Streamflow Transformation for Climate Change

Department of Sustainability and Environment, Vic
& Victoria University

REALM

A program for the simulation of water supply systems

User Guide

21 April 2008

Contents

1. Introduction to the Streamflow Transformation Utility 1

1.1 What is the Streamflow Transformation Utility? 1

2. Investigation of Step Change Methods 3

2.1 Introduction 3
2.2 Factoring of inflows 3
2.3 Yield impacts – Latrobe System case study 5
2.4 Summary of case study findings 7
2.5 Further work 7
2.6 Further work - Outline of findings 7
2.7 Matching flow percentile 9
2.8 Factoring by decile range location 9
2.9 Standardisation 10
2.9.1 Standardising using overall time series properties 10
2.9.2 Standardisation using monthly flow properties 11
2.10 Transformation into a normal distribution 13
2.11 Fitting a Beta distribution to the Flow Duration Curve 15
2.12 Conclusions and Recommendations 17

3. Seasonal Factoring Method 21

4. Comparison of Flow Duration Curves Method 24

4.1 Step-by-step explanation of algorithm 24
4.2 Case using 1 bin 27
4.3 Case using 2 bins 27
4.4 Case using 10 bins 28
4.5 Conclusions 31

5. Analyses for identification of ‘break-point’ date 32

5.1 Single Mass Curve Analysis 32
5.2 Ratio of Averages Analysis 34

1. Introduction to the Streamflow Transformation Utility

This report gives a brief description of the REALM utility "Streamflow Transformation", which can be accessed under the Utilities menu in REALM. The utility is available in REALM beta version 5.12c, which was released in April 2008. This report also contains notes on some of the key algorithms that are executed within this utility.

Section 1.1 below provides an introduction to what the utility does. Section 2 outlines the different options for transformation that were considered for the purposes of water resource modelling under climate change / drought. Sections 3 and 4 describe in more detail the two chosen options. Section 5 outlines two methods within the utility for detection of the most appropriate "Break Point".

Please note that this report is intended to be read in conjunction with the built-in help boxes contained within the utility itself. There are a number of key limitations of the utility, and to ensure user accessibility these limitations are described within the user help notes in REALM. This report is designed to complement these notes, rather than be used as a comprehensive, stand-alone document.

1.1 What is the Streamflow Transformation Utility?

The Streamflow Transformation Utility is a REALM utility that allows users to alter input flows, such that streamflow properties that occur in one period of the flow record (such as flow reductions due to climate change) are replicated throughout the entire flow record.

The utility was developed to evaluate the effect of climate change on water resources modelling. To use the tool effectively in this context, the user must assume "climate step-change". That is, the user must select a date, whereby all flows preceding that date are assumed to be "pre" climate change, and all flows after the date are assumed to be influenced by climate change. Throughout this document, this date is referred to as the "Break Point Date". Note that this utility does not allow representation of climate change as a trend – only as a "step-change". Once the "before and after" periods are defined, the utility will then alter the "before" flows such that their properties match the properties of the "after" flows. This is done in one of two ways:

1) **Matching of seasonal averages:** the user defines a seasonal regime (up to 12 seasons) and the utility adjusts "before" flow values such that, for each season, the average seasonal flow before the break point date matches the average seasonal flow after the break point date.
2) **Matching of flow duration curves (FDCs):** the "before" flows are adjusted such that the FDC for "before" flows matches the FDC for "after" flows. This transformation relies upon

splitting the FDCs into segments (eg. deciles or percentiles – the user can decide) and comparing flow averages, segment by segment.

Information on each of these transformation methods is available in Sections 3 and 4.

2. Investigation of Step Change Methods

The following sections describe initial work undertaken to investigate methods of transforming flows to represent climate “step-change”. Sections 2.1 to 2.4 describe a case study on three potential methods, and Sections 2.5 to 2.12 describe later work done to examine five separate methods.

2.1 Introduction

In 2006, the Department of Sustainability and Environment (DSE) sought to determine if there is a better way of modelling the impact that a continuation of post-1997 drought conditions would have on Victoria’s water resources.

The (then) current method of flow adjustment involved factoring down inflows using a single linear factor to make the average flow conditions prior to July 1997 to be equal to the average flow conditions from July 1997 onwards. There was concern that this method makes already severe single year droughts such as 1982/83 far more severe than they were historically.

A workshop was held on 20 October 2006 to discuss other potential options to model this scenario. One of the options that arose from that meeting was to make an adjustment to streamflows prior to July 1997 such that they would match the flow duration curve from July 1997 onwards. The Latrobe River Basin was selected for trial application of this method.

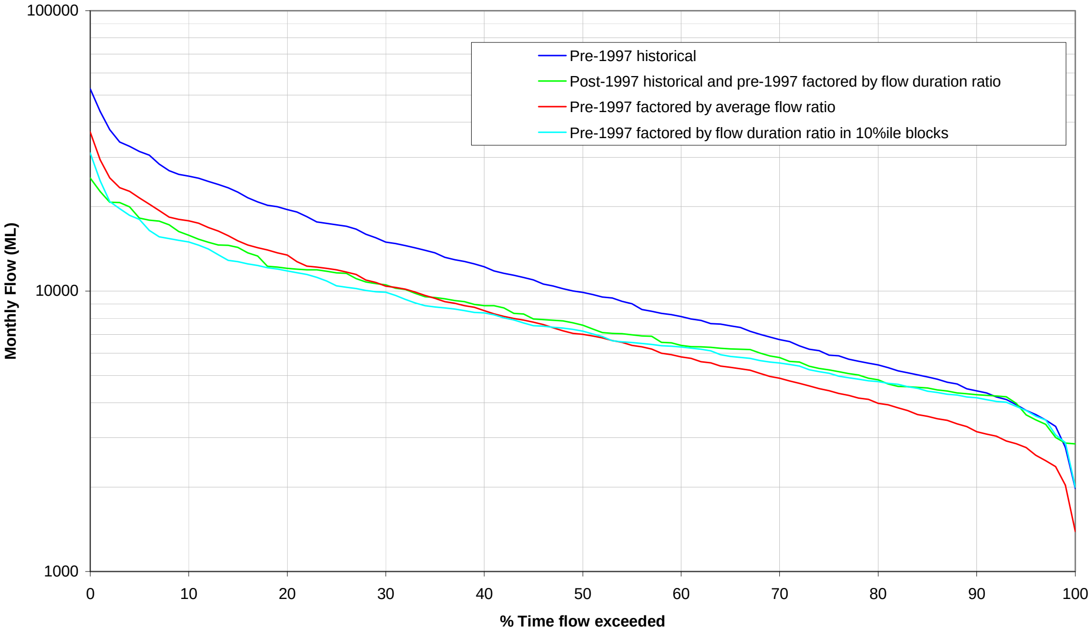
The post-1997 period included data up to 2006 for the five largest inflows to the Latrobe REALM model, with data up to 2004 being used for the other minor inflows.

2.2 Factoring of inflows

Two methods were investigated for the factoring of inflows based on flow duration properties. These were:

a) Transform pre-1997 flow in month *i* so that it is equal to the post-1997 monthly flow with the same probability of exceedance;
b) Transform pre-1997 flow using factors calculated as the ratio of the pre- and post-1997 representative inflows associated with exceedance probability at each 10% interval.

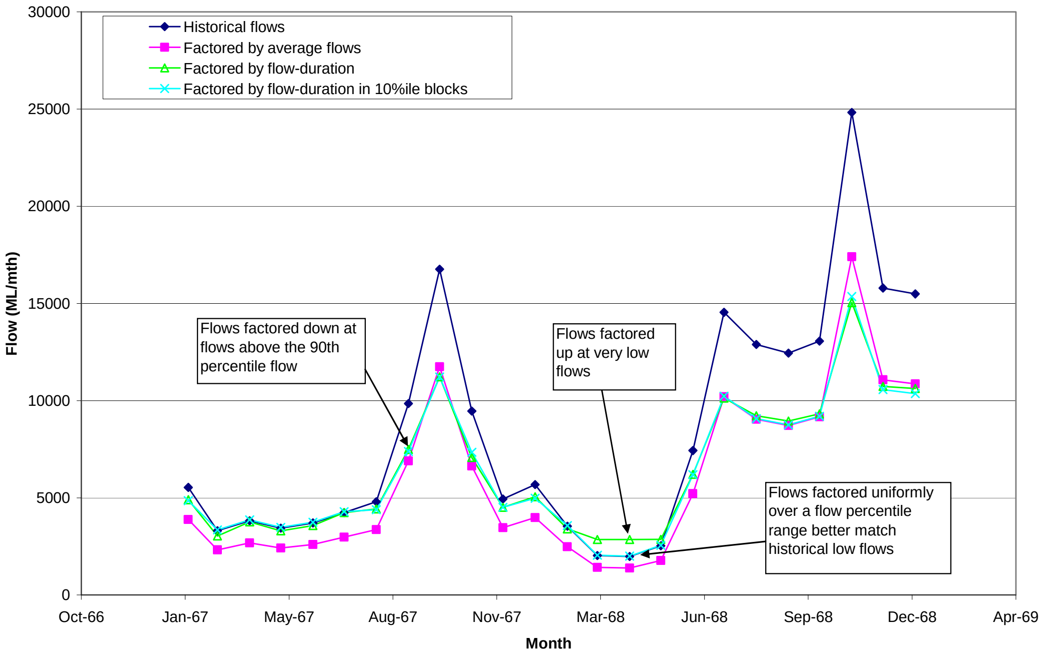
The resulting flow duration curves when using the above methods for inflows to Blue Rock Lake in the Latrobe basin are shown in Figure 2-1. It can be seen from this graph that for this particular inflow, the flow-duration curves for flows between the 90th and 98th percentile low flows are the same pre-1997 and post-1997. This means that these flows are not adjusted. All other flow percentile values are adjusted. This results in a factoring down of all flows above the 90th percentile low flow for both methods and a factoring up of all flows below the 98th percentile low flow for method (a). This effect is not observed to the same extent in method (b), which helps to overcome extreme drought sampling errors when using the relatively short post-1997 period by averaging out the flow duration curve adjustment over a wider flow range.



- **Figure 2-1 Flow-duration curves for inflows to Blue Rock Lake (S216)**

This is illustrated further in a sample of the streamflow hydrographs, shown in Figure 2-2 for the 1967/68 drought. It can be seen that at low flows, the flows adjusted using the continuous flow duration curve are generally above those adjusted by the mean flow. Flows adjusted using the discretized flow duration curve generally better match the historical very low flows.

In other inflow sequences, it is possible that some high flows would be factored up as well when using the continuous flow duration curve to adjust flows. Floods that occurred in some parts of the state in the post-1997 period could be worse than anything observed prior to 1997, which is consistent with climate change predictions for more intense storms. Flood events occur at infrequent intervals and are of highly irregular magnitudes. At high flow percentiles (eg above the 20th percentile flow), this method of factoring inflows based on flow-duration properties suffers from having only a small sample size of flow events. This weakness is common to all adjustment methods. A caveat should therefore be placed on this technique that it is to be used for yield analysis only and is not suitable for analysis of flood events pre and post 1997.



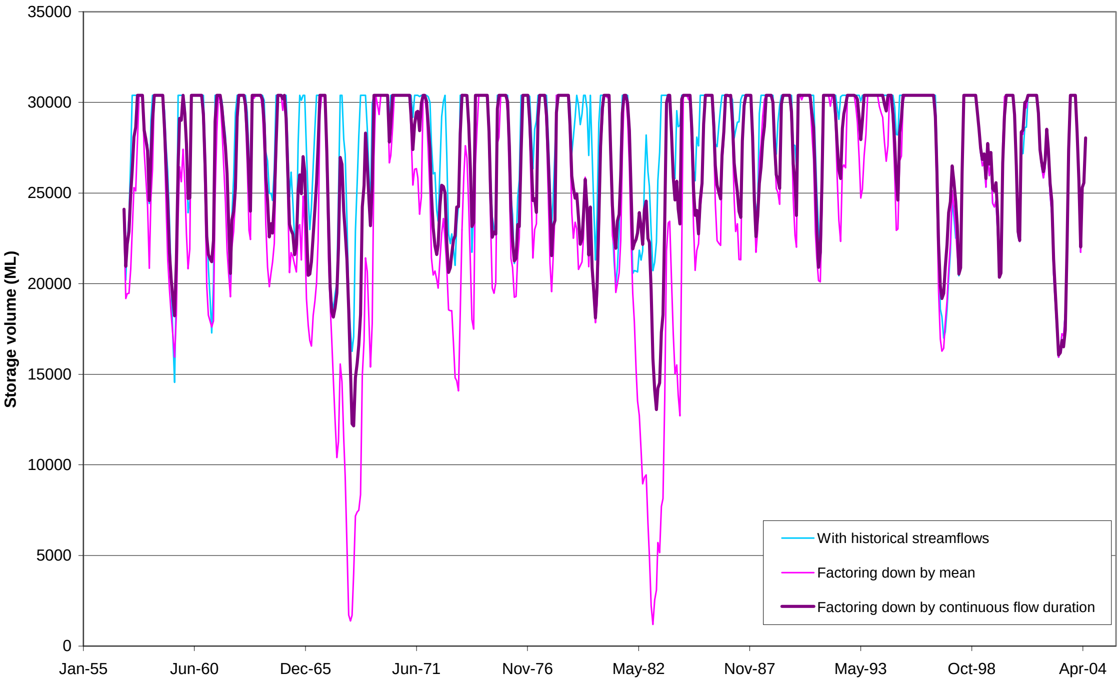
- Figure 2-2 Time series of 1967/68 inflows to Blue Rock Lake (S216)

A further potential minor complexity in the application of both the flow-duration adjustment and the adjustment of flow averages is the treatment of cease to flow conditions. In the work done for the Latrobe River basin, it has been assumed that cease to flow conditions are preserved where they occur in the historical sequence. This is of minor importance for the Latrobe River basin but may be much more significant in ephemeral systems in the west of the state.

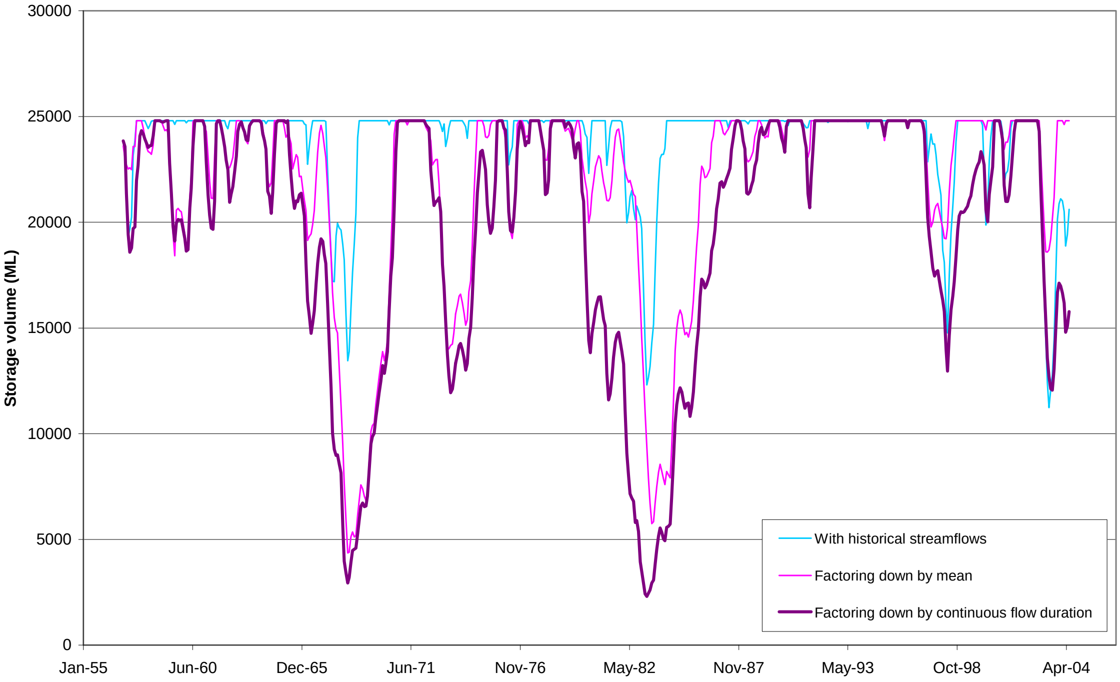
2.3 Yield impacts – Latrobe System case study

The derived streamflow sequences were input into the Latrobe REALM model to determine impacts on yield. It was found that overall yield dropped by 25% when using the flow-duration curve adjustment, compared with 37% when using the adjustment of average flows.

Storage behaviour when using the flow-duration curve adjustment method appears to provide an increase in the frequency of drawdown prior to 1997 rather than the increase in severity observed previously when using the factoring down by means. Drawdown in multi-year storages in multi-year droughts remains similar in both cases. This is illustrated in Figure 2-3 and Figure 2-4.



- **Figure 2-3 Time series of storage volume in Moondarra Reservoir**



- **Figure 2-4 Time series of storage volume in Gippsland Water’s share of Blue Rock Lake**

2.4 Summary of case study findings

On the basis of material presented thus far, it would appear that the flow duration curve adjustment to pre-1997 using a discretised flow duration curve provides a better replication of post-1997 flow conditions during the pre-1997 flow period. It was thus recommended that this method should be adopted in preference to the adjustment based on mean flow conditions.

Further investigations (undertaken at a later date) appear below. These investigations take a more detailed look at five different methods for transformation of flows, including the flow duration curve methods (both by decile and percentile) plus three others.

2.5 Further work

The sections below discuss various approaches to transforming pre 1997 streamflows into a series that has similar characteristics to the flow observed post 1997 in the Latrobe river basin; in this case the distribution, average and standard deviation of the post 1997 flow are thought to be of main importance. Five methods of transforming monthly flows are here discussed:

1) matching of flow percentiles
2) factoring by decile range location
3) adjustment by standardisation
4) adjustment using a normal transformation
5) adjustment using a beta-distribution fit

A short outline of the typical findings is given in Section 2.6, while more detailed descriptions of each method are provided in the later sections.

It is worth noting that methods 1 to 3 are simple to implement automatically while methods 4 and 5 require intervention from the user and are a bit more complex to use.

2.6 Further work - Outline of findings

The transformation of the pre 1997 flows into a time series having similar statistical properties to the post 1997 period flows has here been carried out for the REALM model input representing Sheepwash Creek in the Latrobe River basin.

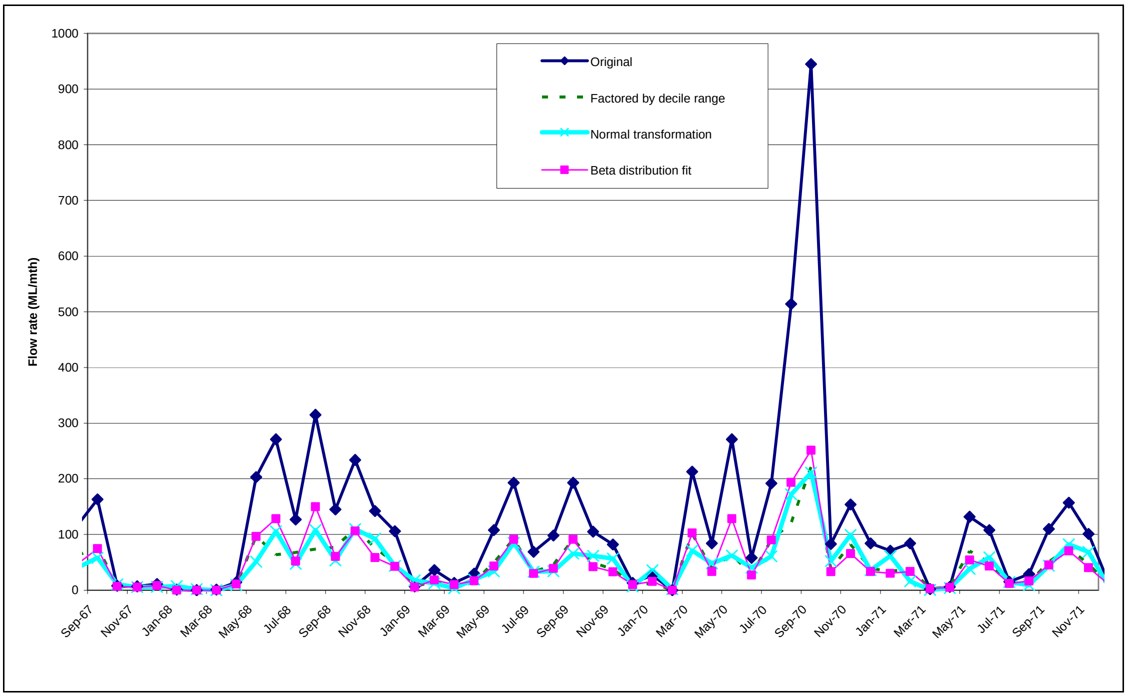
The flow series for Sheepwash Creek is a combination of gauged flow (1974 to 1981) and a calibrated HYDROLOG (a rainfall-runoff model) model output for the period 1957 to 1975 and 1981 to 2004, to which estimated private diversions and farm dam impacts have been added back in order to estimate the natural flow.

Table 2-1 summarises the statistical properties of the pre 1997 flows, the post 1997 flows and the transformed pre 1997 flows under each method used. A time based plot of selected transformed time series is presented in Figure 2-5a and b. They show that the transformed flows are similar in magnitude to each other on average, but the degree to which they follow the pattern of the original series is varied. It would seem that methods 2 and 5 tend to be the ones that follow the pattern of the original pre-97 series the closest without reducing the low flows excessively.

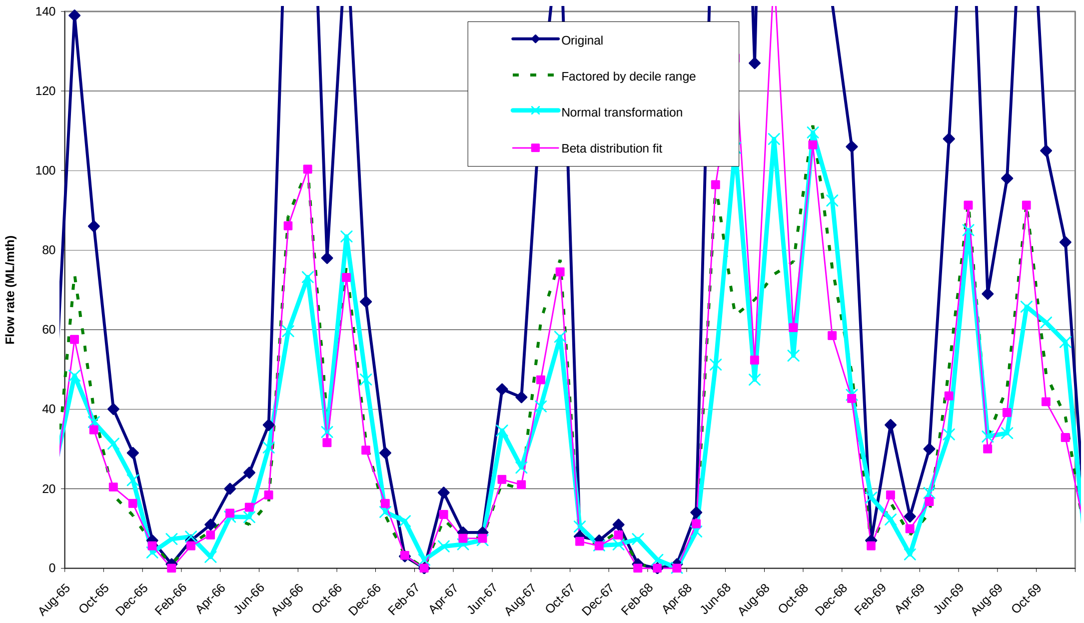
On the basis of ease of implementation, it is here recommended to use method 2 for streamflow transformation.

■ **Table 2-1 Summary of statistics for transformed series**

	Mean (ML/mth)	Standard deviation(ML/mth)
Pre-1997	**145**	**338**
Post 1997	**49.1**	**66**
Percentile adjustment	47.8	60
Factoring by decile range	48.8	80
Standardisation	49.3	66
Normal-transformation	49.4	83
Beta-distribution	49.0	69



■ **Figure 2-5a Comparison of generated transformed time series for selected transformation methods**



- **Figure 2-5b Comparison of generated transformed time series for selected transformation methods (typical low flow event)**

2.7 Matching flow percentile

This method consists of deriving the percentile values for the flow in each month of the pre-97 and post-97 series. The value of the flow in the pre-97 series is then replaced by the flow corresponding to the same percentile in the post-97 time series.

This method is very simple to apply and produces a transformed series that has values of mean and standard deviation that are very close to the ones of the post-97 series.

The only drawback to this method is that the post-97 time series is a very short time series whose statistical properties could be significantly affected by the next data point to be added to the series, and a flow-by-flow replacement can sometimes produce some unexpected fluctuations the resulting time series. This was discussed in Section 2.2 above.

2.8 Factoring by decile range location

This method consists of deriving the average flow values for each decile of the pre-97 and post-97 series. The value of the flow in the pre-97 series is then factored by the post-97 to pre-97 proportion of the average flows for the decile in which this flow is located.

This method is very simple to apply and produces a transformed series that has values of mean and standard deviation that are very close to the ones of the post-97 series.

Similarly to the first method, the only drawback to this method is that the post-97 time series is a very short time series whose statistical properties are yet uncertain, and this method can sometimes produce some inconsistencies in the resulting time series with respect to the original pre-97 series.

2.9 Standardisation

In an attempt to produce a transformed time series with exactly the same mean and standard deviation as the post-97 time series, a method that targets the replication of these properties was trialled.

2.9.1 Standardising using overall time series properties

This approach involves converting a flow point in the target series (pre-97) into a flow point in the base series (post-97) that is as far from the mean of the distribution relative to the standard deviation. Here the standardized value (Z_b) for the flow (Y_b) in each month prior to 1997 is corrected back to a flow value (Y_i') that would correspond to that standardized value (Z_t) in the post 1997 flow distribution, where:

$Z_t = (Y_t - \mu_t)/s_t$ Eq. 1 (post 1997) and,

$Z_b = (Y_b - \mu_b)/s_b$ Eq. 2 (pre 1997)

Where

μ is the mean monthly flow, and s is the standard deviation of flows.

Thus,

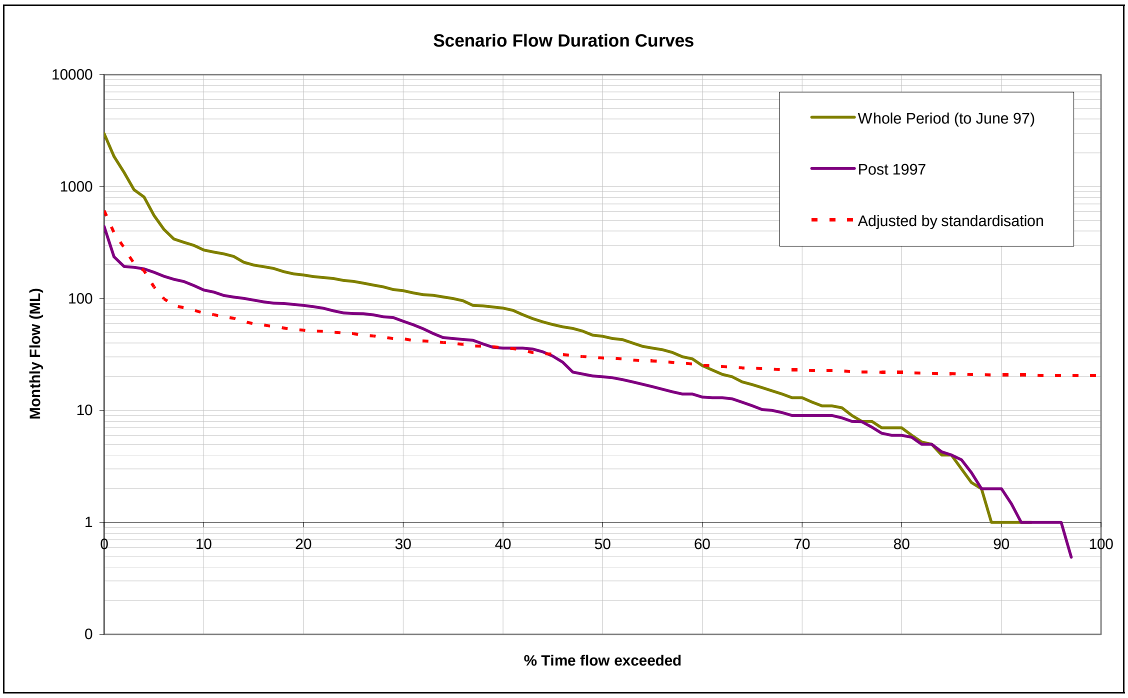
$Y_i' = Z_t . s_b + \mu_b$ Eq. 3

This forces the transformed series to have the same mean and standard deviation.

This method causes a significant compression in the range of the time series around the new mean value, resulting in a reduction in the magnitude of high flow events and a corresponding increase in magnitude of low flows. This causes intuitively unacceptable results at low flows in this case. In particular, zero flows will be transformed into quite high flows if $\mu_t/s_t > \mu_b/s_b$, which is almost always the case if the seasonality of flows is not taken into consideration, as the observed change in the standard deviation of flows is much higher in wetter months than in drier months.

- **Table 2-2: Comparison of mean and standard deviation values for overall Pre-97 and Post-97 flows for Sheepwash Creek**

	Mean (μ)(ML)	**Standard deviation (s) (ML)**	**μ/s**
Pre 1997	145	338	0.43
Post 1997	49.1	66	0.74



- **Figure 2-6 Flow duration curve for transformed series using standardisation**

Table 2-2 for Sheepwash Creek in the Latrobe River basin illustrates the inadequacy of this method where there are frequent occurrences of zero flows and Figure 2-6 shows the flow duration curve for the transformed series.

2.9.2 Standardisation using monthly flow properties

If the effect of seasonality is taken out of this method by carrying out the standardisation process for each month, more intuitively acceptable results are obtained because the relative change between the mean and standard deviation will be preserved for each month. Table 2-3 shows the comparative values of means and standard deviation for pre-97 and post-97 flows.

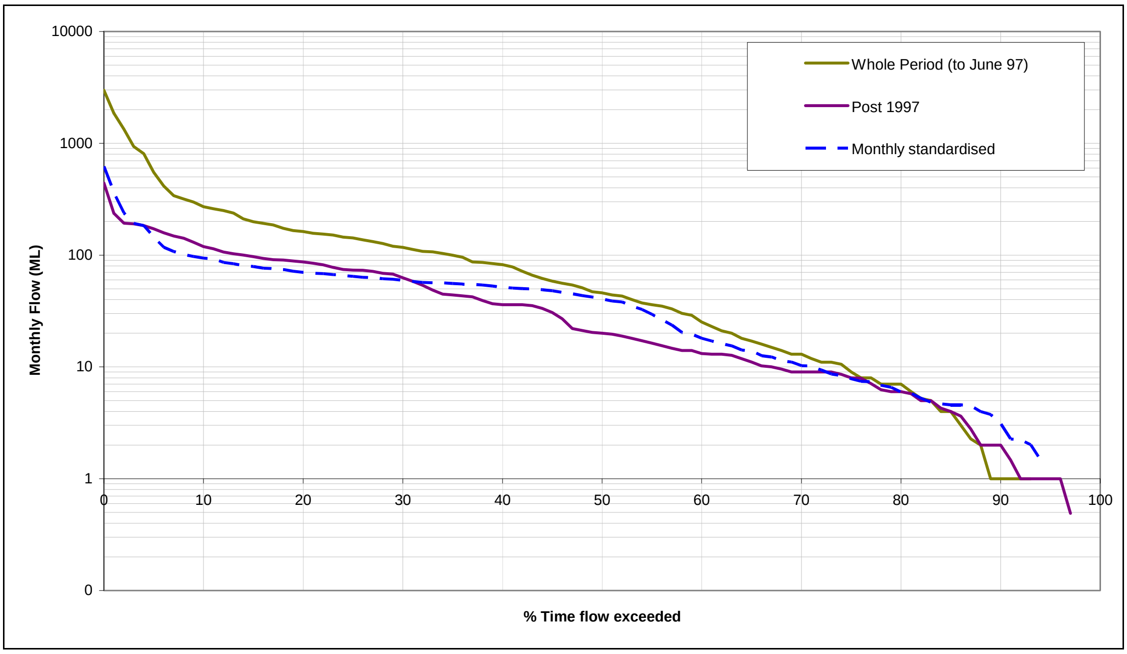
However, as can be seen, the ratio (μ/s) is only lower in the post 1997 case for the months of April and December. This means that in all other months, very low flows will automatically be shifted up. The flow duration curve of the transformed series is a better approximation of the observed post-1997 curve but only because of low flows now being introduced in the months of April and December.

The adjusted flow series has the same mean and standard deviation as the post 1997 series if the negative numbers arising out of applying Eq. 3 are kept in the series. When these numbers are corrected to zero, the resulting mean is 49.3 ML and standard deviation equals 65.9 ML.

The above method is considered flawed as the resulting times series' flow distribution is not a very good fit to the base series (Post -97), as seen in Figure 2-7, although the mean and standard deviation of the transformed series are the same. The above method seems to be an ineffective way of compressing the time series as the transformation to be applied to individual data points is very highly dependent on the monthly averages occurring from a small sample size (10 years of data).

- **Table 2-3: Comparison of mean and standard deviation values for each month of Pre-97 and Post-97 flows**

Month	Pre 1997 mean (μ) (ML)	Pre 1997 std dev. (s) (ML)	Pre 1997 μ/s	Post 1997 mean (μ) (ML)	Post 1997 std dev. (s) (ML)	Post 1997 μ/s
Jan	14	22	0.6	21	22	1.0
Feb	13	25	0.5	7	4	1.7
Mar	18	36	0.5	6	8	0.8
Apr	27	33	0.8	20	33	**0.6**
May	93	115	0.8	27	24	1.1
Jun	170	456	0.4	63	50	1.3
Jul	272	431	0.6	64	44	1.5
Aug	414	641	0.6	125	136	0.9
Sep	373	398	0.9	96	64	1.5
Oct	233	384	0.6	92	61	1.5
Nov	73	68	1.1	49	40	1.2
Dec	37	42	0.9	21	34	**0.6**



- **Figure 2-7 Flow duration curve for transformed series using standardisation on a monthly basis**

Therefore two attempts were made to fit theoretical distributions (with defined statistical characteristics) to the observed flows over the post-97 period such that the target series could then be adjusted back to fit that theoretical distribution based on the property of an individual point within the pre-97 time series. The two methods used were:

1. Transforming the flow into a normal distribution
2. Fitting a Beta-distribution to the Flow Duration Curve

2.10 Transformation into a normal distribution

An attempt is here made by transforming the time series of monthly flows both pre and post 1997 into approximate normal distributions before applying the standardisation method with the seasonal correction.

Logarithmic transformations, power transformations or Box-Cox transformations are commonly used to transform data series into a normally distributed series (Maidment, 1993). In this case the logarithmic transformation cannot be applied because of zero flow values.

The Box-Cox transformation is used here by firstly transforming the pre 1997 and post 1997 series using the formula $Y_t = (X^{\lambda} - 1)/ \lambda$, where λ is selected such that the transformed series has a skewness value of zero. Using Sheepwash Creek as an example, values of λ of 0.211 for pre 1997 and 0.268 for post 1997 were used. Applying this transformation ensures that two normal distributions with defined statistical properties are generated, which enables the replication of the required distribution properties.

▪ **Table 2-4 Comparison of mean and standard deviation values of the normalised flow for each month of Pre-97 and Post-97 flows**

Month	Pre 1997 mean (μ)	Pre 1997 std (s)	Pre 1997 μ/s	Post 1997 mean (μ)	Post 1997 std (s)	Post 1997 μ/s
Jan	1.7	3.3	0.5	3.8	2.3	1.6
Feb	0.3	4.1	0.1	2.2	1.1	1.9
Mar	1.9	3.5	0.5	0.0	3.5	**0.0**
Apr	3.5	2.8	1.3	3.0	2.8	**1.1**
May	6.0	3.3	1.8	4.6	2.1	2.2
Jun	7.4	3.2	2.3	7.0	2.2	3.2
Jul	9.0	3.6	2.5	7.1	2.2	3.2
Aug	10.3	3.6	2.8	8.5	3.6	**2.3**
Sep	10.5	3.3	3.2	8.3	2.4	3.5
Oct	8.6	3.3	2.6	8.1	2.7	3.0
Nov	5.9	2.8	2.1	5.9	2.9	**2.0**
Dec	4.0	3.3	1.2	3.0	2.9	**1.0**

The means, standard deviations and μ/s ratios for the transformed time series are shown in Table 2-4. The units are here transformed ML.

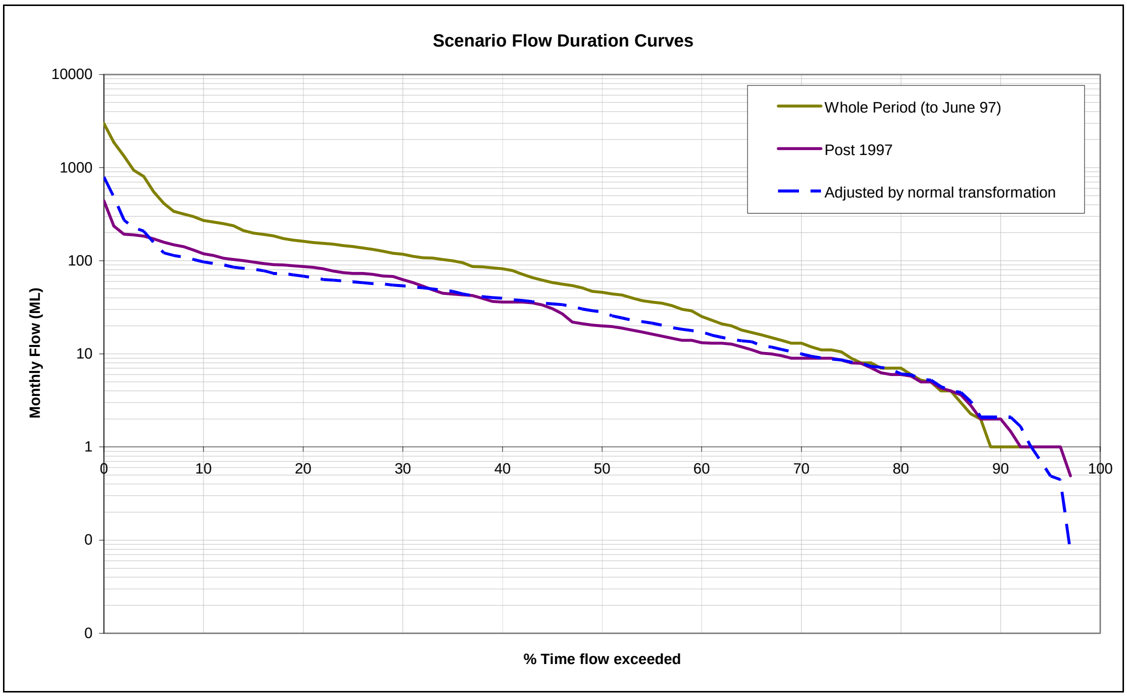
It is seen that this method would be generally effective at ensuring that low flows are preserved in drier months. The discrepancies for January and February can be explained by the fact that zero flows have not been observed in the post 1997 period while there were a number of occurrences in the pre 1997 period.

The transformed series for the pre 1997 period is then firstly standardised (Eq. 1). Then the corresponding value of post 1997 period transformed flow for the same standard value is obtained by applying Eq. 3 at each monthly time step.

$$Y_i' = Z_t \, . \, s_b + \mu_b \qquad \text{Eq. 3}$$

This transformed flow is then turned back into a flow series by reversing the Box-Cox transformation using the parameters obtained for the post 1997 period, i.e.

$$Y_i' = [(Y_t \text{ x } 0.268) + 1]^{(1/0.211)} \qquad \text{Eq. 4}$$



- **Figure 2-8 Flow duration curve for transformed series using normalisation on a monthly basis**

Instances of this equation being inapplicable owing to $[(Y_t \,.\, \lambda) + 1] < 0$ are attributed a transformed flow of zero.

Figure 2-8 shows that the obtained flow distribution is very similar to the post-97 flows, albeit a smoother version as a result of fitting a normal distribution.

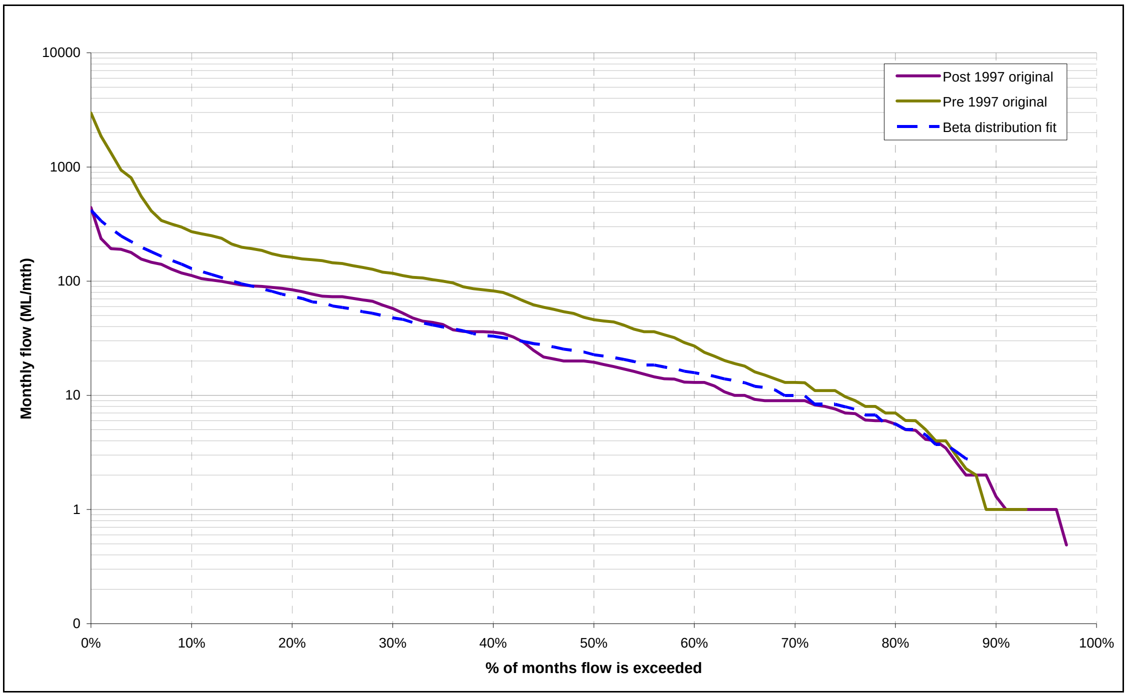
2.11 Fitting a Beta distribution to the Flow Duration Curve

In this method, a theoretical beta-distribution is fitted to the flow duration curve of the selected time series. The method is equivalent to method 2 (matching flow percentiles) except for the fact that a theoretical distribution is matched rather than the actual flow distribution. The beta distribution is commonly used to study variation in the percentage of a variable across samples and is well suited for this purpose. By carefully adjusting four defining parameters: the 90th percentile flow, the 10th percentile flow, the probability of exceedance of the observed post 1997 peak flow and the cease to flow period, a beta distribution can be calibrated to give a good fit to a typical flow duration curve for a natural stream (unregulated and with no diversions).

The advantage of this method is that a judgement can be made on whether to assume that the flow duration curve of the post 1997 period is a steady state curve (in which case it would be required to exactly match it) or potentially a distribution that is likely to smoothen out (if it contains some discontinuities or uneven curvature) with additional data or better data (for any farm dams, stock and domestic, or dairy diversions) added back to the gauged flow. In doing so, the exact match to the mean or standard deviation of the base series would be intentionally forfeited.

The pre-97 time series can then be transformed into a post 1997 period equivalent series by matching the percentile value of the pre 1997 period flow point to the flow value on the beta-distribution.

Figure 2-9 shows the fit obtained for Sheepwash Creek. The 3 parameters used to calibrate this curve, compared to the values extracted from the base series are given in Table 2-5. In this case, the cease to flow period for the transformed series has been set to be slightly more than during the pre 1997 period (6%) rather than the post 1997 period (3%), as it is intuitive that periods of zero flow in the pre 1997 period would also be zero flows under post 1997 conditions. It was also seen that attributing the probability of exceedance of the observed post 1997 peak flow a value of 3% enhances the beta-distribution fit. The means and standard deviations of base series and of the beta distribution are also provided in Table 2-5.



- **Figure 2-9 Flow duration curve for beta-distribution fit**

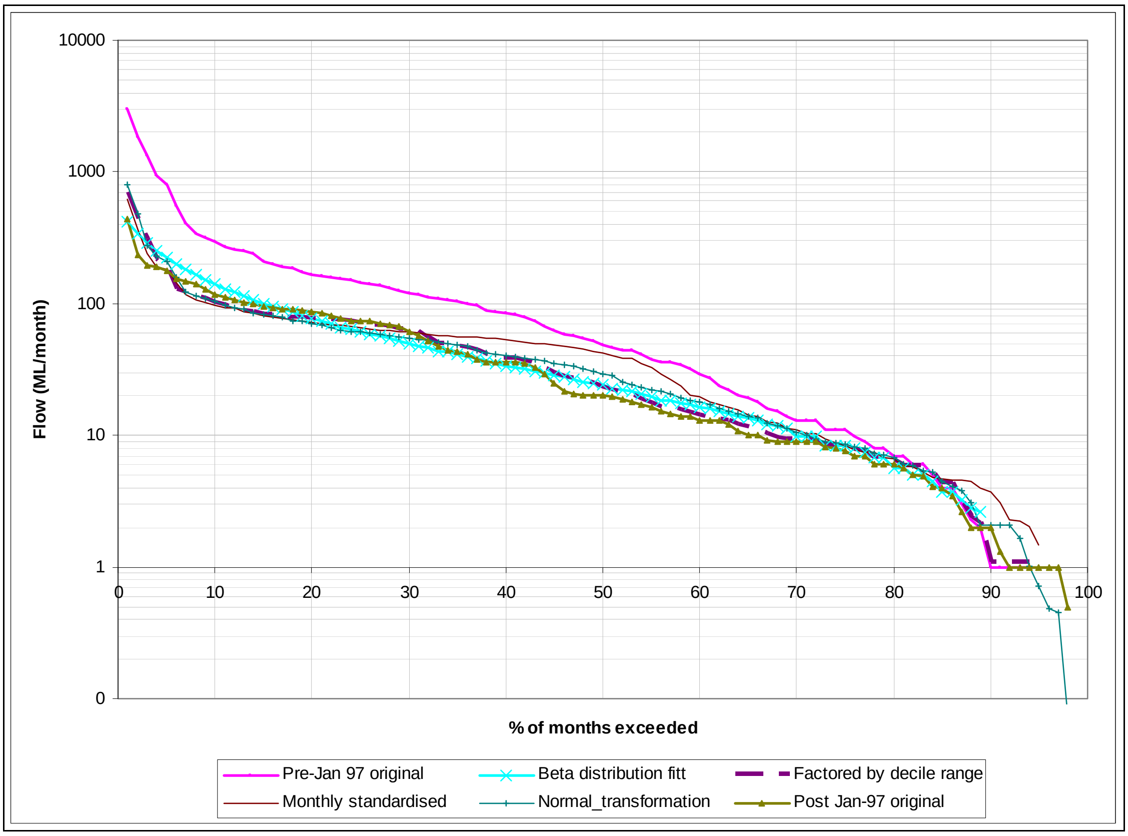
- **Table 2-5 Comparison of statistical properties of observed post-97 flow series and fitted beta-distribution for Sheepwash Creek**

	Base series (post-97)	Fitted Beta-distribution
90% percentile flow (ML)	121	172
10% percentile flow (ML)	3	4.4
% of time with zero flow	4%	6.5%
Mean (ML)	49	49
Standard deviation (ML)	66	69

2.12 Conclusions and Recommendations

The mean and standard deviation of the transformed series were compared with results of other methods in Table 2-1.

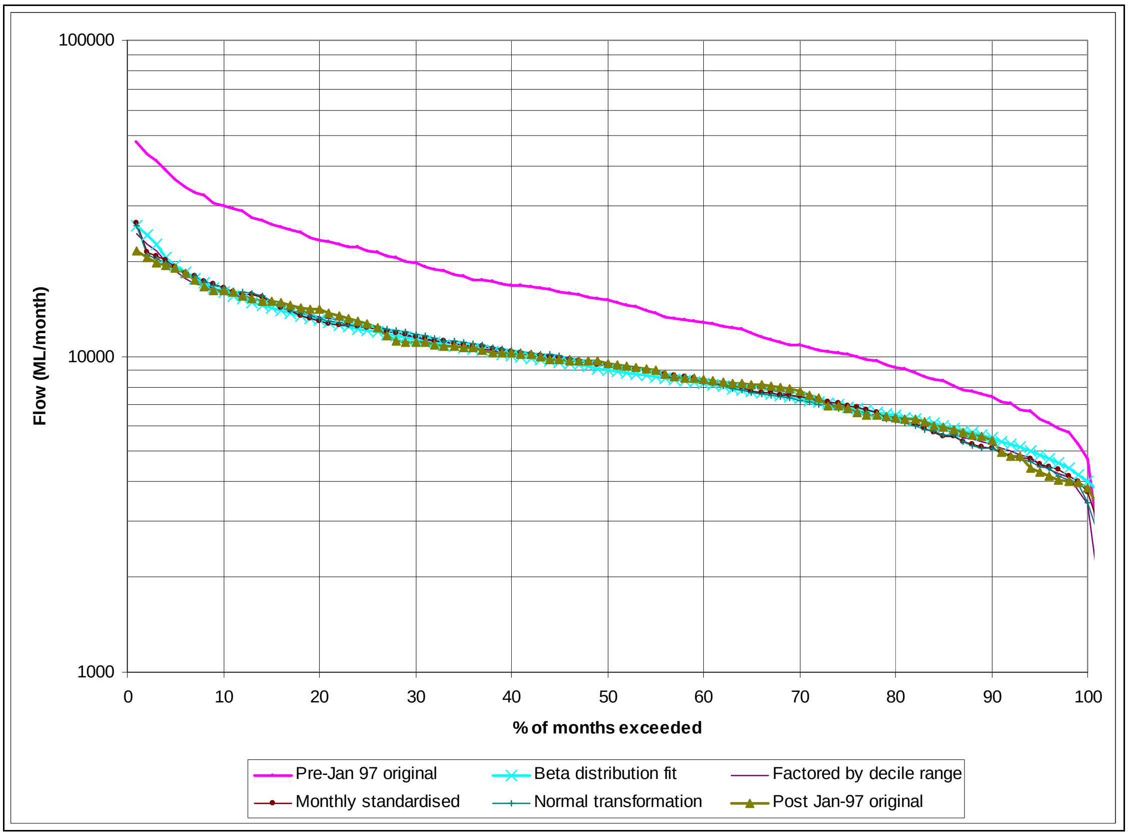
Figure 2-10 provides a comparison of the flow duration curves obtained when applying methods 2 to 5 to the Sheepwash Creek flow time series. It is seen that at low flows, method 5 has the advantage of allowing a preset cease to flow percentage of time to be defined. However, at high flows, all the methods tend to generate higher flows than experienced in the post 1997 period. The bulk of the divergence between these methods occurs in the range of flows between the 20% and 50% flow exceedance percentiles (80 ML to 20 ML). The difference between the flows generated from each method in that flow range was seen to be up to 40 ML, which is quite significant as the average flow is 49 ML/month. In this particular case, method 5 would be conservative.



- Figure 2-10 Comparison of flow duration curves obtained for Sheepwash Creek inflows

- Table 2-6 Comparison statistics for Blue Rock inflows

	Mean (ML/mth)	Standard deviation (ML/mth)
Post 1997	**9919**	**4241**
Percentile adjustment	9924	4145
Decile Range adjustment	9930	4233
Standardisation	9919	4241
Normalisation	9919	4263
Beta-distribution	9900	4275



- **Figure 2-11 Comparison of flow duration curves obtained for Blue Rock inflows at S204**

The same analysis was carried out for the inflows to Blue Rock Lake where there are no zero flows. The statistics shown in Table 2-6 are obtained.

As can be seen from Figure 2-11, the flow duration curves can all be considered as good matches for the original post -97 flows.

From the findings above, it is recommended to use the factoring by percentile range method to transform the flows from pre-97 inflows with post-97 characteristics on the basis of simplicity of method application. It is also believed that the method of fitting a beta-distribution to the flow duration curve of post-97 flows could be a very practical way of overcoming issues related to uncertainty in flow measurements and demand estimations and should be considered for future applications if time is not a factor, as it requires some calibration.

Based on all the investigations described in this section, the recommended methods were factoring of the flow duration curve by percentile or decile. The fitting of a beta-distribution was also a favoured method, particularly if uncertainty in flow parameters is a problem, but is quite labour intensive.

These findings were used as the basis to develop an automated tool that was able to employ the comparison of flow duration curves method (both percentile and decile). The way that this tool applies this method is described in detail in Section 4. Furthermore, the tool was to offer one other method: factoring by season. This method was included for consistency with methods employed by CSIRO, and is described in Section 3.

3. Seasonal Factoring Method

This section gives a brief description of the process of transformation for the simpler of the two available methods: Seasonal Factoring.

In the Streamflow Transformation utility, the user is prompted to enter a “break point date” that will split the flow record into two sections, composed of:

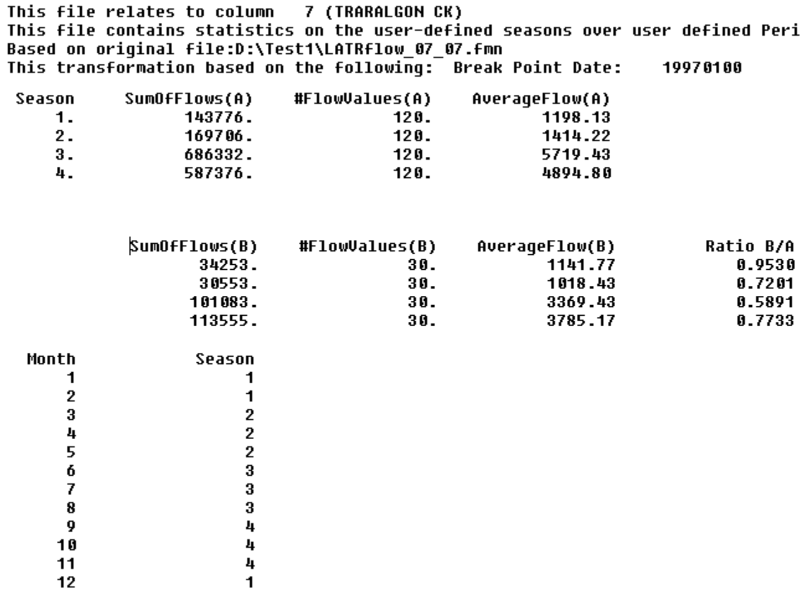
- Period A, the period including all dates prior to the break point date; and
- Period B, the period including the break point date and all subsequent dates.

  (Period A will be transformed to be given similar properties to Period B).

The user can also define the seasonal regime to be used in the transformation. As this definition is on a monthly basis, the maximum number of seasons is 12. Once the seasons are defined, the following steps are followed to complete the transformation:

- For Period A, all the flows belonging to the first season are grouped, and an average flow is found for that season/month under Period A. This is done for all seasons that the user has defined.
- The above step is repeated for Period B.
- The ratio of the $\text{Average}_{\text{Period B}}$ to the $\text{Average}_{\text{Period A}}$ is found for each season/month.
- For each flow value in Period A, a transformed flow is derived by multiplying by the factor that is applicable for that season.

If the user requires information on this transformation process, the utility allows the user to elect to create a “Seasons Information File”. This file contains the statistics calculated for this method, including the sum of flows, the number of flow values and the average flow for each season in each period. The ratio of average flow in period B to period A, which is the factor used to transform the period A flows, is also included. An example of such a file is given in Figure 3-1 below. The corresponding original and transformed flow series are given in Figure 3-2 below.



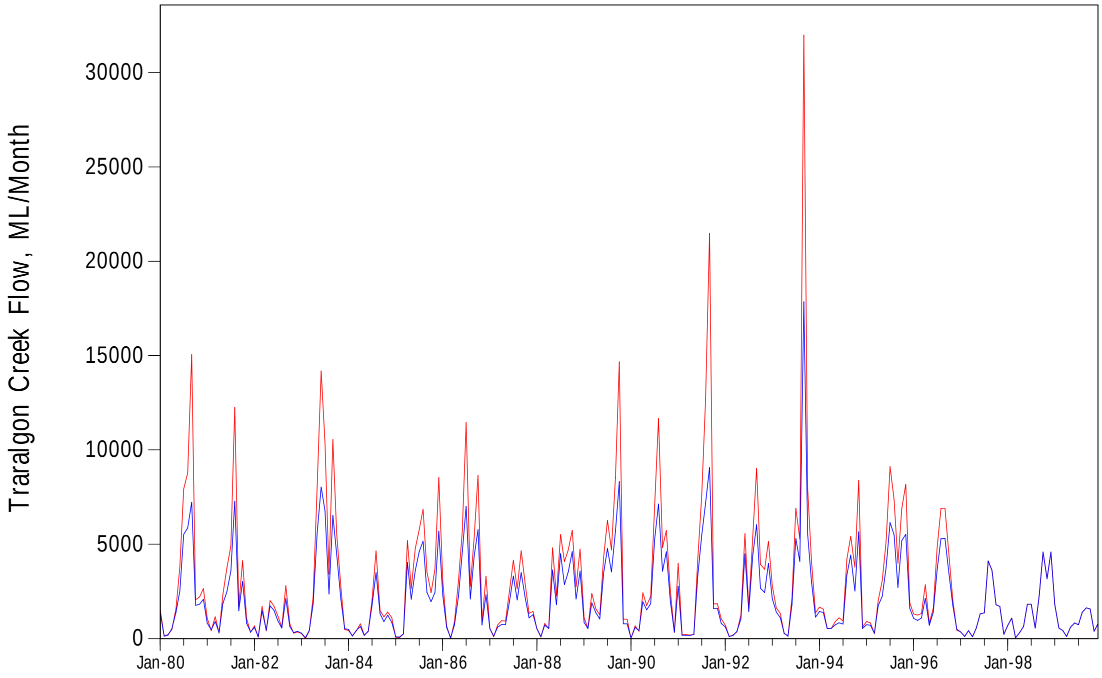
```
This file relates to column   7 (TRARALGON CK)
This file contains statistics on the user-defined seasons over user defined Peri
Based on original file:D:\Test1\LATRflow_07_07.fmn
This transformation based on the following:  Break Point Date:    19970100

 Season      SumOfFlows(A)     #FlowValues(A)     AverageFlow(A)
     1.           143776.               120.           1198.13
     2.           169706.               120.           1414.22
     3.           686332.               120.           5719.43
     4.           587376.               120.           4894.80


             SumOfFlows(B)     #FlowValues(B)     AverageFlow(B)          Ratio B/A
                   34253.                30.           1141.77             0.9530
                   30553.                30.           1018.43             0.7201
                  101083.                30.           3369.43             0.5891
                  113555.                30.           3785.17             0.7733

  Month              Season
      1                   1
      2                   1
      3                   2
      4                   2
      5                   2
      6                   3
      7                   3
      8                   3
      9                   4
     10                   4
     11                   4
     12                   1
```

- **Figure 3-1: Information contained in the "Seasons Information" file (…si.txt), for Traralgon Creek seasonal transformation, 1957 – 2006.**



- **Figure 3-2 Original (Red) and transformed flows (Blue) for Traralgon Creek during the 1990s, subject to the seasonal transformation described in Figure 3-1 above, with a January 1997 breakpoint (this a portion of the series 1957 – 2006).**

It should be noted that the seasonal factoring method includes the ability to specify only a single season. This basically means that flow in period A is transformed by the ratio of mean flow across all of period B to mean flow across all of period A. This is consistent with the methodology that has been commonly used in the past to transform streamflow data across a break point.

Additionally, using only one season avoids potential discontinuities in the transformed flow data, particularly at low flows. For example, if two seasons are selected it is possible that the ratio of period B to period A flows in season one could be quite different to season two. As a result, there could be a significant sudden change in the transformed flow from the last time step of season one to the first time step of season two. It is recommended that the transformed flow data be inspected for such discontinuities. Using the alternative method of transformation by comparison of flow duration curves overcomes this limitation to some extent.

4. Comparison of Flow Duration Curves Method

The aim of this method is to apply a transformation to a streamflow series, with the intention of giving the flows before a pre-defined "break point" date, similar properties to flows after that date. This section explains the process of transformation in detail, for the "comparison of flow duration curves" method. It refers to each individual step undertaken by the algorithm.

4.1 Step-by-step explanation of algorithm

After reading the input flows, the first steps taken are to prompt the user for the "break point date" and then split the record into two sections, composed of:

- Period A, the period including all dates prior to the break point date; and
- Period B, the period including the break point date and all subsequent dates.

  (Period A will be transformed to be given similar properties to Period B).

Next, the user is prompted for the number of segments or "bins" to use in the matching process. Table 4-1 contains remarks about this choice. The table refers to the use of 1, 4, 10 or 100 bins, but the program allows the user to pick any number of bins. The only restriction is that the number of bins must not exceed the number of data points in Period A or Period B.

The following example is illustrative of these principals. *This example is vastly simplified* relative to anticipated scenarios for this program (using only 4 years of data) but will serve the purpose of illustration. Suppose the program was to be applied to the series shown on the left of Figure 4-1, and that the series was subject to the break-point date of June 1993. This would lead to the two periods shown on the right of Figure 4-1. Period A is of length 2 years 5 months, or 29 flow values. Period B is of length 1 year 7 months, or 19 flow values. Thus, factoring by quartiles or deciles is acceptable, but percentiles is not because using 100 bins exceeds the limit of 19 data points. The number of bins can be any of 1, 2, 3, ….. 18, 19.

The following discussion will consider this example, using 1, 2, and 10 bins.

- **Table 4-1: Explanation of input “Number of Segments”**

Number of segments used	**Result**
1	Flow factoring will be based on ratio of flow averages for Periods A and B.
4	Flow factoring will be based on quartiles resulting from splitting the flow duration curves into 4 equal portions. Eg. Factors for Period A top 25% flows will be based on ratio with Period B top 25% flows, and so on for other 3 quartiles.
10	Flow factoring will be based on deciles. Eg. Factors for Period A top 10% flows will be based on ratio with Period B top 10% flows, and so on for other 9 deciles.
100	Flow factoring will be based on percentiles.

Figure 4-1: Monthly Flow series for "Example Creek" case study, in ML/month

DATE	FLOW
199101	3994
199102	215
199103	213
199104	186
199105	223
199106	4149
199107	7587
199108	12648
199109	21472
199110	1837
199111	1842
199112	1051
199201	695
199202	105
199203	189
199204	349
199205	1268
199206	5565
199207	1670
199208	5412
199209	9029
199210	3932
199211	3674
199212	5154
199301	2798
199302	1626
199303	1337
199304	265
199305	134
199306	2161
199307	6912
199308	5275
199309	2981
199310	3944
199311	4093
199312	1366
199401	1666
199402	1554
199403	544
199404	533
199405	893
199406	1095
199407	941
199408	4204
199409	5421
199410	3785
199411	2383
199412	619

Subject to break point date:
June 1993

DATE	FlOW
PERIOD A	
199101	3994
199102	215
199103	213
199104	186
199105	223
199106	4149
199107	7587
199108	12648
199109	21472
199110	1837
199111	1842
199112	1051
199201	695
199202	105
199203	189
199204	349
199205	1268
199206	5565
199207	1670
199208	5412
199209	9029
199210	3932
199211	3674
199212	5154
199301	2798
199302	1626
199303	1337
199304	265
199305	134
PERIOD B	
199306	2161
199307	6912
199308	5275
199309	2981
199310	3944
199311	4093
199312	1366
199401	1666
199402	1554
199403	544
199404	533
199405	893
199406	1095
199407	941
199408	4204
199409	5421
199410	3785
199411	2383
199412	619

4.2 Case using 1 bin

The following case studies refer to Figure 4-1 above. If the user elected to use 1 bin, the transformed flows in Period A would be calculated using the formula

$$PeriodA_{Transformed}(i) = PeriodA(i) \times \frac{AverageFlow_B}{AverageFlow_A}$$

$AverageFlow_A$ = 3400.7 ML/month and $AverageFlow_B$ = 2651.1 ML/month from the figures above, so each of the Period A flows above are transformed by multiplying by the factor 2651.1 / 3400.7 = 0.7796

4.3 Case using 2 bins

Next we will consider the case using 2 bins. Let us examine the 19 flow values in Period B:

2161 6912 5275 2981 3944 4093 1366 1666 1554 544 533 893 1095 941 4204 5421 3785 2383 619

The method requires that we divide these into two bins: the top 50% and the bottom 50% of flows. To do this, we first sort the flows.

533 544 619 893 941 1095 1366 1554 1666 2161 2383 2981 3785 3944 4093 4204 5275 5421 6912

The division of this set into two bins results in the situation shown below. The first nine data points are put in the first bin, and the last nine in the second. However, if we wish to include all points in our analysis, then we must apportion the median flow (boxed) between the two bins.

$$\underbrace{533\ 544\ 619\ 893\ 941\ 1095\ 1366\ 1554\ 1666}_{\text{BIN 1}}\ \boxed{2161}\ \underbrace{2383\ 2981\ 3785\ 3944\ 4093\ 4204\ 5275\ 5421\ 6912}_{\text{BIN 2}}$$

The average flows for each bin are then calculated as:

$$Average_{BIN1,PeriodB} = \frac{533 + 544 + 619 + 893 + 941 + 1095 + 1366 + 1554 + 1666 + ½(2161)}{9.5}$$

$$Average_{BIN2,PeriodB} = \frac{½(2161) + 2383 + 2981 + 3785 + 3944 + 4093 + 4204 + 5275 + 5421 + 6912}{9.5}$$

Once the equivalent Period A averages have been calculated, the transformed flows can be calculated as shown below:

If the flow to be transformed is less than the median flow (bottom 50%):

$$PeriodA_{Transformed}(i) = PeriodA(i) \times \frac{AverageFlow_{BIN1,PeriodB}}{AverageFlow_{BIN1,PeriodA}}$$

If the flow to be transformed is greater than the median flow (top 50%):

$$PeriodA_{Transformed}(i) = PeriodA(i) \times \frac{AverageFlow_{BIN2,PeriodB}}{AverageFlow_{BIN2,PeriodA}}$$

If the flow to be transformed is the median flow:

In the case of the median flow, the factor used will be the average of the two factors above.

4.4 Case using 10 bins

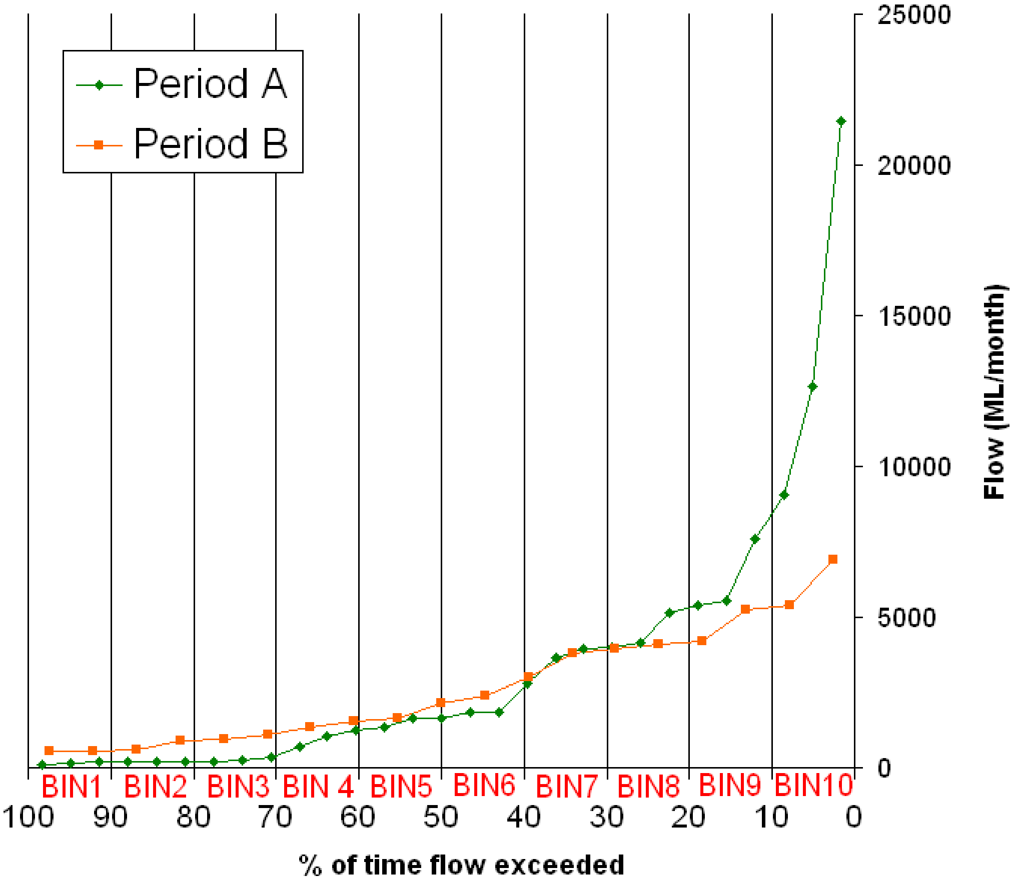
To explain the case using 10 bins, it becomes more convenient to refer to the flow duration curve of the given flow series. The flow duration curves for each of Period A and Period B are given below in Figure 4-2. Note that the x-axis is displayed in reverse order compared to the usual way of viewing flow duration curves.

Note the method of plotting the x-coordinates of the flow duration curve: the first point is not plotted at 100% and the last point is not plotted at zero. Instead, each of the 29 green points is allotted a portion of the x axis ($1/29^{th}$) and the point is plotted **in the centre** of this portion.

In the case of Period A, the task involves apportioning 29 points of data into 10 bins. The example of Period A will be used to explain the process that the program uses. It involves the following steps:

Step 1.
As with the method for 2 bins, the first step is to sort the flows by magnitude.

Step 2
Allocation to each of the 29 flow values a portion of the x-axis of the flow duration curve.
For example, the smallest flow value in Period A gets the space between 0 and 1/29, the next smallest gets the space between 1/29 and 2/29 etc., until finally the largest value gets the space between 28/29 and 1.



- **Figure 4-2: Flow Duration Curve for the example flow series. Note that the x-axis is reversed and divided into 10 equal segments or "bins".**

Step 3.

Flow values are apportioned to a single bin, or between two bins, as required.

By comparing the space allocated to a flow in Step 2 with the bin space(s). For example, the smallest flow in Period A is allocated the space 0 to 1/29 (0.0345), and the first bin is allocated 0 to 0.1. The following equations

$lower\ bound_{flow\ value} \geq lower\ bound_{bin1}$[1] and also

$upper\ bound_{flow\ value} \leq upper\ bound_{bin1}$

---

[1] Note on the terminology: "lower bound flow value" represents an x-value that is the *lower bound of the space in the x-domain allocated to the given flow value.* Likewise "lower bound bin 2" refers to the *lower bound of the space in the x-domain allocated to bin 2* (in this case, 0.1).

are both true for the first value, so the first (smallest) flow value is apportioned wholly to the first bin. The same is true for the second flow value, for which lower bound flow value = 1/29 (0.0345) and upper bound flow value = 2/29 (0.0690). However, the third flow value has lower bound flow value= 2/29 (0.0690) and upper bound flow value= 3/29 (0.1034).

Therefore:

$\text{lower bound}_{\text{flow value}} \geq \text{lower bound}_{\text{bin1}}$ gives $0.0690 \geq 0$ (true)
$\text{upper bound}_{\text{flow value}} \leq \text{upper bound}_{\text{bin1}}$ gives $0.1034 \leq 0.1$ (false)

This false statement means that the third flow value needs to be apportioned between bin 1 and bin 2. The portion in bin 2 is as follows:

$$\frac{\text{upper bound}_{\text{flow value}} - \text{lower bound}_{\text{bin2}}}{\text{upper bound}_{\text{flow value}} - \text{lower bound}_{\text{flow value}}} \text{ which gives } \frac{0.1034 - 0.1}{0.1034 - 0.0690} = 10\%$$

Another way of expressing the above mathematics is to say that the third flow value is 10% in Bin 2. The contribution to Bin 1 will be 90% (the complement of 10%), as follows:

$$\frac{\text{upper bound}_{\text{bin1}} - \text{lower bound}_{\text{flow value}}}{\text{upper bound}_{\text{flow value}} - \text{lower bound}_{\text{flow value}}} \text{ which gives } \frac{0.1 - 0.0690}{0.1034 - 0.0690} = 90\%$$

Because this third flow value is partially in Bin 1 and partially in Bin 2, in the Streamflow Transformation Utility, this value would be referred to as a "**partial**".

Step 4.

Calculation of flow averages for each bin.

For example, the flow average for Bin 1, Period A will be:

$$\frac{\text{flow value}\,1 + \text{flow value}\,2 + 0.9(\text{value}\,3)}{2 + 0.9}$$

Given the smallest three flows in Period A are 105, 134, and 186 ML/mth, this gives

$$\frac{105 + 134 + 0.9(186)}{2.9} = 140.14 \text{ ML/month}$$

Similar calculations are then done for bins 2 to 10 for Period A. The whole process (Steps 1 to 4) is repeated for Period B, bins 1 to 10.

Step 5.

Transformed flows are calculated using the ratio of averages as the conversion factor.

This step has already been examined in the case using 2 bins above. The only difference is that instead of 2 ratios, 10 ratios are calculated (1 for each bin). For each flow to be transformed, the applicable ratio is the one calculated for the applicable bin. For example, if the flow is from bin 4, then the applicable ratio will be:

$$\text{Ratio}_{\text{BIN4}} = \frac{AverageFlow_{bin4,PeriodB}}{AverageFlow_{bin4,PeriodA}}$$

If the flow to be transformed is a *partial* (shared between two bins) then a weighted average of ratios will be used. For example, if the flow is 80% in bin 4 and 20% in bin 5 then the following conversion factor will be used:

$$(0.8 \times \text{Ratio}_{\text{BIN4}}) + (0.2 \times \text{Ratio}_{\text{BIN5}})$$

4.5 Conclusions

Case studies for 1 bin, 2 bins and 10 bins are given above. The description for 10 bins gives insight into how the utility:

- Apportions flows between bins, where necessary; and
- Calculates ratios for each bin.

The utility offers the user any number of bins, provided the number of bins does not exceed the number of flow values in Period A or Period B. The same principles used to do the above tasks with 10 bins can be extended to any number of bins (and much longer flow series than those used in this example).

5. Analyses for identification of 'break-point' date

The Streamflow Transformation utility provides the option to run two different analyses that may assist in choosing the most appropriate break point. The options are available through Utilities > Streamflow Transformation, then on the second window ("Enter Break Point Date") pressing the button "Break Point Date Analysis". Using this feature produces one file for each of the two analyses, containing data that can be plotted in REALM for interpretation by the user.

It should be noted that these analyses are intended to act as a guide only to assist the user in break point date identification. Selection of the appropriate break point can be a complex and subjective task and may require additional statistical analysis beyond what is provided by the utility. In the past, statistical measures such as a paired t-test have been successfully used to compare hydrological properties across a predetermined break point.

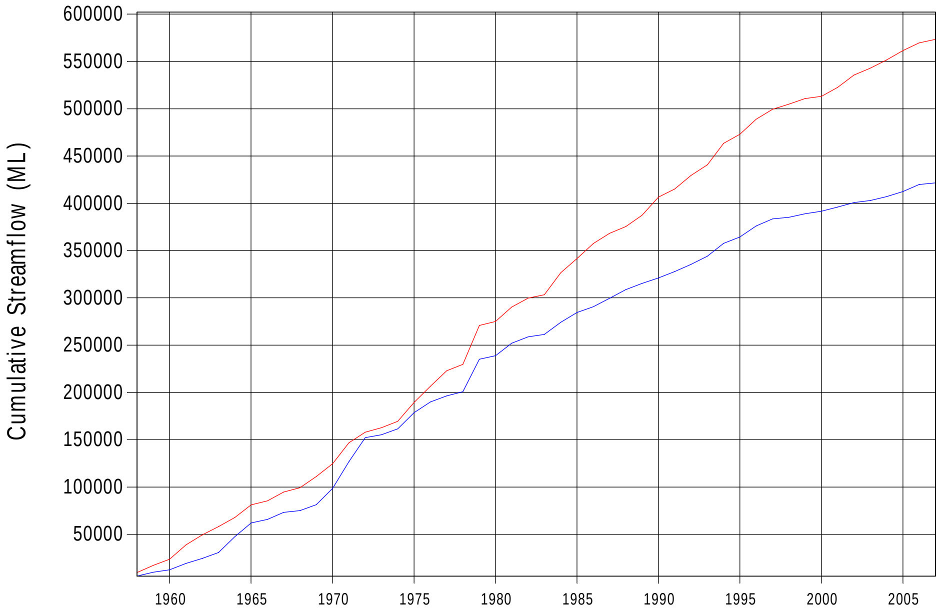
The two analyses provided by the utility are:

1) Single Mass Curve Analysis
2) "Ratio of Averages" Analysis

The following discussion provides some information on these two methods of analysis.

5.1 Single Mass Curve Analysis

This common analysis involves plotting cumulative streamflow versus time. The vertical axis of a single mass curve measures volume (eg. ML or GL) and the line itself represents the integration ("area under the curve") of the flow series. Example curves are given in Figure 5-1 below.



- **Figure 5-1: Plotting of a single mass curve file (…SMC.txt) in REALM, for Flynn's Creek (red) and Eaglehawk Creek (blue) in the Latrobe Catchment.**

The gradient of a single mass curve over a certain period is related to how wet or dry that period of time is. Steeper gradients indicate that flow was accumulating more quickly, meaning greater flow rates. Streamflow tends to vary according to the annual seasonal cycle; therefore, plotting single mass curves on a daily or monthly time-step tends to result in a jagged-edge, "saw-tooth" line. For this reason, the Streamflow Transformation utility creates all Single Mass Curve files (…SMC.txt) on an annual time-step for the calendar year (January to December). A possible future enhancement to the utility could be to allow the user to select the annual period over which such calculations are undertaken (eg July to June instead of January to December).

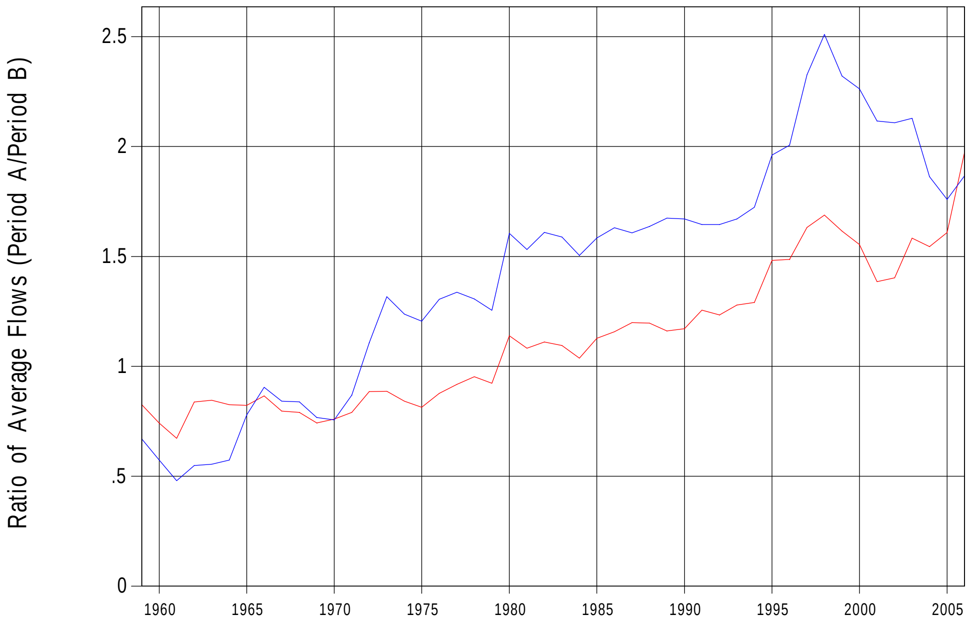
From Figure 5-1 above, it can be seen that the gradients of the single mass curves fluctuate with wet and dry periods (typically lasting a few years). However, it is important to note that the long-term average gradient, over 10 years or more, remains reasonably constant during the 60s, 70s and 80s. The more steady the long-term gradient over a period, the less appropriate it is to place a break point within that period. The most appropriate point to put a break point is where there is a *visible, long term change in gradient*. In the above figure, it can be seen that there is a period of around 10 years at the end of the flow record which appears to have a different (lower) gradient than the preceding years. Therefore, in this case, the most appropriate place for a break point is at the beginning of this later period (around 1997/98).

By default, the plot of Single Mass Curves produced by the utility: (1) contains only the series' selected by the user for transformation; and (2) begins at the start year of the file. At times it may be useful to have the curves start at a different (later) date, and this can be achieved by loading in a shortened REALM format file containing only the period that the user wishes to plot.

5.2 Ratio of Averages Analysis

This analysis assumes that the aim when choosing a break point date is to choose a date which maximises the differences in flow properties between Period A and Period B. The Ratio of Averages analysis focuses on only one flow property: average flow. For each date in the flow series, the utility calculates the average flow for the preceding period and the average flow for the following period. This information is written to a file (…ROA.txt) that can be plotted in REALM (an example is given in Figure 5-2 below). The user can then manually inspect the dates under consideration as break-points, to see how much greater the average flows before the date (Period A) are, compared to the average flows after the date (Period B). As for the single mass curves, the streamflow transformation utility calculates the ratios on an annual basis, for calendar years (January to December).

If the aim is to pick the date which provides maximum contrast between before-and-after, then the user should pick the date *corresponding to the maxima* of the Ratio of Averages plot. For example, in Figure 5-2 below the plot indicates that the most appropriate break-point date for this series is in 1998.

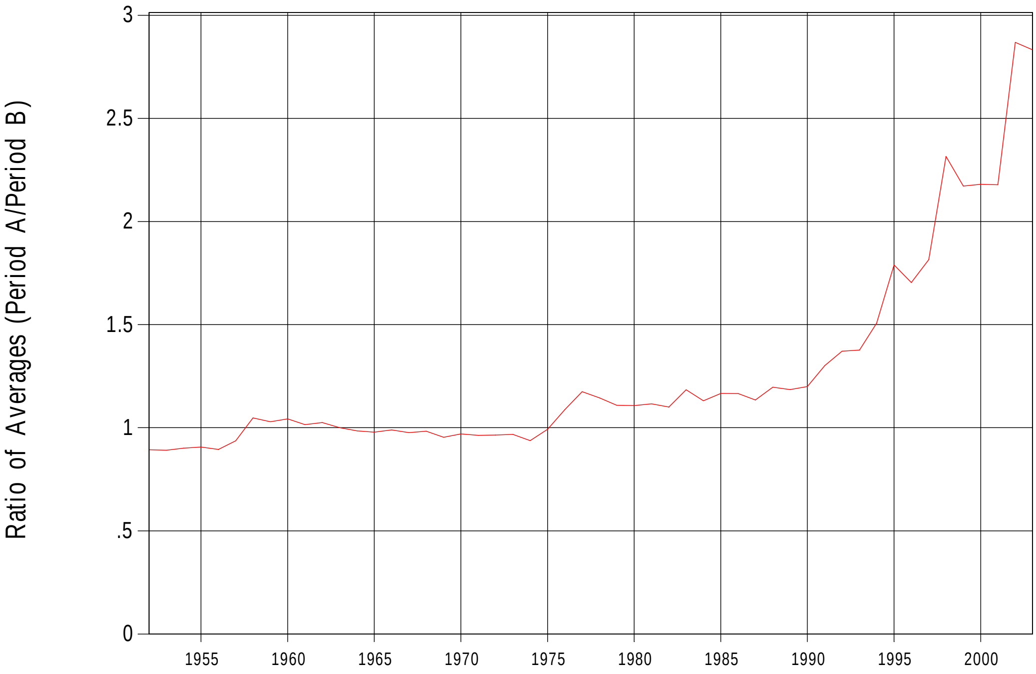


- **Figure 5-2: Plotting of a ratio of averages file (…ROA.txt) in REALM, for Flynn's Creek (red) and Eaglehawk Creek (blue) in the Latrobe Catchment.**

The Ratio of Averages analysis has a number of limitations, as follows:

- The periods at the very start and the very end of the file, do not provide reliable Ratio of Averages values. This is because the stability of the ROA values decreases as the number of values in either Period A or Period B declines. Although the final few years may contain the highest value of ROA, this value will be based on a very short period and, as such, is very sensitive to an update of data (say, when an extra year is added in a routine REALM update). Help notes in REALM outline various reasons why 10 years is the recommended minimum length for Period A or Period B.

- The assumption that climate change can be characterised by a single "step" change is a crude simplification. If the observed changes in flows are better characterised as a gradual change rather than a single step, this may lead to a Ratio of Averages value that continually increases with time, rather than coming to a global maxima. An example of this is shown in Figure 5-3 below. This can increase the difficulty of choosing a break point, and in many cases, it may be that no appropriate break point date exists due to this phenomenon occuring.



- **Figure 5-3 Plotting of a ratio of averages file (…ROA.txt) in REALM, for Hollands Creek in the Broken River Catchment. In this case, the successively higher peaks in ROA value make it difficult to choose a breakpoint using this method.**