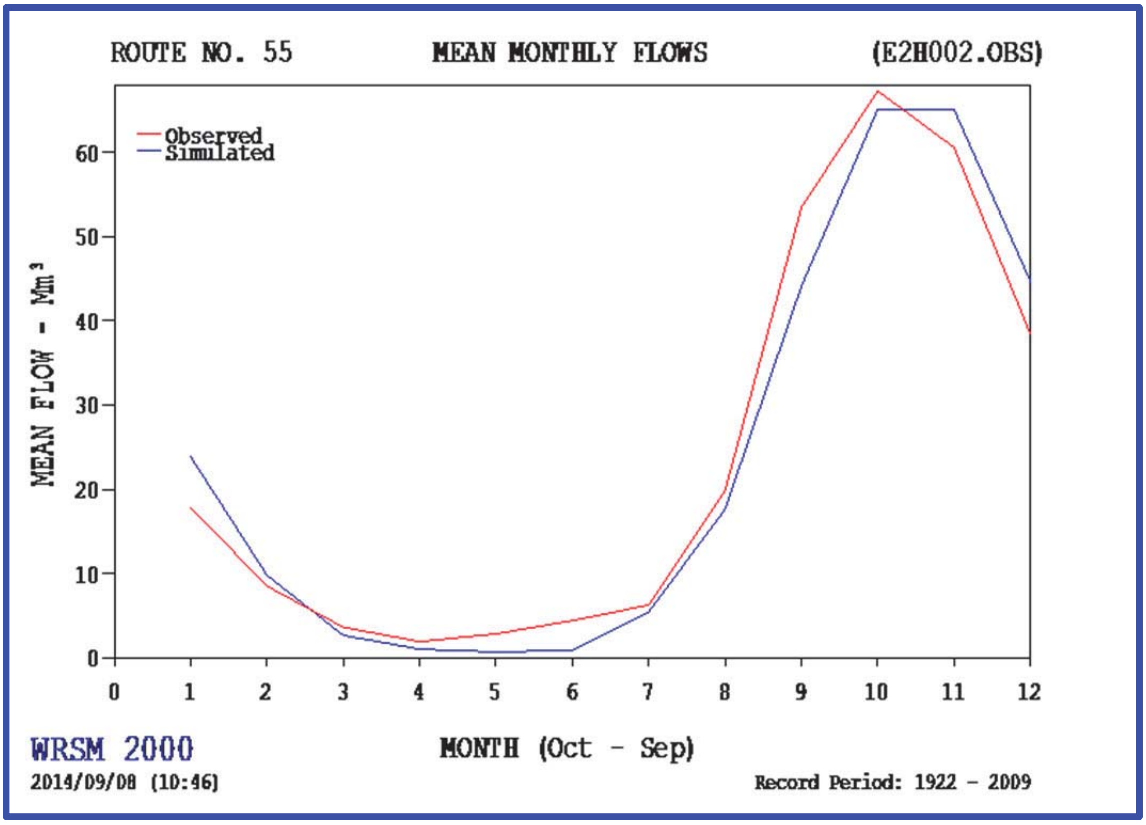

Figure A.4: Mean Monthly Flows for E2H002

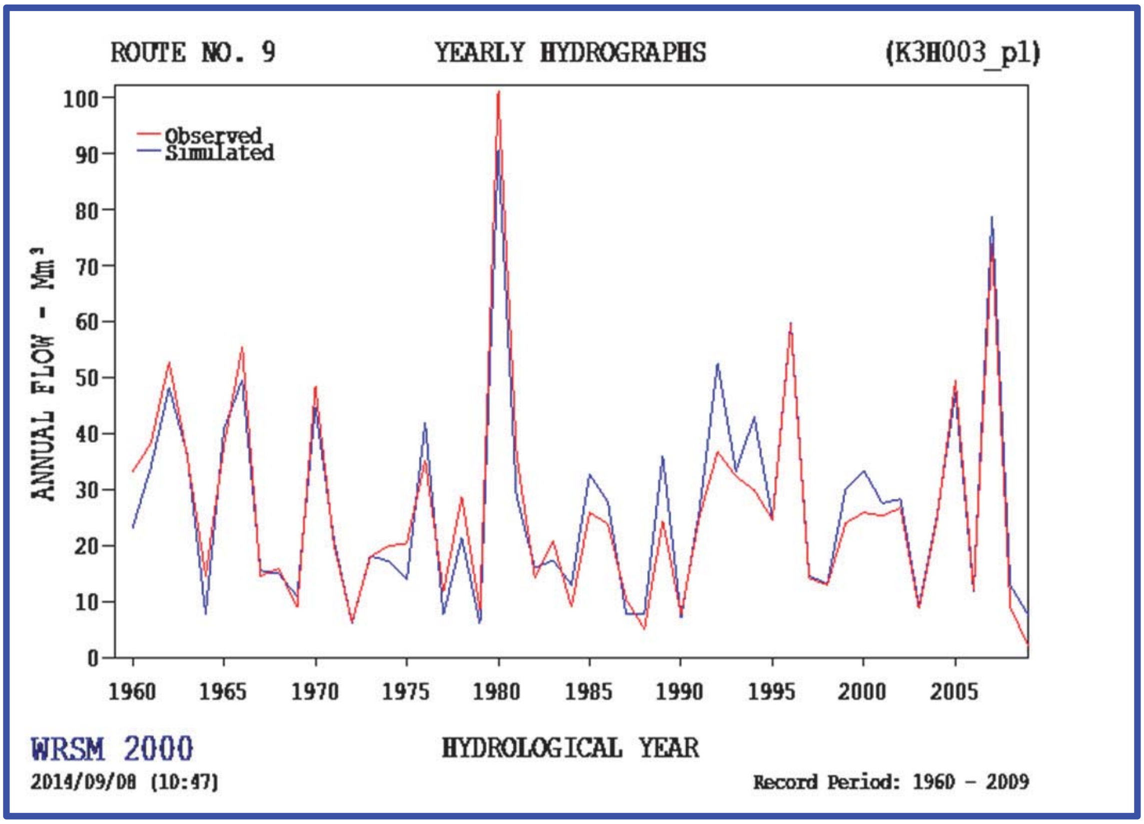

Figure A.5: Yearly Hydrograph for K3H003

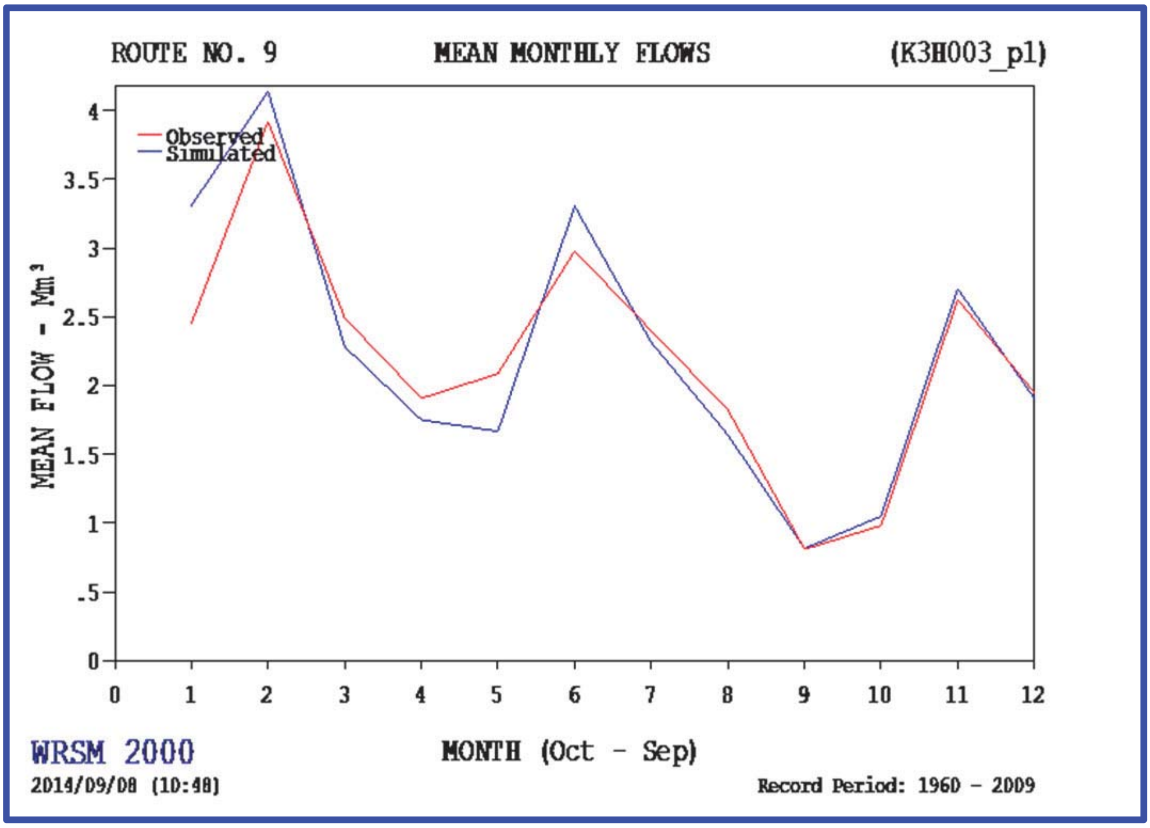

Figure A.6: Mean Monthly Flows for K3H003

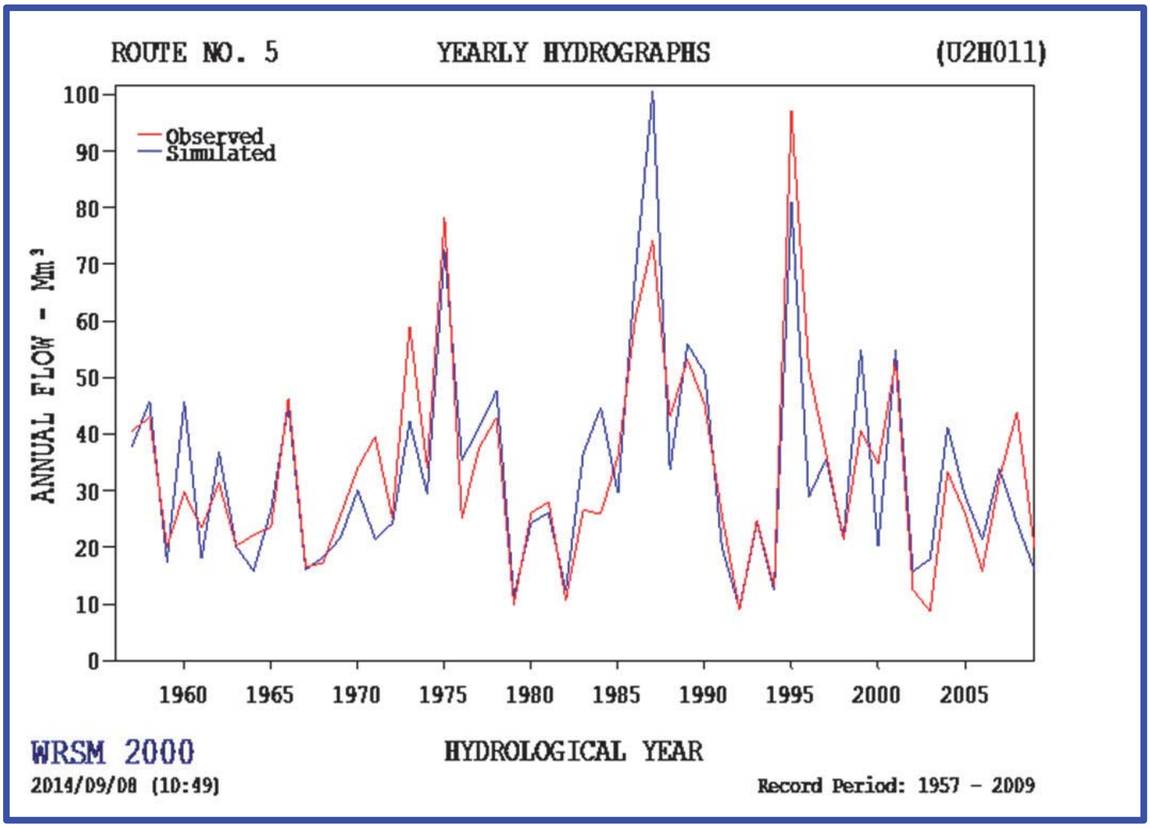

Figure A.7: Yearly Hydrograph for U2H011

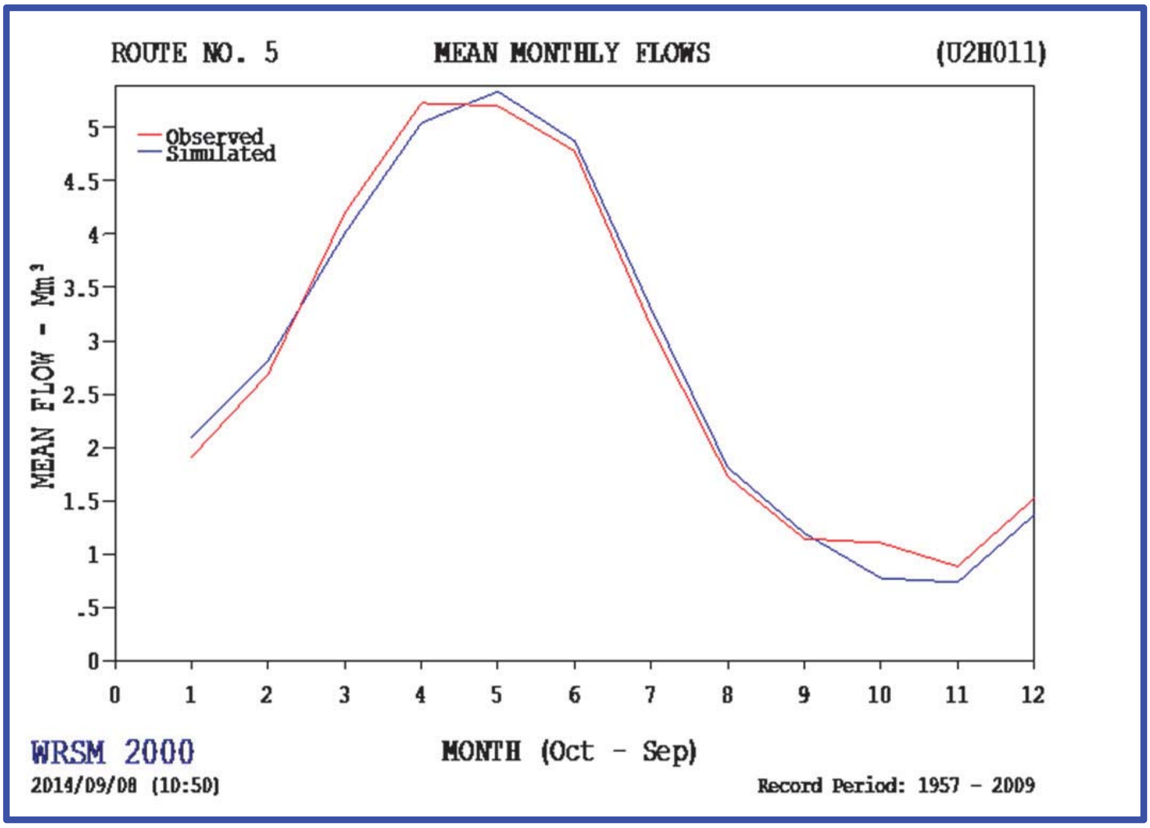

Figure A.8: Mean Monthly Flows for U2H011

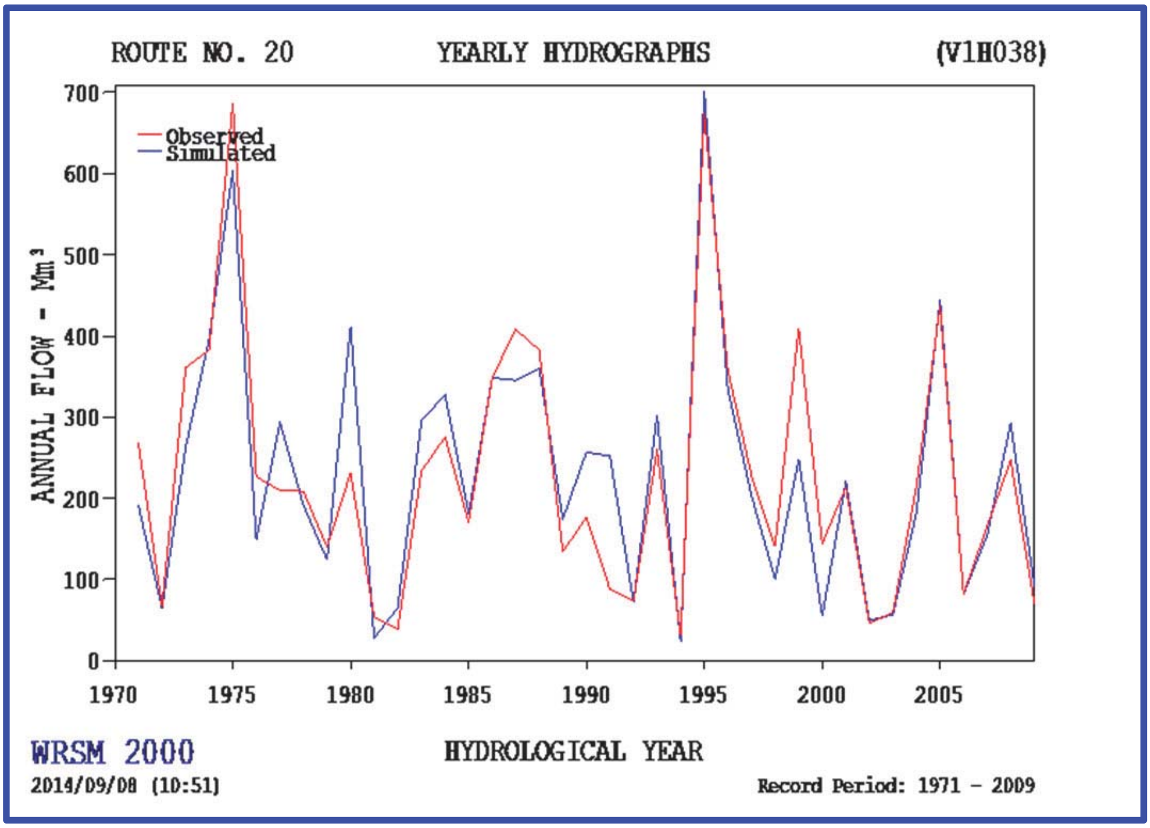

Figure A.9: Yearly Hydrograph for V1H038

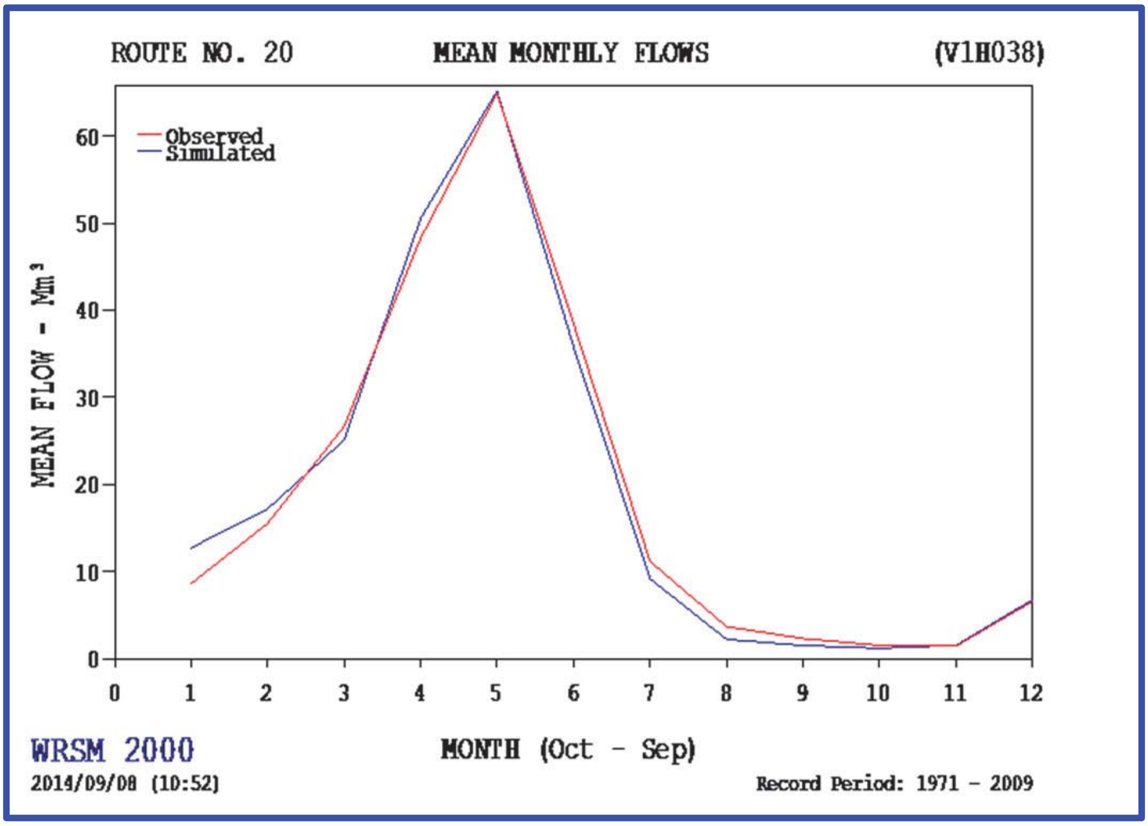

Figure A.10: Mean Monthly Flows for V1H038

9781431208449